Peter H. Cole · Damith C. Ranasinghe
Editors

Networked RFID Systems and Lightweight Cryptography

Raising Barriers to Product Counterfeiting

First edition

Springer

Peter H. Cole
University of Adelaide
School of Electrical
and Electronic Engineering
Auto-ID Lab
5005 Adelaide
Australia
cole@eleceng.adelaide.edu.au

Damith C. Ranasinghe
University of Adelaide
School of Electrical
and Electronic Engineering
Auto-ID Lab
5005 Adelaide
Australia
damith@eleceng.adelaide.edu.au

ISBN 978-3-540-71640-2 e-ISBN 978-3-540-71641-9

DOI 10.1007/978-3-540-71641-9

Library of Congress Control Number: 2007934348

Cover design: KünkelLopka, Heidelberg

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

Preface

The rapid growth of RFID use in various supply chain operations, which has arisen from the development of Electronic Product Code (EPC) technology, has created a need for the consideration of security issues in the adoption of that technology.

As the originators of EPC technology, the Auto-ID Center laboratories, established at MIT in 1999, and extended in subsequent years to become an association of seven laboratories around the world, have taken a keen interest in the workings of EPC in practical applications. The laboratories, now called the Auto-ID Laboratories, have adopted all questions surrounding security of these applications as a principal research interest. Their research has been primarily concerned with the ability of RFID to combat the widespread counterfeiting that has emerged in many supply chains and that is not adequately suppressed by non-RFID security technologies. This book is the outcome of that research.

The Auto-ID Laboratories network, whose members have provided the chapters of this book, consist of laboratories at The Massachusetts Institute of Technology in the USA, Cambridge University in the UK, The University of Adelaide in Australia, Keio University in Japan, Fudan University in China, The University of St. Gallen and The Swiss Federal Institute of Technology in Switzerland, and The Information and Communications University in Korea. Together, they have been and continue to be engaged in assembling the building blocks needed to create an "Internet of things". This global infrastructure leverages the global connectivity of the Internet and makes it possible for computers to identify any object worldwide. This Internet of things will not just provide the means to feed reliable, accurate, real-time information into existing business applications; it will usher in a new era of innovation and opportunity. More detail on the formation, functions and expertise of the Auto-ID Laboratories network, and its relation to world standards bodies, can be found in Chapter 1.

This book contains eighteen chapters divided into four sections. Section 1, entitled "Anti-counterfeiting and RFID", provides an introduction to EPC networks and the theory of security and authentication. Section 2, entitled "Security and Privacy Current Status", explains the current status of security and privacy concepts, some vulnerabilities of RFID systems, and defines a suitable evaluation framework for security objectives. Section 3, entitled "Network Based Solutions", explores the role of networks in achieving security and privacy objectives. Section 4, entitled "Cryptographic Solutions", shows how specific features built into RFID

tags and readers can enhance security and privacy objectives, and describes novel anti-counterfeiting technology.

It is not necessary for the chapters to be studied in a particular order, however, it should be noted that Chapter 1 provides a comprehensive outline of what is found in each of the subsequent chapters.

Each chapter is written by one or more acknowledged experts in the field. It has been a great pleasure to work with these authors in the production of this book.

I wish to sincerely acknowledge the efforts of my co-editor Damith C. Ranasinghe, who has not only assumed some of the significant burdens of editing, but has also made contributions to many of the chapters. In addition, I wish to express my appreciation to all of the members of the Auto-ID Laboratories who are responsible for the quality of this work. Additionally, I would like to thank the editorial staff of Springer Publishing, who have been unfailingly helpful throughout the production process.

Adelaide, Australia Peter H. Cole
12. September 2007

Contents

Chapter 1
Introduction from the editors

Structure of this book

This introduction describes the structure of the book, and in particular how it is divided into sections and chapters. It gives an outline of what can be found in each chapter, and gives a description of the origin and structure of the organisation known as the Auto-ID Laboratories whose members have studied the anti-counterfeiting problem and have provided the material for this book.

The four sections of the book are, as shown in the Table of Contents, entitled: 1: "Anti-counterfeiting and RFID" with four chapters; 2: "Security and Privacy Current Status" with four chapters; 3: "Network Based Solutions" with three chapters and 4: "Cryptographic Solutions" with six chapters.

The Auto-ID Laboratories

The Auto-ID Labs is the research-oriented successor to the Massachusetts Institute of Technology (MIT) Auto-ID Center, originally founded by David Brock and Sanjay Sarma of MIT with funding from Procter and Gamble, Gillette, the Uniform Code Council, and a number of other global consumer products manu-facturers. The MIT Auto-ID Center was created to develop the Electronic Product Code (EPC), a global RFID-based item identification system intended to replace the UPC bar code. In October 2003 the Auto-ID Center was replaced by the combination of the newly founded research network the Auto-ID Labs, and EPCglobal, an organization charged with managing the new EPC Network. The Auto-ID Labs are responsible for managing and funding continued development of the EPC technology.

From its foundation in 1999, the Auto-ID Center grew to become a unique partnership between almost 100 global companies and six of the world's leading research universities: the Massachusetts Institute of Technology in the US, the University of Cambridge in the UK, the University of Adelaide in Australia, Keio University in Japan, the University of St. Gallen in Switzerland, and Fudan University in China. Together they were and still are engaged in assembling the building blocks needed to create an "Internet of things" which is a global infra-structure – a layer on top of the Internet – that will make it possible for computers to identify any object anywhere in the world instantly. This network will not just provide the means to feed reliable, accurate, real-time information into existing business applications; it will usher in a whole new era of innovation and opportunity.

The Auto-ID Labs in March 2005 added Daejoen ICU University in Korea to their network, thus completing their organisation as the leading research group in

the field of networked radio-frequency identification (RFID) and emerging sensing technologies. The labs now consist of seven research universities located on four different continents. The areas of expertise range from hardware through software to business research related to RFID.

The research can be grouped into three main areas: hardware, software and business layer. On the autoidlabs.org website, the Auto-ID Labs continuously publish their research results and provide an archive with over 150 whitepapers and academic publications. The following shows how the network and the research are organized.

- Members

The research network now consists of the following seven research institutions:

– The University of Adelaide (Australia)
– The University of Cambridge (United Kingdom)
– Fudan University (China)
– The Information and Communications University (South Korea)
– Keio University (Japan)
– The Massachusetts Institute of Technology (USA)
– The University of St. Gallen/ETH Zurich (Switzerland)

The research is organised as follows.

- Business processes and applications

– Focus group: The University of St. Gallen/ETH Zurich, Keio University, The University of Cambridge, The Massachusetts Institute of Technology, The University of Adelaide
– Business cases
– Business applications
– Privacy and security aspects
– Fundamentally new business processes and industries which include payment, leasing, insurance, quality management, factory design, 3PL-managenemt, brand protection, and anti-counterfeiting amongst others

- Software and networks

– Focus group: Keio University, The Massachusetts Institute of Technology
– Future system architecture
– EPC network
– Middleware
– Integration with existing systems

- Hardware

– Focus group: The Massachusetts Institute of Technology, Fudan University, The Information and Communications University, The University of Adelaide
– RF and chip design

- Class 2 and higher tags
- Tags with memory, battery, sensors and actuators
- Enhanced reading rates in challenging environments

- External links

External web links related to the Labs are

- The Auto-ID Labs website is at http://www.autoidlabs.org/
- The EPCglobal website is at http://www.epcglobalinc.org/home

Section 1 Anti-counterfeiting and RFID

Chapter 2: "Anti-Counterfeiting and Supply Chain Security"

In Chapter 2 "Anti-Counterfeiting and Securing Supply Chains" can be found an overview of the anti-counterfeiting problem and what is needed to secure supply chains, and how this may be achieved. As its title suggests, the chapter deals with two issues: the problems raised by counterfeit products and the methods by means of which supply chains might be made secure.

The chapter makes at the outset an emphatic statement about intellectual property rights and their role in sustaining innovation and underpinning economic growth and employment.

The challenges for affected enterprises raised by the violation of intellectual property rights are described in detail. The requirements for Auto-ID based anti-counterfeiting solutions are derived from detailed studies of firstly attack models by means of which the behaviour of illicit actors may be understood, and the secondly the capabilities of low cost RFID transponders that may be used to counter such attacks.

A number of solution concepts, employing both RFID and optical technologies, are identified. These range from various forms of using unique serial numbers, through plausibility checks based on track and trace, to object specific security systems that are discussed in more detail in Chapter 13, to secure authentication systems based on enciphered responses to reader challenges. Such approaches are seen as providing motivation for the research that is the major topic of this book.

Then follow two chapters "Networked RFID Systems" and "EPC Network Architecture" that provide basic background on the context in which anti-counterfeiting security solutions must be devised.

Chapter 3: "Networked RFID Systems"

In Chapter 3 "Networked RFID Systems" the authors seek to identify concepts and operating principles of a modern RFID system. Although a wide range of operating principles for such a system, such as use of microelectronic labels, surface

acoustic wave labels, labels using multiple resonances to encode data and so on, are identified and referenced, the material presented in this chapter considers in detail RFID systems based on using microelectronic devices. It is noted that in general the operating principle and operating frequency are driven principally by the application of the labelling system and by the constraints provided by electromagnetic compatibility regulations, environmental noise, and the ability of fields to permeate a scanned region of space or to penetrate intervening materials.

All modern RFID system infrastructures are seen as consisting of the three primary components: (a) RFID labels (transponders); (b) RFID label readers or interrogators (transceivers); and (c) backend networks (electronic databases). The RFID labels can be distinguished based on their frequency of operation: (a) LF; (b) HF; (c) UHF; or (d) microwave, the latter category being considered to cover the frequency bands at 2.45 GHz and 5.8 GHz. Advantages and disadvantages of each of these bands are listed.

Labels are also categorised in terms of their powering techniques of: (a) passive; (b) semi-passive; or (c) active, and the general features and applications of each type are identified. In considering communication between labels and interrogators, it is noted that there are similarities and differences in the way communication is achieved in both the far and the near field by a label antenna, and it is also explained that the role of label quality factor changes significantly between the two situations.

The EPC concept is briefly described on account of its close relation to emerging applications, and a hierarchy of label functionalities is also introduced. A method, by which the dilemma of diverse functionalities may be resolved by means other than a rigid hierarchy of functionalities, is described.

In considering back end systems it is pointed our that the general design principle in EPC based RFID systems is to off load silicon complexity of the label to backend systems and to the reader in order that the cost of the labels may be kept to a minimum, but the discussion of such systems is left to a further Chapter 4.

The important aspect of anti-collision that arises in multiple label reading applications is considered and it is noted that as RFID labels are constrained by limited computational power, and memory, and the anti-collision algorithms embedded in multiple tag reading protocols take note of this, and that anti-collision methods used in RFID must consider the wireless and ad hoc nature of RFID networks along with the necessity to recover from sudden power loss in the almost invariably used passive RFID systems.

Among the anti-collision algorithms both deterministic and probabilistic schemes are recognised. In addition to features which reduce the frequency of collisions, the capacity to detect collisions is seen as a powerful addition to an anti-collision algorithm. The role of line coding schemes is analysed and those which may or may not detect invalid symbols cased by collisions of label reply signals of differing strengths is identified. The role of CRC schemes in detecting collisions is also discussed.

Also influencing the performance of tag reading protocols is the issue of tag confusion, under which tag receive conflicting command or response signals from more than one interrogator, and so-called ghost reads (a reader reporting an EPC

of a tag that does not exist in its tag reading range) can occur. The features of well designed protocols that reduce this phenomenon are described.

The chapter concludes with a summary of the issues covered and reminds readers that the following chapter will elaborate on the integration of backend systems to RFID technology, developed under the Auto-ID Center vision of a "Networked Physical World".

Chapter 4: "EPC Network Architecture"

In Chapter 4 "EPC Network Architecture" the authors provide an outline of the structure and usage of the ubiquitous item identification network that originated at the former Auto-ID Center, now called the Auto-ID Labs, and currently managed by a number of working groups at EPCglobal Inc. The Auto-ID Center vision was to create a "Smart World" by building an intelligent infrastructure linking objects, information, and people through a ubiquitous computer network, leveraging the Internet for global connectivity.

Contrary to the component based EPC Network architecture developed initially by the Auto-ID Center, the more modern version is based on an N-tier architecture with an emphasis on defining interfaces. The interfaces define the required standard functionalities and methods by which optional functionalities can be accessed rather than defining components and their associated functionalities.

The N-tier layered service oriented architecture approach fits naturally with an object oriented modelling of the architecture because objects encapsulate information and state while offering functionalities through their interfaces. The modules also have a loose coupling due to the independence of different modules. This reduction in dependency implies that the system is easier to manage and enhance.

Web services are one method of implementing the Service Oriented Architecture (SOA) over standardised protocols and interfaces. There is a strong tendency and a technological trend driving the EPC network architecture towards a web services based SOA.

The EPC Network can be separated into six primary modules, some physical, some logical: (1) RFID tags; (2) RFID tag readers; (3) EPC; (4) middleware; (5) Object Name Service (ONS); and (6) EPC Information Service (EPCIS).

Middleware system provides real time processing of RFID tag event data. Conceptually the middleware occupies the space between a Reader (or multiple Readers) and the application systems.

The middleware has several fundamental functions, some of which are: data filtering of received tag and sensor data; aggregation and counting of tag data; and accumulation of data over time periods.

The middleware possesses two primary interfaces that allow it to communicate with external systems: the Reader Interface and the Application Level Event Interface. The former provides an interface between the middleware and readers, and the latter between the middleware and external applications.

An middleware is composed of multiple Services, each with their own functionality. The services can be visualised as modules in the middleware. The multiple services modules can be combined to perform certain functions for specific applications. Hence one or more applications may make method calls to the middleware resulting in an operation being performed (e.g. collection and return of temperature readings from a sensor), and the return of results.

Event management is a primary service provided by the middleware services. A common event management function is filtering, which is particularly useful in situations where there is heavy data traffic.

However, recent developments have retreated from such a rigidly defined schema to the characterization of two instances: ECSpec and ECReports instances using a standard XML depiction. Thus requests to the middleware are sent as ECSpec object, while data from the middleware is returned as an ECReports object.

The core XML schemas for these objects are defined with extensions and rules to accommodate application or manufacturer specific XML schema (such as that suited for a specific sensor application) or a number of such schemas to allow the capture and reporting of physical world events and measurements.

The functionality provided by the ONS system is similar to the services provided by the Domain Name System (DNS); however instead of translating host names to their underlying IP addresses for user applications, ONS translates an EPC into URL(s). The Object Name Service (ONS) in an EPC Network identifies a list of service endpoints associated to the EPC and does not contain actual data related to an EPC. These service endpoints can then be accessed over a network. However, unlike the DNS, ONS is authoritative, that is the entity that retains control over the information about the EPC placed on the ONS is the same entity that assigned the EPC to the item.

In the event that the local ONS server is unable to satisfy the requests it is forwarded to a global ONS server infrastructure for resolution.

It should be noted here that the ONS does not resolve queries down to the level of fully serialised EPCs. The depth of the query stops at the Object Class level (product type) of the EPC.

A possible interface for an EPCIS can be implemented by adopting web services technology. A web services technology based interface allows applications in the wider area network to utilise services provided by local EPC Information Services using a remote method invocation paradigm. Such an architecture has the advantage of leveraging standardised XML messaging frameworks, such as that provided by the Simple Object Access Protocol (SOAP), and a description of the available services defined in terms of a Web Services Description Language (WSDL) file.

Hence an application requiring information is able to access a WSDL file which has a description of the available service methods, the required input and output parameters to the methods and information to invoke those methods.

EPCIS provides a model for the integration of RFID networks across the globe. However it is important that EPCIS provides a secure communication layer so that local EPC Networks can retain the authority to determine access to information. WS-Security is a candidate proposal for enhancing web services security that

describes enhancements to SOAP messaging to provide message integrity and message confidentiality while proposed architectural extensions to the existing WS-Security profile could provide access control as well as a federated security model for EPCIS.

As stated above, the ONS does not resolve to the serial number level of the EPC and the DNS technology upon which the ONS is based also does not allow the fine grain resolution down to serial number levels. Resolution down to serial EPC level (to a specific object) is handled by the EPCIS Discovery Service (EPCIS-DS).

EPCIS-DS is best described as a "search engine" for EPC related data. EPCIS-DS provides a method for custodians of a particular RFID tag data to update a register within the EPCIS-DS to indicate that that they are in possession of data related to an EPC.

The chapter also considers briefly supply chain management issues such as product recall, grey-market activity and counterfeiting, and describes the concept of the "electronic pedigree", a term that has been coined to label the electronic history of an item's life throughout the supply chain. However it is made clear that issues related to security are considered in more detail in later chapters of this book.

Chapter 5: "A Security Primer"

Finally, in this introductory section, there is provided in Chapter 5 "A Security Primer" an overview of state of the art cryptography that can be applied to communication over insecure channels. The chapter describes the range security objectives to be sought, the fundamental Kerckhoffs' Principle that must be observed in designing defences, the types of attack that can be mounted by persons of ill will against cryptosystems, and gives a classification of the security levels that can be attained. Unkeyed and keyed cryptographic primitives are defined, the latter including both public key and secret key systems, and their use both in securing messages against eavesdropping and in detecting that messages are authentic is explained. The burdens of providing the computational resources for the implementation of known effective schemes are discussed and found to be excessive the RFID context, and the chapter concludes with a statement that resource constraints in RFID tags have introduced a need for new lightweight cryptographic primitives to be used in RFID technology.

Section 2 Security and Privacy Current Status

This second section of the book contains four chapters that describe the current status of attempts to produce security and privacy. They begin in Chapter 6 with a more detailed treatment of security and privacy concepts than has been presented in the Primer.

Chapter 6: "Addressing Insecurities and Violations of Privacy"

In Chapter 6 "Addressing Insecurities and Violations of Privacy" the authors examine the vulnerabilities of current low cost RFID systems and explore the security and privacy threats posed as a result of those vulnerabilities, and the quality of defences that may be deployed.

The chapter formulates a framework for defining the problem space constructed around low cost RFID systems, and considers the challenges faced in engineering solutions to overcome the relative defencelessness of low cost implementations. Security issues beyond and including interrogators are not considered, as such concerns may be easily resolved using existing technology and knowledge. There is a concentration on the systems that are advocated by EPCglobal as Class I and Class II, both in respect of published standards at UHF and emerging draft standards at HF.

It is noted that for a low cost tag any additional hardware required to implement security needs to be designed and fabricated, this incurring additional cost. Reducing dice sizes to very small levels is not seen as feasible to compensate for such costs as the increase in cost of handling smaller die must be considered. A more practical avenue for reducing costs is seen as the use of obsolete IC manufacturing processes and filling up such fabrication pipelines with RFID IC chips.

It is concluded that as, due to cost constraints, low cost tags do not utilize anti-tampering technology, the long-term security of label contents cannot be guaranteed.

In emerging standards labels within reading range are reported as having a means of revealing their presence, but not their data, when interrogated by a reader. The labels then reply with a non-identifying signal to an interrogation by using a randomly generated number. However, for HF tags, there is no such prevalent standard, although EPCglobal is currently developing an HF specification to complement its UHF air interface protocol. The existing standards most commonly in use for HF tags, other than the ISO 18000, are listed.

Two important and related performance parameters are the number of label reads per second and data transmission speeds. Performance criteria of an RFID system demand a minimum label reading speed in excess of 200 labels per second.

As near and far fields scale differently with distance, each frequency band is seen to provide its own set of advantages and disadvantages.

Anonymity desired by persons is discussed. The most important concept is probably the concealment of the identity of a particular person involved in some process, such as the purchasing of an item, visit to a doctor or a cash transaction. Another is the concept of untraceability (location privacy).

There is a discussion of "killing" a label. Killing involves the destruction of the label thus rendering it inoperable. An alternative idea to killing that has been entertained involves the removal of the unique serial number of the EPC code in articles that allows the label owners to be tracked, albeit with difficulty in practice. This does not remove all the privacy concerns as tracking is still possible by associating a "constellation" of a label group with an individual. This implies that a particular taste in clothes and shoes may allow an individual's location privacy or anonymity to be violated. However "killing" a label will eliminate

privacy concerns and prevent access by unauthorized readers when combined with a password to control access to the kill command.

The security risks that arise with low cost RFID labels are seen as arising from: (a) communication between a tag and a reader taking place over an insecure channel; (b) tags being accessible by any reader implementing the air interface protocol; (c) tags being not tamper proof and allowing a channel for physical access to tag contents and circuitry (as a result, tags cannot be expected to secure information for long periods); (d) integrated circuit designs being constrained by cost and being thus minimalist implementations; (e) air interface protocols being designed to reduce tag complexity; and (f) design flaws in reader implementations due to cost constraints.

It is noted that transmissions from a reader and a tag take place over a clear communication channel which may be observed by a third party. In this context the classification of eavesdropping range concepts: (a) operating range; (b) backward channel eavesdropping range; (c) forward channel eavesdropping range; and (d) malicious scanning range is offered. Passive eavesdropping and scanning (active eavesdropping) concepts are discussed.

Attacks on security are classified as: (a) cloning; (b) man-in-the-middle; (c) denial of service; (d) and code injection. Communication layer weaknesses of: (a) physical attacks; (b) non-invasive attacks; (c) invasive attacks; (d) privacy violations; (e) profiling; and (f) tracking and surveillance concepts are also defined and explained.

In addressing vulnerabilities, sources of unreliability are identified as: (a) effects of metal and liquids; (b) effects of permeability of materials on tag antennas; (c) interference and noise from other users; (d) tag orientation; (e) reading distance; (f) Electromagnetic Compatibility (EMC) regulations; and (g) cost and power constrained implementations of tag chips.

In addressing security issues the list of security objectives identified and explained are: (a) confidentiality; (b) message content security; (c) authentication; (d) access control; (e) availability; and (f) integrity. Tag and interrogator authentication is addressed. Tag and product authentication issues are also discussed.

In the context of addressing violations of privacy, relevant concepts are elaborated as: (a) privacy of personal behaviour; and (b) privacy of personal data. The number of privacy violations RFID technology can potentially cause are said to be numerous, so the reader is referred to specific literature. Significant issues that must be dealt with by policy formulation or amendment in relation to RFID practice are stated as those generated by the following items: (a) unique identification of all label items; (b) collection of information; (c) dissemination of that information; and (d) mass utilization of RFID technology.

Achieving the privacy objectives discussed so far is seen as to be sought by the deployment of cryptography, so discussion of how those objectives may be achieved begins with a discussion of cryptographic tools. Concepts identified and elaborated into subcategories are (a) primitives without keys; (b) symmetric key primitives; and (c) asymmetric key primitives.

Attacks on cryptographic primitives are classified as: (a) ciphertext-only attacks; (b) known plaintext attacks; (c) chosen plaintext attacks; (d) adaptive chosen ciphertext attacks; and (e) adaptive chosen ciphertext attacks. Attacks on protocols are classified as: (a) replay attacks; (b) known key attacks; (c) im-personation attacks; and (d) dictionary attacks.

In considering levels of security the levels are defined as: (a) unconditional security; (b) computational security; (c) ad-hoc security; and (d) provable security. An explanation is given for each.

The chapter then turns to a consideration of low cost RFID cryptography for which the challenges are defined as: (a) cost; (b) regulations; (c) power consumption; (d) performance; and (e) power disruptions. The impact of all these challenges is discussed.

This chapter also provides a survey of appropriate solutions. These include: (a) use of cryptographic hash functions; (b) use of linear and non linear feedback shift registers; (c) the NTRU cipher; (d) the Tiny Encryption Algorithm (TEA); (e) the Scalable Encryption Algorithm (SEA); and (f) an unorthodox re-encryption mechanism for securing a banknote, employing on the label a cipher text and a random number and on the banknote a serial number and a digital signature.

Lightweight cryptography and lightweight protocols then receive consideration, this material leading to a discussion of minimalist cryptography. Concepts identified and explained are (a) pseudonyms; (b) one time pads and random numbers; (c) exploiting noise; (d) distance implied distrust; and (e) authentication protocols and particularly the YA-TRAP protocol, in various versions, that provides location privacy and allows the authentication of the tag by using monotonically increasing timestamps stored on the tag which are in synchronicity with timestamps on a secure backend database.

The chapter concludes with the notion that security comes in many flavours and strengths, but that low cost implies that we find mechanisms that are generally "good enough" as deterrents rather that mechanisms that are impossible to crack. However, the use of one time codes does allow a great strength over a limited number of reader and tag authentications.

Chapter 7: "Security Vulnerabilities in RFID Systems"

Then Chapter 7 "Security Vulnerabilities in RFID Systems" outlines a weakness, known as an SQL attack that is present in some simple software systems. It is shown is that this weakness can fortunately be avoided by good software design, and generally now is. Other forms of attack, such as engineering buffer overflow are also studied, but it is shown is that the architecture commonly adopted for an RFID reader will provide protection against such attacks. Finally, various denial of service attacks, that may be mounted through the introduction of broadband noise or unauthorised transmissions, are considered and are warned against. For such attacks the appropriate defence appears to be the identification of their sources and their silencing through the deployment of appropriate legal means.

Chapter 8: "An Evaluation Framework"

In Chapter 8 "An Evaluation Framework" there is presented a framework against which the success of attempts to provide security is later to be evaluated. The problem space is constructed around low cost RFID systems, so as to enable the engineering of solutions to overcome the defencelessness of low cost RFID systems and to be able to evaluate those solutions for their effectiveness. The chapter develops simple evaluation criteria for security mechanisms and a simple, yet sufficient model of a low cost RFID system for analysing security mechanisms.

The chapter provides an outline of low cost RFID system characteristics according to: (a) class; (b) length of unique identifier; (c) read range; (d) read speed; (e) hardware cost; and (f) power consumption. The chapter summarises the important aspects of low cost RFID, that need to be understood and provides reasonable assumptions that need to be made prior to implementing any cryptosystems to address the vulnerabilities implied by the non-achievement of defined security objectives.

There is provided a security evaluation matrix to appraise the suitability of various mechanisms for providing security and privacy to low cost RFID and various applications constructed around low cost RFID.

In the matrix there are to be achieved security objectives of: confidentiality; message content security; tag authentication; reader authentication; product authentication; access control; availability; and integrity. In the matrix there are also to be achieved privacy objectives of: confidentiality; message content anonymity and untraceability. In the matrix there are also cost and performance estimates of: tag implementation cost; back end resource requirements (on line or off line); overhead costs (initialisation cost or time); time estimates (time to complete a process or clock cycles); and estimated power consumption.

Criteria for evaluating security mechanisms and hardware costs are alo given. In estimating hardware costs it is common to express the area evaluation in terms of the number of gates (NAND) required. Implementing a NAND gate in hardware requires at least four FETs. Typical cost estimations in terms of the gate count are given for various cryptographic hardware elements.

It is recognised that it is difficult to implement a security mechanism without the aid of proxy systems or a secure backend system for storing secret information such as keys. Security mechanisms of this kind are recognised as requiring online and real time access to secure resources. The monetary and time cost of implementing such mechanisms is considered in the evaluation process. Observing constraints placed on RFID security mechanisms may require expensive database system implementations and expensive networking infrastructure. Backend resource costs are expressed as those requiring online access or those that can be performed off-line.

Overhead costs may result from the need for initializing tags with secure information, or the need for performing some operations prior to their use, or periodically during their use. For instance a security mechanism may require the replenishment of secret keys on a tag.

Under power consumption costs it is recognised that any security mechanism design will eventually involve an IC implementation. Currently, static CMOS is the choice of most digital circuit designs built for low power consumption and robustness.

An important aspect of the design process and the establishment of its suitability is to ensure that the power dissipation of the integrated circuits do not exceed that outlined in the consideration of low cost RFID system characteristics.

Chapter 9: "From Identification to Authentication"

Finally in this group of papers there is presented in Chapter 9 "From Identification to Authentication" a description of how RFID can be used for product authentication in supply chain operations. A review of existing approaches is provided. These approaches are analysed in the context of anti-counterfeiting needs, and fields where future research is needed are identified. It is pointed our that the effort that an illicit actor has to undertake to break or by pass the security mechanisms implemented has a major impact on the cost of product authentication system.

The general requirements of authentication systems in supply chain applications are identified. They include that: the system needs to be used by multiple parties from multiple locations; authentication of products that are unknown to the system should be supported; the cost and effort to perform a check need to be low; and the optimal solution should allow also the customers to authenticate products.

The general attack scenarios of illicit players are described, and range from: taking no explicit action, but relying instead on consumer demand for counterfeits; through the use of misleading bogus security features that are designed to deter closer inspection; through also the removal of authentic security features for genuine products and re-applying them to fake products; to the cloning and imitation of security features.

RFID product authentication techniques are discussed in detail. Particularly promising are the methods that use the unique factory programmed chip serial number (TID) of EPC Class-1 generation-2 tags. However, it is shown that such schemes are not proof against attackers who have access to hardware manufacturing. Tags with cryptographically protected secrets, particularly where the secrets are shared within groups of tags, are vulnerable to those secrets being stolen and sold out by insiders.

There is also discussion of other forms of attack such as denial of service attack of the types discussed in earlier chapters, but such attacks are not considered as realistic threats against RFID product authentication which is mostly performed under the surveillance of authorised personnel or by the customer.

The chapter contains and extensive review of product authentication approaches and their advantages and disadvantages, a section deducing tag requirements for authentication, and concludes that the role of standards is of primary importance in product authentication and should be taken into account in solution design.

This chapter is notable in that it contains 67 references and, in an Appendix, a comprehensive table summarising the requirements of different product authentication approaches.

Section 3 Network Based Solutions

Section 3 contains three chapters that outline solutions to the authentication problem by exploiting characteristics that can be introduced into the communication network.

Chapter 10: "EPC System for Safe and Secure Supply Chain and How it is Applied"

The material in Chapter 10 "EPC System for Safe and Secure Supply Chain and How it is Applied", while being drawn from Japanese experience, can be considered to be applicable everywhere. The overall aim of the chapter is to explain how EPC systems improve safety and security.

The chapter begins with gray markets and black markets being defined, paths into an out of legitimate market being identified, and short term issues (expired products, wrong handling of products) and long term issues (product recall arising from later discovery of defects) issues being described.

Five stages of the supply chain from manufacturer, through wholesaler, repackager, and retailer to the consumer are defined.

The view of the Chapter is summarised in six tables all well supported by text argument.

Tables 1, 3, and 4 all consider threats classified as: fake label; adulteration; relabelling; substitution; fake product; stolen; gray market; scrapped; and recall; and these nine items are grouped into the three classes of: counterfeit; illegal trade; and wrong status.

The six tables describe in order: threats and entry points; basic applications (one physical and three informational) for securing a supply chain; threats and entry points, now revised to exclude out of scope items such as fake labels or adulteration by the manufacturer; measures that may be deployed to secure the supply chain, grouped as to whether they are covered by the EPC system or not; mapping of security measures to EPC systems components such as EPC, Tag, Reader, Middleware, EPC-IS or ONS; and network availability influence of security measures.

The Chapter considers that EPC components being standardised currently may not be sufficient to realise all the security measures required. Potential research topics arising from that fact include: ID encryption; access control to the tag; management of exposure of tag identifiers; electronic document validation (not yet sufficiently pervasive); business processes to manage product status beyond EPC-IS; need for ONS security; and need for a tamper evident tags.

Chapter 11: "The Potential of RFID and NFC in Anti-Counterfeiting"

In Chapter 11 "The Potential of RFID and NFC in Anti-Counterfeiting", the authors investigate how RFID and Near Field Communication (NFC) could improve current customs processes to fight illicit trade.

In current import processes, customs officers have to evaluate which consignments are inspected and, when an inspection takes place, whether intellectual property rights have been infringed. The authors propose and evaluate new micro processes that leverage the dual-existence of products and logistic units in order to enable easier, faster and more reliable inspection of goods.

The significance of the work rests on the fact that the majority of counterfeit products in the Western countries are imports and the most important means of transport of counterfeit products is by sea.

Customs are responsible for about 70% of all seizures of counterfeit products in the world [2]. The role of customs is especially important in protecting the European Union and the U.S. because the vast majority of counterfeit products in those markets are imports and, after entering the market, subject to free circulation within the community.

Customs authorities fail to seize large amounts of counterfeits either because they do not know how to recognize the fakes or because the process of gathering statements from trademark owners is too time-consuming.

While controlling the trade, however, customs also work to facilitate the trade and seek not to disturb import and export. These two objectives conflict, and thus customs always have to balance between control and facilitation. Given also that the vast majority of goods that pass through customs are legal and should not be disturbed, finding counterfeit goods is not among customs' top priorities.

Customs use RFID also to strengthen the security of consignments. To guarantee the integrity of cargo, shippers install electronic seals, or *e-seals*, into their containers. The role of RFID in the e-container is to provide connectivity and real-time telemetry.

One consequence of this trend is the emerging of *green lane* programs where shipping companies gain lighter inspections when they conform to certain additional regulations, such as in the Smart & Secure Tradelanes (SST) initiative or the Customs-Trade Partnership Against Terrorism (C-TPAT).

Customs conduct risk analysis to identify high-risk consignments in pre-hand. Regarding counterfeiting, the country of origin is the most important criteria in the risk-analysis and, consequently, it is often attempted to be disguised by the carriers of counterfeit goods. Careful selection of inspected containers can provably provide considerable improvements in the detection rates of counterfeit products.

The authors propose the use of Near Field Communication devices, and in particular RFID tags operating at 13.56 MHz. The devices apply touch to read principle which makes communication easy and intuitive, and the typical reading ranges vary from 0 to 20 cm. Besides reading NFC tags, the protocol allows for secure two-way communication between the reader devices. This differentiates

NFC from RFID technology used in supply chain applications, where the goal is mostly to read multiple tags at once without line of sight.

The authors propose new micro processes that can be used to improve the existing customs import process to find and seize more counterfeit goods. The enabling technology of the proposed processes is any hand-held NFC device with a network connection, such as already available NFC mobile phone. This device allows the customs officers to read tagged items in their field work. It is taken into account that in a modern customs process, the flow of information and the flow of goods are separated and therefore the customs officers need to move to the warehouse to conduct the physical inspections. In a very lean and automated import process, the time that the products spend in the customs warehouse can be very small and measured in tens of minutes, which can set rigid time-constraints for the inspections.

The process steps are the following: (i) identify the product by reading the tag; (ii) obtain the network address of the authentication server using a network address resolution mechanism (e.g., Object Naming Service); (iii) establish a secure connection with the authorized server (e.g. EPC PAS); (iv) establish which authentication protocol, if any, the tag supports; (v) automatically authenticate the product (tag) using the supported protocol; and (vi) verify the tag-product integrity.

It should be kept in mind that usually it is actually the tag that is authenticated and not the product itself. Therefore verification is required to make sure that the authenticated identity really matches the physical product (step vi). Omitting this verification makes the system vulnerable to simple attacks where fake goods are equipped with any authentic tags.

Though RFID is already used in customs logistics in different ways today, it still has unused potential to help customs in the fight against illicit trade. In this Chapter, the authors have presented how, together with NFC enabled mobile reader devices, RFID enables product authentication applications that make inspection of tagged cargo faster and more reliable.

Chapter 12: "Improving the Safety and Security of the Pharmaceutical Supply Chain"

Chapter 12 "Improving the Safety and Security of the Pharmaceutical Supply Chain" discusses various techniques that may be used to combat counterfeiting in the pharmaceutical supply chain. These include the use of electronic pedigrees (to ensure the integrity of the supply chain), together with mass-serialization (to provide for a unique lifecycle history of each individual package) and authentication of the product (to check for any discrepancies in the various attributes of the product and its packaging are as intended for that individual package). Management of the pedigree process and product authentication is discussed in some detail, together with various other learnings from the Drug Security Network, including identification of some remaining vulnerabilities and suggestions for tightening these loopholes.

The Drug Security Network (DSN) was formed as a forum for a number of major players in the pharmaceutical industry to consider the major changes and challenges to business practices which will result from the enforcement of pedigree legislation and introduction of mass-serialization, which are being introduced imminently in order to make the pharmaceutical supply chain safer and more secure.

The paper discusses in turn the primary deliverables (three papers) of the DSN activities.

The purpose of a pedigree is stated as providing legal proof of a secure chain of custody from the originator of the pharmaceutical package (usually the manufacturer or wholesaler) through to the organization that sells or dispenses the pharmaceuticals.

Three key issues needing to be considered are: Pedigree Data Content/Format; Pedigree Processing; and Pedigree Transmission Mechanism.

A number of key requirements are identified for a standardized format for electronic pedigrees. These are: completeness; global scope; suitability for legal or government audit; and integrity, authentication and non-repudiation.

A number of key requirements are identified for the transmission mechanism for electronic pedigrees. These are: timely access to data for verification and certification processes; robust access to data for verification and certification processes; authentication, integrity and non-repudiation; and suitability for legal/government audit.

The Propagating Document Approach and the Fragmented Data Approach are identified with the former being the most favoured.

In that approach, each subsequent custodian verifies the signed content of previous custodians, then amends and re-signs the data, before transmitting the pedigree to the next custodian when the goods are shipped onwards. As the pedigree document moves across the supply chain, additional outer layers are added. This approach offers a double-linked chain of security, since each custodian can verify all the inner layers of the pedigree document, then signs to confirm that they have done so (the reverse link). At the time of shipping, they then add additional data about the next recipient and sign this (the forward link).

It is pointed out that a pedigree document primarily records a chain of transactions. It does not warrant that the package itself is the genuine product. For this, authentication is required. Two kinds of authentication are discussed: authentication of the identity, since the identity provides the 1–1 link to the pedigree data; and authentication of the product itself, in case the identity of the package has been copied or the details about the product have been falsified.

It is explained that a key feature of the Safe and Secure Supply Chain is the emphasis on authenticating the object, as well as the pedigree trail. A networked information system, such as one complying with the future EPC Information Services (EPCIS) standard, would provide a mechanism for a manufacturer or labeller (or other authoritative party) to be able to validate a number of properties specific to a particular serial number. These might include an independent hard-coded read-only tag ID, the product class and/or details of customized security features, either covert or overt.

It is further explained that when validating the authenticity of the product, it may be necessary to check the following criteria: authenticity of the tag; authenticity of the pedigree ID; authenticity of the serialized identifier; authenticity of the product's packaging; checking the current state; and authenticating the trail.

Three groups of use cases are considered.

In the discussion on security of business documents in general, the following five key security requirements are identified: authentication; authorization; confidentiality; integrity; and non-repudiation

The concept of a Pedigree Business Document is introduced, and the risks of paper pedigree are considered in some detail. The paper identifies a number of potential loopholes of paper-based pedigree documents. These include that: a fraudulent wholesaler can sell counterfeit items with legitimate paper-based Pedigree documents; and a fraudulent wholesaler may forge paper-based Pedigree documents and sell counterfeit items saying they are returns from the retailer. Thus it is explained that using paper-based Pedigree documents increases the risks of entry of counterfeit drugs.

Cross-border shipments and diversion are also discussed. Vulnerabilities in the form of potential loopholes in the security of proposed pedigree legislation are discussed, and the need for certification authorities is also established. Enforcing a change of serial ID and labeller code on repackaging is seen as essential.

Section 4 Cryptographic Solutions

Section 4 consists of six chapters that describe solutions to the provision of authentication services by exploiting cryptographic concepts that may be introduced within RFID labels.

Chapter 13: "Product Specific Security Features Based on RFID Technology"

In this chapter, the authors propose a security solution based on Radio Frequency Identification (RFID) technology, using low-cost transponders that contain item-specific information to avert removal-reapplication attacks. The proposed solution aims at providing unique and secure authentication.

The approach utilizes RFID technology in which transponders hold unique and cryptographically secured data that uniquely binds a given instance of product to a given tag, and thus makes duplication or re-application of tags difficult.

A solution based on signed product characteristics is proposed. The main components of the architecture are an RFID tag containing product specific validation data introduced by a branding machine, explained below, and a product verifier containing an RFID reader, a crypto engine and a communications interface to a key data base.

The system allows setting up a secure and authentic binding between a product and a passive RFID tag residing on that product.

The Branding Machine is mainly responsible for computing and writing of the unique and secure product validation data to the tag. The component called product verifier is able to determine whether the product validation data delivered by an RFID tag is authentic and thus indicates the tagged product's authenticity. The product verifier will have the modules RFID reader, a crypto engine, and a communication interface.

In operation, the RFID reader component requests the RFID tag for the product validation data stored on that tag. The crypto engine is responsible for checking the authenticity of the product validation data read by the RFID reader and also for determining whether the product validation data has been altered by an impostor which event would be an indicator for a faked product. The communication interface can be used to determine authentic cryptographic keys from the (optional) component called key database. The usage of the key database can be eliminated by storing known verification keys either on tags or on product verifiers.

The unique product identifier contains, as well as cryptographic parameters, a bit sequence that uniquely characterizes the given product. Typically, this information is determined by the product's vendor. Depending on the specific type of product, different physical, chemical, etc. properties that can be verified, i.e. detected or measured, by a (human or machine) observer. Example properties that – either altogether or in a subset – can uniquely characterize a product with a certain high probability are weight, electric resistance, geometry, or a serial number printed on the product itself or its packaging, etc. This data will typically be written on the tag by the product's vendor before product delivery, for example during packaging. It is also possible to place a reference here, such as an URI that specifies a dataset stored on a remote database. This may help to save tag resources, but will make product validation dependent on the availability of that external storage.

In summary in this Chapter, the authors propose an anti-counterfeiting security solution based on RFID and EPC technology, which is applicable for passive, hopefully low-cost transponders. The exceptional feature of the approach is that the tags contain verifiable, item-specific information. Thus, a tag which is applied to a product is tightly bonded to that item, providing a measure to avert cloning attacks. The solution is also adaptable for offline checks if no network connection is available. However, the applicability of the proposed solution depends very much on the availability of unique, product specific properties which are easy to observe.

Chapter 14: "Strengthening the Security of Machine-Readable Documents"

Chapter 14 "Strengthening the Security of Machine-Readable Documents" considers the on-going trend towards turning paper documents that store personal information or other valuable data into machine-readable form, an example being the electronic passport that will become common in the near future.

The Chapter shows how the security of these machine readable documents could be improved by combining RFID with optical memory devices, integrating an optical memory device into the RFID enabled smart document to produce methods by means of which these two storage media can be combined to secure the document against threats like illicit scanning, eavesdropping and forgery.

The approaches described make use of the optical document-to-reader channel which is considered to be more secure than the radio-frequency communication interface. They are relevant to numerous applications where tagging physical documents would be interesting. Besides e-passports and other travel documents, customs freight papers, security papers (e.g. gift certificates, jewellery appraisals), driver's licenses and vehicle registration papers that would benefit from being machine readable through radio-frequency (RF) communication provide examples. A common factor of these documents is that they all relate to a physical entity that is not very well suited to being tagged to become a data carrier itself.

Four different approaches, defining how this combination of an optical and rf channel could be used to overcome existing security threats of machine readable documents, in terms of more secure communication protocols and resistance against forgery and cloning, are described. Instead of establishing security based on sharing secrets between the reader device and document before the communication, the approaches make use of optical memory devices which cannot be read or eavesdropped without a line of sight.

Machine readable documents are defined and discussed. All physical documents that carry a digital memory device are considered as machine readable documents. The typical instance of these kinds of documents is an RFID tagged paper, but another way to make documents machine readable is to use optical character recognition (OCR) to read data printed on the document. Machine readable travel documents (MRTD) comprise e-passports, visas and special purpose ID/border-crossing cards. Because of their similar nature, driver's licenses are also included within this group.

The benefits of having RFID transponders in physical documents come from the simple and fast read process that does not demand a line of sight connection. Depending on the grade and price of the chip, the contactless memory device can also contain support for re-writable memory and logical functions like cryptographic primitives. Therefore machine readable documents can provide high level of security and counterfeit resistance.

The significant components of an RFID enabled machine readable document application are the document itself, the reader device and the reader's control and crypto unit. Typically the transponder stores at least a unique identifier (UID) number. In addition, the transponder can provide logical functionalities like access control (through key comparison), random number generation and data encryption. Thus, the transponder serves as more than a mere barcode label.

Without specific encrypted addressing, the RFID air interface is not secure and the transponder is vulnerable to clandestine scanning (or *skimming*) and eavesdropping.

The reader device is responsible for the wireless communication. It is connected to the control and crypto unit through a closed, secure channel.

The role of the digital memory devices in the authentication processes of travel documents is twofold: on the one hand they help authenticating the traveller and on the other hand they help proving the authenticity of the document itself.

E-passport design has to address needs for individual privacy and national security and thus it poses severe security and privacy requirements. First of all, the integrity and authenticity of the data the passport stores has to be guaranteed. Second, the data has to be kept confidential from non-authorized parties. Third, the passport must not pose privacy threats for its carrier and, furthermore, all these have to be fulfilled in a public system during up to 10 year long life-span of the passport.

The authors describe the studies of Juels et al. who have discussed the security issues of e-passports and in which the following four threats, among others, were brought into light: (a) clandestine scanning; (b) clandestine tracking; (c) eavesdropping; and (d) cryptographic weaknesses. Moreover, those authors concluded that the e-passports do not provide sufficient protection for their biometric data.

The last of these concerns is relevant regarding forgery because without this connection, the system can be fooled, for example by putting a valid transponder into a fake paper.

Scarce resources on the chip limit the use of cryptographic primitives and the goal of the design is often low-cost low-security features.

The chapter then considers how optical memory devices can be combined with RFID to overcome some of the security threats of machine readable documents. The addressed security issues comprise: (a) no connection between chip and paper; (b) data integrity; (c) clandestine scanning; (d) clandestine tracking; and (e) eavesdropping. The first two issues are seen to relate to the security of the overall system and the latter three to the unsecured wireless communication.

Integrating an optical memory device into the RFID enabled machine readable document is proposed. What is common to all optical memory devices is that they need a line of sight connection for reading, making them resistant against clandestine reading and eavesdropping. Therefore it can be assumed that this channel is secure. The addition of an optical memory device extends the communication channel between machine readable document and a reader device to provide the combination of an insecure (RFID) channel and a secure (optical) channel.

Four approaches showing how the combination of RFID and optical memory devices can be used to increase the security of machine readable documents are described. The first two approaches address data integrity and bind the chip and the document, while the two other approaches aim at securing the communication.

Because machine readable documents often relate to a physical entity, it is assumed that data of interest that the document stores relates to this entity. That data is denoted as *object specific data* and it can be used for example in authentication. In addition, the documents can store any other application specific data which is merely referred to as *other data*. This other data can be static or dynamic. The way in which data can be used are: (a) storing object specific data in the optical memory; (b) storing a hash of object specific data in optical memory; (c) storing access keys in the optical memory; or (d) storing session keys in optical memory.

In the first approach static object data is stored in both the RFID transponder and the optical memory. This mirroring of the data increases the reliability of the overall document and provides a mechanism to tell if one or other device has been

tampered with. However, the optical memory does not provide access control and is vulnerable to skimming if the document falls into the wrong hands.

In the second approach the optical memory device stores only a hash value of the data. The used hash function needs to be known by the party performing the data integrity check so the specification of the hash function is stored on the chip. A smaller optical storage is needed.

In the third approach the RFID transponder does not reveal the object specific data if no correct access key has been transmitted in advance, which approach prevents clandestine reading. The access key is stored on the optical device and can only be read with line of sight connection. The means that, proved the document is safeguarded by its owner, the document owner can control who has access to the contactless memory

In this approach the RFID reader initiates a reading session by asking the transponder for an index i between 1 and N, where N indicates the number of access keys stored in the document. Because single access keys can be still obtained by eavesdropping the radio channel between the reader and the transponder, the number of access keys N needs to be large enough to make the malicious use of compromised keys infeasible and spoofing of access keys difficult.

In the fourth approach the transponder challenges the reader for a response to be read from the optical device. A challenge response operation is initiated by the tag transmitting a pseudo random challenge and an integer in the range 1 to N, where N is the number of session keys held in the optical memory device. To authenticate itself to the transponder, the reader device sends a response which is the challenge encrypted by the session key obtained by reading the optical device. The session key is also used to encrypt the following wireless communication of object specific data.

Most importantly, the session key is never transmitted in the insecure radio channel as this key is only optically accessible, which overcomes the weakness of an approach that may involve compromised access keys. On the other hand, this approach requires the transponder to support data encryption.

The authors see the main benefits of combining RFID with optical memory devices as lying in the field of document security. The use of two memory devices adds complexity to the system and thus makes the documents harder to be cloned or forged. Even though this conflicts the fundamental security doctrine of Kerckhoffs which says that the security of a system should depend on its key, not on its design obscurity, it is believed that it can provide effective ways to combat counterfeits.

Chapter 15: "Enhancing Security of EPCglobal Gen-2 RFID against Traceability and Cloning"

In Chapter 15 "Enhancing Security of EPCglobal Gen-2 RFID against Traceability and Cloning" the authors present a synchronization-based communication protocol for EPCglobal Class-1 Gen-2 RFID devices. The proposed protocol is secure in a sense that it prevents the cloned tags and malicious readers respectively from impersonating and abusing legitimate tags. In addition, the protocol provides that

each RFID tag emits a different bit string (pseudonym) or meta-ID when receiving each and every reader's query. Therefore, it makes tracking activities and personal preferences of tag's owner impractical and thus ensures the user's privacy.

As background, the authors observe that despite many prospective applications, RFID technology poses several security and privacy threats which could harm its global adoption. Ironically, the security weakness of RFID technology comes from the most basic operation of an RFID tag, that is to release a unique and static bit string known as the Electronic Product Code (EPC) identifying the object associated with the tag upon receiving a query request from a reader. Using the unique EPC as a reference, someone (equipped with a compatible reader) can track the moving history, the personal preferences and the belongings of a tag's holder. Even worse, absence of secure authentication results in revealing the EPC to malicious readers under a skimming attack. Once capturing an EPC, an attacker can duplicate genuine tags and use the cloned tags for its malicious purposes. A natural solution to the aforementioned security problems is to employ cryptographic protocol in the RFID system. Unfortunately, the cost of manufacturing a tag has to be extremely low, e.g., less than 30 cents (according to RFID journal, one RFID tag is expected to cost 5 cents by 2007). Therefore, the computationally intensive security protocols widely known in cryptographic literature cannot be incorporated into a small chip with tightly constrained computational power (at least in the foreseeable future).

The authors point out that the latest RFID standard ratified by EPCglobal is named EPCglobal Class-1 Gen-2 RFID specification version 1.09 (Gen-2 RFID for short). with the properties: (a) tags are passive; (b) communication is in the UHF band with range up to 10 m; (c) tags support on-chip PRNG and CRC computation; (d) privacy protection mechanism is to make the tag permanently unusable once it receives the kill command with a valid 32-bit kill PIN; and (e) read/write to memory is allowed only after it is in secure mode after receiving access command with a valid 32-bit access PIN

The authors argue that that privacy protection mechanism suggested in the specification is inappropriate, and many scenarios the tag should never be killed. Therefore, in designing a new protocol, they have avoided this kind of mechanism and, make a new use for the *kill* PIN.

For the *access* PIN, the authors point out that it is ineffective from a security point of view since the 16 bits of PIN is XORed with a 16-bit pseudo-random number sent by the tag in a session. Just by eavesdropping the 16-bit pseudo-random number and the XORed PIN, an attacker can easily recover the access PIN. Losing the access PIN is very dangerous because it allows a malicious reader to read/write the entire memory of a tag.

Although researchers have proposed protocols making use of a hash function, such protocols are seen as still beyond current capability of low-cost RFID tag. Thus, the authors have sought another solution which should use only the available functionalities of current RFID standards.

Although Juels has suggested such a scheme to prevent the legitimate tags from being cloned, it is pointed out that his protocol does not take eavesdropping and privacy issues into consideration, and thus provides no protection against privacy invasion and secret information leakage. In their paper, the authors present another scheme targeting most of security features for a RFID system including authentication,

traffic encryption, and privacy protection. An important feature is that the only trusted party in the proposed RFID system is the backend server and all the secrets are kept only at the tags and the backend server's database. In addition, the RFID reader is not be able to learn any secret information including PIN and EPC itself from data called *meta-ID* sent by a tag.

In the proposed protocol the meta-ID is forwarded to the backend server and the backend server can retrieve detail object information keyed by that meta-ID. The advantage of this approach are described as being: (a) the approach enables easy accountability and access control; and (b) instead of reader to tag authentication, the protocol requires the reader to authenticate to the backend server before sending a meta-ID.

There is a discussion of random numbers and their generation leading to the conclusion that the Gen-2 standard should support 32-bit PRNG to take full advantage of 32-bit PIN currently supported by Gen-2 specification. In the proposed protocol, use is also made of checksum code to provide security and resolve possible collisions at the backend server's database. To avoid a problem with check sums on all zero strings, a bit string is required to start with a bit 1.

In explaining their main ideas, the authors describe first protecting data transmitted between the tag and reader against eavesdropping. The obvious way to them is to utilize encryption/decryption and the most simple encryption function useable is XORing. The problem now turns to the key management issue: that is to ensure that a new encryption key is used in every session. Solving this issue turns out to be a solution to privacy protection as well since RFID tag can XOR EPC with different key in every session, thus, preventing malicious readers from tracking the tag. They suggest that the simplest, yet most efficient way of key sharing in this scenario is to use the same PRNG with the same seed at both RFID tag side and backend server. The session key can be computed by generating a new pseudo-random number from a current session key after every session. Importantly, this computation is required to be done at both RFID tag and reader/backend server in a synchronous way. Otherwise, subsequent traffic cannot be understood by both sides.

The next security problem discussed is the need for authentication. It is argued that in most cases, a reader just needs to know EPC stored in a tag and then eventually contact the backend server to get/update information about the object carrying the tag. Keeping this in mind, it is proposed that reader-to-tag authentication can be delegated to tag-to-backend server authentication. More specifically, a reader can only receive an EPC from an RFID tag in an encrypted form. It needs to authenticate itself to the backend server first, and then, depending on its privileges, that backend server can decide what kind of information to send back to reader (for example, in case of a public reader, only information describing what the referenced object is; and in case of a manufacturer's reader, the actual EPC and PIN associated with that tag can be sent). Actual reader-to-tag authentication needs to be carried out when reader wants to access (read/write) other sections of tag's memory bank. To do so, it is stated that the protocol can use a PIN-based approach just like in the original Gen-2 RFID specification.

A detailed discussion of the protocol and its sub protocols is provided. A possible synchronization issue with the protocol is recognised in that a false

'End Session' message might be sent by a malicious reader, and a method to prevent such interference is proposed.

In summary, the authors have presented a simple communication protocol for RFID devices, especially EPCglobal Class-1 Gen-2 RFID devices. The proposed protocol achieves desirable security features of a RFID system including: implicit reader-to-tag authentication, explicit tag-to-reader authentication, traffic encryption and privacy protection against tracking. The scheme makes use of only PRNG and CRC which are all ratified in the current Gen-2 RFID specification. It is claimed that there should be little overhead in adapting the proposed protocol into the Gen-2 RFID specification.

Chapter 16: "A Random Number Generator for Application in RFID tags"

In Chapter 16 "A Random Number Generator for Application in RFID tags" the authors observe that in current RFID technologies, pseudo random number generators (PRNG) serve as random number sources, but claim that their output numbers can show poor statistical properties, and that less than properly random secret keys reduce the security of data transmission. The objective is to show that an oscillator-based Truly Random Number Generator (TRNG) provides a better solution.

The TRNG exploits thermal noise of two resistors to modulate the edge of a sampling clock. The white noise based cryptographic keys prevent potential attackers from performing any effective prediction about the generator's output even if the design is well-known. The objective of this chapter is to discuss how to realize a TRNG in an RFID tag system.

Due to the confidential nature of most cryptographic systems, relatively few hardware RNG designs have been published. Designs available in the literature reveal three different IC-compatible methods for producing truly random sequences: (a) direct amplification; (b) oscillator sampling; and (c) discrete-time chaos. In the work reported here, several TRNGs were modelled, analysed and compared with the method of direct amplification. The results show that the oscillator-based TRNG is almost free from $1/f$ noise and periodic influences of substrate and power supply. These advantages make the oscillator-based TRNG a desirable solution for RFID tag application.

In an oscillator-based TRNG, a jittered low-frequency clock is used to sample a high-frequency clock. Two resistors' thermal noise is amplified to dither the edges of the low-frequency clock. The high-frequency clock is derived from the tag's analog front-end. According to EPCglobal RFID Class-1 Generation-2 Protocol, it should be n times of 1.28 MHz, where n is an integer.

An operational amplifier is used as a noise amplifier. The noise of a noise resistor is amplified by the operational amplifier and added to a triangular wave. The amplified noise output follows a Gaussian probability density and of variance proportianal to the value of its resistance.

The Chapter contains extensive analysis of circuits and probabilities of output values. In designing the oscillator-based TRNG, a main factor which influences the

statistical properties of the output sequence is the sample rate. In the finite bandwidth system proposed, the highest sample rate is limited by the noise amplifier.

In order to eliminate the correlation between the 16 bits of the output random number, it is shown that the sample rate must be roughly less than 1.5 times the bandwidth of the noise amplifier. Because the bandwidth of an operational amplifier is a function of power consumption, the highest sample rate is actually limited by the total power available.

The lower limit of sample rate is determined by the period of tag to reader communication cycle. The TRNG must provide 16 bits of truly random number within this period of time. According to EPCglobal RFID Class-1 Generation-2 Protocol, the minimum time of this period equals to 465 microsecond.

To implement TRNG in the RFID tag, there exist two main constraints: power consumption and chip area. In the proposed scheme, the most important factor is to keep the total power consumption of the low-frequency oscillator at around 1 microwatt. With the state of art, at a power supply voltage to 0.8 V, the total current consumption should be no more than 1.3 microamp.

For low power consideration, the output white noise needs to be as small as possible. In respect that the oscillator-based TRNG shows good quality against $1/f$ noise and some periodic influences, the lower limit of the output white noise is the resolution of the hysteresis comparator. Therefore, it is recommended that the noise magnitude be higher than 3 mV. The difference between the threshold voltages of the hysteresis comparator should be big enough to overcome the input offset, but it may not be too big as to increase the power consumption. Here, the value of 50 mV is chosen.

In exploring trade-offs of power consumption and chip area some new structures of low power operational amplifiers using subthreshold techniques were explored and four options emerged. There exists a trade-off between power consumption and a resistance. In order to control the current consumption of the operational amplifier, more chip area is needed for a large resistance value. Fortunately, accurate absolute resistance is not a rigid requirement, so well resistors with relatively high resistance can be used.

The constraints relating power consumption and chip area appeared to lead to a large chip area and a lower random number output rate. In order not to make these sacrifices, two system level optimization methods were employed to improve the overall performance.

One is a method of combining a TRNG and PRNG in which a 1-bit truly random number is added in the cycle ring of a PNRG so that the output sequence of the LFSR will also be as unpredictable and irreproducible as a TRNG.

By incorporating a 1-bit truly random number in the random number seed instead of generating 16 bits within the time limit, the lower limit of sample rate can be decreased to 2.2 kHz, thus remarkably cutting down the power consumption.

Power-on generation is another solution to deriving high quality random numbers with limited power consumption. The basic idea of power-on generation is to generate all the random numbers that will be used according to security protocols before other circuit blocks are awoken. Right after power on, the tag is set to random number generation mode. During this period of time, the TRNG is turned on, and most of the other circuit blocks in the tag are in sleep mode. The

tag will not respond to the "Query" command sent by the reader until all the random numbers are prepared.

In summary, this Chapter introduced the principle of an oscillator-based TRNG. By characterizing the TRNG's power consumption, sample rate, chip area, and the quality of the output, the authors show that it is possible to implement a TRNG in the RFID tag system as a solution to security problems. Finally, two system level optimization methods were proposed to reduce the power consumption of the TRNG.

Chapter 17: "A Low Cost Solution to Cloning and Authentication Based on a Lightweight Primitive"

In Chapter 17 "A Low Cost Solution to Authentication Based on a Lightweight Primitive" the authors explain how a Physically Uncloneable Function (PUF) generated within an RFID tag may be used for lightweight encryption, permitting both reader and tag authentication.

The Chapter observes that the lowest cost tags pose, because of their wide scale deployment, the greatest threat to security and privacy because constraints of silicon area and circuit complexity pose severe limitations on the possible solutions to the implementation of cryptographic primitives.

A primary concern with current low-cost RFID systems is the cloning attack, which operates to defeat the track and trace features so often desired in RFID systems. This is particularly so when the there is no mechanism for a reader to verify that it is communicating with a genuine label, not a fraudulent label. It is noted that labels and readers are constantly in an un-trusted environment where the integrity of messages is doubtful.

The paper discusses briefly the use in authentication of public and private keys, and hash functions operating on strings of plain text, but is unable to offer mechanisms that can be implemented with sufficiently small silicon area.

The goal of an authentication scheme in RFID is seen as preventing an adversary from creating a fake tag to misrepresent a legitimate tag.

Challenge and response protocols are discussed. While it is possible to construct a challenge and response protocol using a variety of cryptographic tools, the currently available solutions are expensive in terms of silicon area.

Physically Uncloneable Functions (PUF) are then defined. They have the properties that it is easy to compute a response to a challenge, but it is difficult to model the behaviour of the circuit that generated them. It is easy to construct on standard CMOS processes circuits that generate those functions. The PUF generating circuits are stated as being immune to discovery by physical attacks as they would be destroyed by the attack. They are also seen as having the property that it is not possible to tamper with the measurement data.

The idea behind the circuits to make use of process variations, which are beyond the control of the manufacturer, in wires and transistors on an integrated circuit to obtain a characteristic response from each circuit when a given input is applied. Circuit implementations for an arbiter-based POF are shown.

It is pointed out that even if a single bit responses from different circuits on a single wafer are somewhat correlated, about 800 challenge response pairs are sufficient to distinguish one billion chips with a probability of $1-5\times10^{-10}$. Such an identification scheme can be implemented with less than 1000 gates.

A block diagram of a label with a puff circuit block is provided. In the security engine there is an input and output buffer for storing the next challenge to be sent to the circuit while the output buffer will buffer up to 100 response bits. The idea is to compare responses to randomly selected challenge sets with expected responses that have been obtained from a secure database.

Some techniques in which XOR operations are conducted between different bits of a long response to produce a shorter response are described. These can have the effect of making it difficult for an attacker to analyse the content of the challenge-response pairs unless that attacker can arrange to eavesdrop on a number of different authentications in different environments.

Authentication schemes to authenticate a reader and for a reader to authenticate the tag are also described. In this scheme the tag stores a one-time pad and a secret key and the reader has access to the tag related information stored on a secure database. Again about 800 challenges are need to be able to effectively identify billions of chips.

There is also description of the tag authentication scenario when the physical implementation of the hash function achieves a critical cost effectiveness required for low-cost RFID, but regrettably it appears that there are no suitable candidates for such hash functions, and that material is seen as supporting the argument for the PUF circuit solution.

There is an evaluation matrix for schemes based on physically uncloneable functions. In that matrix the security objectives are tag authentication and reader authentication. Items quantified are: gate count in the combination of puff block, buffers and a challenge set storage; and performance in terms of the authentication speed showing about 400 milli seconds for completion. Reducing the number of challenges to 128 instead of 800 and using a LFSR arrangement to feed the PUF circuit allows the authentication speed to rise to about 600 tags per second.

In considering overhead costs it is noted that tags need to undergo a verification phase prior to deployment to generate an adequate number of challenge-response pairs for each tag, and of course that implies storage costs within the database.

Practical issues such as sensitivity to environmental conditions are also considered. It has been shown that the circuits are robust against environmental variations for realistic changes of temperature and regulated voltage. The output noise remains below 9%, such variations being significantly less than the between-chip variation.

References to probable attacks are given, but the attacks themselves are not discussed.

In conclusion it is stated that the PUF is considered to provide a cost-effective solution to authentication in low-cost RFID systems. It is considered that future work should be on elaborating common RFID protocols to allow the incorporation of authentication commands of the type discussed in this chapter.

Chapter 18: "Lightweight Cryptography for Low Cost RFID"

This chapter proposes a number of practicable solutions, based on lightweight cryptography, that address the security objectives and privacy goals outlined in Chapter 6 of this book and in the low cost RFID framework outlined therein. The proposed solutions are then evaluated for their merits using the evaluation framework developed in Chapter 8.

The majority of the proposals aim at removing complexity from the label to other proxy systems and limiting any security related computation on the chip to simple operations.

Implementations of the mechanisms are considered in the context of the C1G2 air interface protocol.

Notational aspects to improve the clarity of the discussions are reviewed, and the properties of a cycling redundancy check (CRC) of a number are discussed in detail.

Four essential approaches to stream cipher design are listed. The term linear complexity, which is an important concept in the study of stream ciphers, is defined and discussed. It is stated that since no mathematical proof of security can be found for feedback shift register based key-stream generators, system theoretical designs based on established guidelines and testable security properties are to be reviewed.

Linear Feed Back Shift Registers (LFSRs) are discussed in detail. It is noted that the output bit string of LFSRs are not secure even if the feedback scheme is not known.

Based on a system-theoretic approach the most common practices in making LFSRs secure is to use a nonlinear Boolean function to generate nonlinearity in the output, or employing irregular clocking of LFSRs. Two generators based on the previous ideas and suitable for RFID applications are considered.

Various stream ciphers based on non linear feedback shift registers, linear combination generators (the use of several LFSRs to build a single stream cipher), nonlinear filter generators and clock controlled generators which eliminate the linearity properties of the LFSRs, are discussed. The shrinking generator, implemented using a pair of simple shift registers, is considered to provide such a stream cipher and is said to be secure provided that it is implemented prudently. Its properties are reviewed. It is stated that these generators have survived much public scrutiny and they can be concluded to be computationally secure.

A number of practical guidelines that should be followed to avoid stream ciphers based on LFSRs falling to the prey of adversaries are discussed.

Two generators based on LFSRs, the nonlinear filter generator and the clock controlled generator, and are discussed. The key stream generator called the knapsack generator based on the summation of a set of weights selected based on the register values of a LFSR to generate an integer sum S, is discussed and it is stated that provided that such a sub set exists the problem has been proven to be NP-hard.

In clock controlled generators the idea is to use a combination of LFSRs so that the output of one LFSR controls the clocking of a second LFSR. This stream cipher attempts to defeat attacks based on the regular clocking of LFSRs. The

previously mentioned shrinking generator provides an example of a clock controlled generator that is suitable for implementation on an RFID label. The security of the generator has survived many known attacks on LFSR based systems, especially due to the very long period of the generator.

The order of complexity of known attacks is a function of the length of both LFSRs in the generator and has exponential time complexity. Shrinking generators are considered resistant against efficient cryptanalysis attacks due to the difficulty of the attack scenarios and the time order complexity of the algorithms. It is stated that for maximum security the following implementation considerations should be satisfied: (i) use of secret connection polynomials that are not sparse; (ii) use of maximum length LFSRs; and (iii) the lengths of LFSR should have no common divisor other than 1. The irregular output of the shrinking generator may be solved by buffering the key stream prior to its use.

Techniques for reducing power dissipation in CMOS circuits are discussed.

Although the underlying concepts and the use of Physically Uncloneable Functions (PUF) have already been illustrated in Chapter 17 of this book, the use of PUF is further extended in this chapter to provide a confidentiality service.

It is shown that the combination of a PUF circuit block with a stream cipher can create a practicable and a powerful solution capable of delivering both an authentication service and an encrypted communication channel. The mechanism is suitable for both Class I and Class II tags, especially Class II tags requiring an encrypted communication channel.

It is observed that an RFID label implementing the C1G2 protocol will scroll out its EPC when queried after being singulated by any transceiver implementing the C1G2 air interface protocol. This unique identity carried by the RFID label poses various security threats and privacy violations illuminated in Chapter 6 of this book.

Two methods to achieve anonymity and untraceability, the use of pseudonyms and re-encryption, are discussed. It is stated that the proposed mechanisms are able to satisfy a majority of the security objectives and all of the privacy objectives considered necessary in Chapter 6 and outlined in the framework provided in Chapter 8.

The use of random tag identifiers provides anonymity by altering the tag response to a query command and thus never transmitting a predictable response. It is shown that a mobile adversary who may be passive or malicious and who is able to collect all the relevant information still has the task of breaking the stream cipher given only the ciphertext.

It is noted that the chapter has outlined a scheme for providing anonymity and untraceability, and separately outlined methods of authentication. A scheme aimed at combining the previous solutions to provide, in addition, a product authentication service is then described. The proposal introduces the concept of an electronic maker. Each tag attached to a product will contain an Electronic Product Authentication Code (EPAC) with the various data fields: (i) Product Identifier, (ii) Product Signature; (iii) Signature Calculation Method; and (iv) Signature Verification Key. An evaluation of the product authentication mechanism defined here is provided.

In its conclusions the chapter states that it has used lightweight hardware and lightweight protocols to address various vulnerabilities identified in Chapter 6, as strong cryptographic solutions are too area or power hungry to satisfy the limitations of RFID systems, and much of the encryption hardware available for smart card technology is therefore inapplicable.

Part I
Anti-counterfeiting and RFID

Chapter 2
Anti-Counterfeiting and Supply Chain Security

Thorsten Staake[1], Florian Michahelles[1], Elgar Fleisch[1], John R. Williams[2], Hao Min[3], Peter H. Cole[4], Sang-Gug Lee[5], Duncan McFarlane[6], and Jun Murai[7]

[1]Auto-ID Lab St.Gallen, University of St.Gallen and ETH Zurich, Switzerland
{thorsten.staake,elgar.fleisch}@unisg.ch, fmichahelles@ethz.ch
[2]Auto-ID Lab MIT, Massachusetts Institute of Technology,
Department of Civil and Environmental Engineering, Cambridge, MA, USA. jrw@mit.edu
[3]Auto-ID Lab Fudan, Fudan University, Shanghai, China. hmin@fudan.edu.cn
[4]Auto-ID Lab Adelaide, School of Electrical & Electronic Engineering,
The University of Adelaide, South Australia 5005, Australia. cole@eleceng.adelaide.edu.au
[5]Auto-ID Lab ICU, Information and Communications University,
Daejon, Republic of Korea sglee@icu.ac.kr
[6]Auto-ID Lab Cambridge, Centre for Distributed Automation and Control,
Institute for Manufacturing, Department of Engineering,
University of Cambridge, Cambridge UK. dcm@eng.cam.ac.uk
[7]Auto-ID Lab Keio, Keio University SFC Research Institute, Tokyo, Japan
jun@sfc.keiu.ac.jp

Abstract: Counterfeit trade developed into a severe problem for many industries. While established security features such as holograms, micro printings or chemical markers do not seem to efficiently avert trade in illicit imitation products, RFID technology, with its potential to automate product authentications, may become a powerful tool to enhance brand and product protection. The following contribution contains an overview on the implication of product counterfeiting on affected companies, provides a starting point for a structured requirements definition for RFID-based anti-counterfeiting systems, and outlines several principal solution approaches that are discussed in greater detail in the subsequent chapters.

1 Counterfeit trade and implications for affected enterprises

Intangible assets constitute a considerable share of many companies' equity. They are often the result of extensive investments in research and development, careful brand management, and a consistent pledge to high quality and exclusiveness. However, the growth of markets in Asia where these intangible assets are difficult to protect, the trend in favour of dismantling border controls to ease the flow of

international trade, and the increasing interaction of organizations in disparate locations require new measures to protect these assets and safeguard companies from unfair competition. Especially product counterfeiting, the unauthorized manufacturing of articles which mimic certain characteristics of genuine goods and which may thus pass off as products of licit companies, have developed into threats to consumers and brand owners alike.

Counterfeit trade appears to affect a wide range of industries. Alongside the traditionally forged items such as designer clothing, branded sportswear, fashion accessories, tobacco products, and digital media, customs statistics show a considerable growth of fakes among consumer products as well as among semi-finished and industrial goods including foodstuff, pharmaceuticals, fast moving consumer goods, electrical equipment, mechanical spare parts, and electronic components (e.g. TAXUD 2004). The implications are numerous and wide ranging. Counterfeiting undermines the beneficial effects of Intellectual Property Rights (IPR) and the concept of brands as it affects the return on investment in research, development, and company goodwill. Producers of reputable products are deterred from investing within a national economy as long as their intellectual property is at risk. National tax income is reduced since fake goods are largely manufactured by unregistered organizations. Social implications result from the abovementioned costs: the society pays for the distorted competition, eventually leading to fewer innovative products and a less secure environment as earnings from counterfeiting are often used to finance other illegal activities (ICC 2005). However, for selected emerging markets, the phenomenon also constitutes a significant source of income and an important element of their industrial learning and knowledge transfer strategy. As a consequence, not all governments determinedly prosecute counterfeiters, which often renders legal measures to eradicate the source of illicit goods ineffective.

For companies, counterfeit trade can lead to a direct loss of revenue since counterfeit products, at least partly, replace genuine articles, a reduction of the companies' goodwill as the presence of imitation products can diminish the exclusiveness of affected brands and the perceived quality of a product, and to a negative impact on the return on investment for research and development expenditures which can result in a competitive disadvantage to those enterprises which benefit from free-ride effects. Moreover, counterfeit trade can result in an increasing number of liability claims due to defective imitation products, and may facilitate the emergence of future competitors as it can help illicit actors to gather know-how which may enable them to become lawful enterprises in the future. These implications explain the vivid interest in organizational and technical protection measures – especially since established security features have apparently not been able to prevent the increase of counterfeit occurrences. RFID technology has the potential to overcome the shortcomings of the established technologies and may become a powerful tool for product and brand protection. However, the wide range of affected products and industries, the large number of stake holders, and last but not least the considerable reengineering capabilities of many illicit actors require a thoughtful solution design and thus a careful requirement definition.

2 Requirements for Auto-ID based anti-counterfeiting solutions

The specification of Auto-ID-based anti-counterfeiting technologies is strongly influenced by the security related requirements as well as by the design parameters which stem from an integration in the desired production and inspection settings. The security related requirements can be deduced from an attack model, whereas the additional practical requirements stem from interviews with industry experts at various workshops (e.g. from Special Interest Group Anti-Counterfeiting). Both aspects are discussed below, including a description of the potential capabilities of an attacker and the security-related constraints of RFID-based systems.

2.1 Attack model

A critical design parameter of anti-counterfeiting technologies is their desired level of security which can be defined as the cost and effort that is required to compromise or bypass the system.

Since the level of security strongly influences the cost of the solution, it should be carefully adjusted to the risk (i.e. the damage and probability of occurrence) imposed by counterfeit goods. Formal attack models allow for structuring the requirement analysis. In cryptography, such models usually take the form of an "experiment," a program that intermediates communications between a fictional adversary, and a runtime environment containing the system components (often referred to as oracles) (c.f. Juels 2006). Security models have to accurately reflect real-world threats (i.e. the capabilities of illicit actors) as well as the actual system characteristics. With respect to RFID, appropriate models should not only address the top-layer protocols, but also include the basic characteristics of RFID transponders down to the bit-level. The latter may lead to a less formal description but is necessary in order to capture potentially challenging threat scenarios like power analyses (and other side channel attacks) or destructive reengineering tests; in fact, while purely algorithmic models may help to evaluate cryptographic primitives and communication protocols, they do not sufficiently capture less standardized hardware attacks which impose realistic threats to RFID systems. Therefore, the attack model that is outlined below consists of a non-formal description of the system characteristics, the capabilities of the illicit actors as well as the identification and evaluation of the potential attack scenarios.

System Capabilities

Low-cost RFID transponders are limited with respect to their maximum transistor count (as the chip size influences the transponder cost), the available energy (due to restrictions of the transmitting power of readers, the size of the antenna, and the often required considerable distance between tag and reader devices), and the frequency spectrum. This ultimately results in limited computational power, confines the memory size and communication bandwidth, and hampers the integration of sophisticated pseudo random number generators or sensors against hardware attacks.

Since cryptographic operations usually rely on computationally intense primitives, complex encryption procedures are difficult to realize in low-cost transponders. Even promising proposals that outline a lean integration of established security standards in RFID devices (e.g. Feldhofer et al. 2004) would dramatically increase the energy consumption and the required communication bandwidth, and thereby lead to lower read ranges and reduced bulk reading capabilities.[1]

More complex – and thus more expensive – transponders allow for more sophisticated cryptographic measures. In principle, the complexity of the design can escalate up to those of battery powered smart cards (i.e. active tags) with public-key crypto systems. When evaluating a potential migration path towards more secure systems, it should be considered that the silicon chip only constitutes one cost factor of the device (besides the cost for packing and the antenna) and that doubling the gate count does not necessarily double the price of the of the transponder.

Another relevant characteristic of RFID results from the radio connection between tag and reader. Connectivity is connectionless and communication is provided over an unreliable channel. This allows illicit actors to listen to the data exchange and, for example, detect existing identification numbers. Moreover, conflicts have to be considered when sharing the channel. Due to the limited power of the readers and computational constraints among tags, a more powerful sender can easily jam legitimate readers (Walters et al. 2006). An intentional violation of the tag-to-reader communication protocol, e.g. by continually transmitting messages in an attempt to generate collisions, can also disable a meaningful data exchange, which gives rise to several potential attacks.

Capabilities of Illicit Actors

The computational power and hardware complexity of low-cost RFID transponders is rather limited compared to the potential capabilities of illicit actors. Moreover, the unattended and distributed deployment of RFID transponders makes the devices highly susceptible to physical attacks. In fact, the access of the adversary to the system is a critical parameter of the attack model. Most cryptographic security analyses base on the assumption that illicit actors are able to experiment extensively with the elements of the system (e.g. Bellare et al. 1998), and thus are able to submit a large number of "oracle" queries to expose weaknesses of the design or to "guess" secret information. In this context, the limitations of RFID systems also restrict the capabilities of the attackers; illicit actors may have unlimited access only to selected transponders (e.g. after purchasing original articles with the security feature still in place), but limited access to arbitrary components. The latter is the case since attackers can only read tags which are in close proximity to their reader devices, or listen to tag reader communications which are within eavesdropping range (see Juels (2006) for a definition of various read ranges). However, in most supply chain related applications, the vast majority of transponders are hidden to other parties most of the time.

[1] In some scenarios, however, the application of such systems is nevertheless meaningful. Integration in existing standards is discussed later in this section.

With respect to transponders that are in the possession of the attackers, a wide variety of tools is available. Potential steps include power analyses and the exact measurement of response times, the application of different clock speeds or the elimination of the air interface in order to increase the frequency of queries, and hardware attacks (e.g. opening the packaged IC) in the attempt to directly read out key registers on the circuit or to reverse engineer the underlying algorithms. Therefore, when designing RFID-based anti-counterfeiting features, care must be taken that compromising accessible transponders does not affect the security of the remaining system. The protection should base on secret keys which are different and non-related among the tags rather than on secret algorithms that a large number of transponders may have in common.

Attack Scenarios

A simple attack model for low-cost RFID devices is provided by Juels (2004), who mainly addresses threats to data security, authentication, and privacy. With respect to anti-counterfeiting features, however, the focus of potential attacks is shifted to an extended set of threats. Interviews with brand protection experts conducted during this research revealed the relevance of the following issues: tag cloning which is strongly related to tag authentication, obfuscation and deception, tag omission, removal-reapplication, and, new in the context of product security features but frequently discussed in computer security, denial-of-service attacks. Each issue is addressed below.

Cloning refers to the duplication of security features such that they are likely to pass of as authentic during inspection. With respect to RFID, tag cloning may be defined as the replication of a transponder with the duplicate being able to emulate the original tag's behavior. In a system with cloned entities, investigators (or reading devices) can no longer ensure that the distinguishing mark they observe originates from the correct source; moreover, without taking the existence of duplicate features into account, observers would even falsely certify the authenticity of bogus components. Large scale tag cloning attacks can severely compromises anti-counterfeiting solutions and therefore should be addressed during the system design.

Obfuscation connotes the use of misleading protection technologies. In practice, licit companies frequently change security features to prevent counterfeiters from copying or cloning their protection technology. While following this paradigm of "creating a moving target", the licit parties unintentionally complicate the inspection process. Especially third parties can be overwhelmed by the coexistence of different, mostly visual security features. Consequently, counterfeit producers can often rely on the lack of knowledge (and the lack of time and motivation to acquire it) during inspection processes. A very common attack stems from the application of security mechanisms which are not related to the genuine product, such as the use of holograms instead of micro printings or flip colors instead of complicated packaging designs. However, the need to change anti-counterfeiting primitives when they become ineffective as well as their user-friendliness given the limited resources during inspection translates into the requirement of a flexible security system with a static user interface.

In anti-counterfeiting systems that rely on more than one component, threats may not only originate in bogus product security features but also in malicious backend systems. When a barcode, a micro printing, or an RFID transponder references a database containing track and trace information or advanced shipment notices, the authenticity of the relevant source has to be verified.

Tag Omission, i.e. the abdication of the security features by counterfeit producers even if the corresponding genuine articles are equipped with protective measures, relies on low inspection rates among many categories of goods. The phenomenon shows the need of large scale and consequently low-cost inspections. Preferably, inspections can be automated even in loosely guided processes as given in many warehouses, at customs, or at retail stores.

Removal-Reapplication attacks refer to the application of genuine security features from (mostly discarded) genuine products to counterfeit articles. This constitute a potential threat for tagging technologies where security features are being attached to an object (like holograms or RFID transponders) rather than being an inherent part of it (such as chemical markers). The consideration of this attack is of importance especially when protecting high value goods like aviation spare parts which are, when out of service, often still accessible to illicit actors. When relying on tagging technologies, a defense is to tightly couple the security feature to the object, e.g. by tamper-proofing its physical package or by establishing a logical link between object and tag.

Denial-of-Service Attacks may be defined as "any event that diminishes or eliminates a network's capacity to perform its expected function" (Wood and Stankovic 2002). Since established anti-counterfeiting technologies usually do not rely on network resources, this attack is new to the brand and product protection domain. However, when authentication processes involve entities in disparate locations, the access to these resources may be disturbed. With respect to RFID devices, attacks can cut off the connection between individual transponders and reading devices. When illicit actors target major distribution centers or customs, e.g. at harbors or airports, denial-of-service attacks may severely slow down inspection processes and thus interfere with the unobstructed flow of goods.

Eliminating any possibility of such attacks is difficult on a technical level given the limited functionality of low-cost transponders. However, providing tools for detecting attacks and localizing the illicit device is not a major issue. In actual systems, the operator would have to physically remove or deactivate the attack device.

2.2 Practical Requirements

The attack model led to a set of security related requirements. They include measures to avert a duplication of security features; the design of a stable, easy to use interface; the necessity to efficient inspection processes at low cost even in loosely guided processes; a tight coupling of the security feature to the object; and measures against denial of service attacks. In addition to this set, a number of – partly interrelated – conditions stem from the practical requirements on anti-counterfeiting

and supply chain security solutions which are not directly related to breaches of security:[2]

- *Different levels of security:* The desired level of security has a major impact on the fixed and variable costs of the solution. It can be determined i) by the risk or cost resulting from a compromised system, and ii) by the lifetime of the object which is to be protected. Risk or cost can be classified in terms of the potential health and safety hazards for consumers, or the incremental financial losses of licit manufacturers and brand owners. Depending on the probability of individual occurrence, health and safety hazards may require highly secure systems. In the context of RFID, these can be realized by the application of complex cryptographic primitives (e.g. public-key-based authentication mechanisms implemented on certified RFID transponders); for critical spare parts in the aviation industry, for example, the cost of RFID transponders may be as high as 10 EUR or above. However, if illicit products primarily cause incremental financial losses (e.g. due to dissatisfied consumers and substitution effects), a detailed cost-benefit analysis is helpful in order to select an appropriate protection mechanism.
- *Migration path:* Anti-counterfeiting technologies often constitute a barrier for illicit actors only for a limited, unknown period of time. Holograms, for example, have been considered highly secure features when introduced and are now widely available on the market. Consequently, it is desirable to have the opportunity to change the underlying security primitive at low cost, i.e. without the need to alter the technical infrastructure or to require the user to get accustomed to new checking procedures. RFID technology, if properly designed, allows for separating user interfaces and underlying technologies, and may therefore constitute a sustainable solution.
- *Manufacturing requirments:* Existing manufacturing settings are often highly optimized with respect to throughput and down times. The addition of supplementary process steps can severely impact the key performance measures of the production facilities. This is especially the case in high volume production environments e.g. in the pharmaceutical or fast moving consumer goods industry, where the required line speeds severely limit the technology choice. Process steps that are necessary to integrate security features have to be as non-intrusive as possible.
- *Product specific requirements:* Prouct related characteristics can impose a number of additional constraints on the technology choice. Restrictions may result from the available size for such features, the object's material, and operating conditions such as temperature, electrical discharge, abrasion etc. When the security features are to be deployed at an early stage of the production process, aggravated conditions may apply. The product specific requirements should be analyzed on a case-by-case basis at an early stage of the design process.

[2] The practical requirements result from a group work undertaken during the second Special Interest Group workshop in Hamburg at July 1, 2005.

- *Invariance of the product design:* I order to enhance the level of security, it is desirable to integrate the security features in the product and not to rely on tagging its packing. However, companies seem to be rarely willing to subordinate product design to anti-counterfeiting measures. This limitation may further complicate the tag-in-product integration.
- *Technology specific requirements:* Idividual security technologies may be chosen due to the specific advantages they exhibit such as the possibility to automate checking processes, which may have to be defined in greater detail. With respect to RFID, bulk reading (i.e. the number of tags which can be read quasi-simultaneously; read rates (i.e. the share of transponders which is actually detected during a bulk read); read ranges (i.e. the maximum distance between tag and reader during the inspection process); data standards, etc. are to be considered.
- *Confidentiality:* Last but not least, securiy features shall not reveal confidential information of the manufacturer (e.g. on production output) nor infringe the privacy of the user or consumer.

Depending on the actual application, several solutions concepts are applicable which are outlined below.

3 Solution concepts

RFID technology comes at various levels of complexity and offers several functionalities which make it applicable as anti-counterfeiting measures in various application scenarios. The following section discusses in greater detail the usability of unique serial numbers, a technique to avert removal-reapply-attacks, and the usability of tags with authentication capabilities in a standard reader environment.

3.1 Using unique serial numbers

Marking objects with unique identifiers, i.e. on item-level rather than for individual project categories only, helps to monitor the flow of goods and thus to detect illicit trade activities. If designed carefully, a numbering system can significantly reduce counterfeit trade. The latter is possible if an approach is chosen which is difficult to apply for illicit actors, but whose identifiers are easy to check for supply chain partners or end-users. In an ideal scenario,

- the number is assigned in a random way, with the numbering space significantly larger than the number of items to be identified, so that illicit actors are unlikely to simply guess valid IDs,
- the validity of the number can be easily checked by the supply chain partner or, if desired, by the consumer,
- the number can be read automatically by authorized persons, allowing for large scale searches for invalid or duplicate identifiers, thus increasing the chance to seize illicit goods,

- a duplication of the number carrier is unreasonably expensive, and
- the number carrier cannot be removed nor can illicit actors overwrite the number which would allow them to disguise the identity of the object.

The basic operating principle of a unique ID system is quite simple: The manufacturer generates a random number, writes it to the data carrier and stores it in a database. When the product ID is checked e.g. in a store or at customs, a reader device retrieves the product ID, sends it to a service offered by the manufacturer (or an IT provider) which looks up the number in the database and returns the result to the reader device. An operational implementation, however, should provide additional features such as a system for user access management that prevents illicit actors from discovering licit numbers or competitors from monitoring the flow of goods. When the system is applied by a larger number of vendors, it becomes impractical to store the access information of individual service providers on the reading devices; therefore, the system should contain a lookup system which allows the readers to retrieve the corresponding addresses from a known online source.

3.2 Plausibility checks based on track and trace

A track and trace system is a potentially powerful tool as it can provide an enormous degree of supply chain visibility. In principle, information on an object's location and the corresponding time, possibly together with data on the owner, its status, the object's operating conditions, etc., is recorded and stored for further processing. If such measurements are repeated over time, they allow for plausibility checks of the recorded products history. Heuristics can be applied as for example done by credit card companies which routinely freeze cards if they inhabit a suspicious transaction history.

Track and trace systems rely on the ability to uniquely identify individual articles. In order to facilitate meaningful analyses, numerous data points have to be collected, which requires an efficient way to capture supply chain events. In this regard, RFID can be seen as an enabling technology. Though the operating principles of track and trace systems may appear simple, an actual implementation of the infrastructure is a considerable challenge. From a technical perspective, especially access management in non-predetermined supply chains constitutes a major hurdle. However, even bigger barriers seem to be organizational issues on the ownership of the data, the distribution of system costs, and the lack of interest among some industries to provide a higher degree of supply chain visibility for their customers. In fact, the solution requires numerous stakeholders – often with conflicting interests – to work together, which renders it impractical in many application scenarios. However, track and trace is likely to become the dominant solution in highly regulated industries where powerful stakeholders can enforce an adoption (e.g. the Food and Drug Administration with respect to pharmaceuticals).

3.3 Object specific security

Security solutions that are based on tagging technologies have a system specific drawback: when checking an object, it is still the tag (e.g. a hologram or a simple RFID transponder) which is authenticated and not the object or document the tag is attached to. The link between tag and object is often not strong enough. In theory – and also in practice if the solution is not designed properly – a tag can be removed from an original article and attached to another object, thereby compromising the security system.

In contrast to most other tagging technologies, RFID can overcome this shortcoming. Even low-cost RFID tags with a certain amount of memory can store data that binds a tag to a given product, as a picture in a passport binds a passport to its holder. As a result, illicit actors are detained from simply removing a tag from a legitimate product and reapplying it to a counterfeit article in a way that the fake is not detected during product validation. An in-depth description of this technique is given in the chapter 13, "Product Specific Security Features Based on RFID Technology".

3.4 Secure authentication

Concepts that allow for a proof of identity (or authentication) are common in computer systems. However, establishing efficient means of authentication in RFID infrastructures constitutes a major challenge. The lack of cryptographic functionalities of basic RFID transponders is a big impediment to current designs. Since serial numbers are usually not read protected, an attacker can obtain an identifier from a tag and program it into another transponder, or emulate the tag using other wireless device. If done at a larger scale, duplicate devices render track and trace or anti-counterfeiting solutions ineffective.

Challenge-response protocols can avert tag cloning as they allow for a comparison of secret keys at disparate locations without transferring them over a possibly insecure channel. In a carefully designed system, third parties are prevented from reconstructing the secret, even if it is used numerous times. These properties qualify challenge-response protocols for an application in an RFID environment, where the channel must be regarded as insecure.

Critics of this approach frequently mention the increasing tag costs which may result from the integration of the required cryptographic unit in RFID transponders. However, Feldhofer et al. (2005) showed an implementation of an 128 bit version of the Rijndael cipher (Daemen and Rijmen 2002) using less than 4,000 gates, which, given an approximate gate count of current EPC Gen2 tags of 15,000, would only lead to a small increase in tag cost.[3] This motivates further research on the actual integration of authentication protocols in RFID systems and thus is a major topic in the remainder of this book.

[3] This is especially true since the cost of the actual chip is only one component of transponders which is also made up by packing, the antenna, assembly, etc.

References

1 Daemen, J. and Rijmen, V. 2002. The design of Rijndael: AES – the Advanced Encryption Standard. Berlin, Germany: Springer

2 Feldhofer M., Dominikus S., and Wolkerstorfer J. (2004). Strong authentication for RFID systems using the AES algorithm, Proceedings of the Workshop on Cryptographic Hardware and Embedded Systems (CHES '04), pp. 357–370

3 Feldhofer, M., Wolkerstorfer, J., and Rijmen, V. 2005. AES implementation on a grain of sand. In Proceedings, Information Security, 152(1): 13–20

4 ICC International Chamber of Commerce (2005). Current and emerging intellectual property issues for business − A roadmap for business and policy makers. Document no 450/911 Rev. 6 (Paris, France: ICC, March 2005), p. 2. www.insme.info/documenti/Roadmap-2005-FINAL.pdf.

5 Juels A. (2004). Minimalist cryptography for low-cost RFID tags, RSA Working Paper, www.rsasecurity.com/rsalabs/staff/bios/ajuels/publications/minimalist/Minimalist.pdf.

6 Juels, A. (2006). RFID security and privacy: a research survey, IEEE Journal on Selected Areas in Communications, 24(2): 381–394

7 TAXUD European Taxation and Customs Union (2004). Breakdown of the number of cases registered and the number of articles seized by product type: EU – 2004, http://ec.europa.eu/taxation_customs/resources/documents/customs/customs_controls/counterfeit_piracy/statistics/counterf_comm_2004_en.pdf.

8 Walters, J. P., Liang, Z., Shi, W., and Chaudhary, V. 2006. Wireless sensor network security: A survey. In Xiao, Y. (Ed.), Security in Distributed, Grid, and Pervasive Computing: Chapter 17. Sound Parkway, NW: CRC Press

9 Wood, A. D. and Stankovic, J. A. 2002. Denial of service in sensor networks. Computer, 35(10): 54–62

Chapter 3
Networked RFID Systems

Damith C. Ranasinghe[1], and Peter H. Cole[1]

[1] The Auto-ID Lab, School of Electrical and Electronic Engineering,
The University of Adelaide, SA 5005, Australia. {damith, cole}@eleceng.adelaide.edu.au

Abstract. A combination of Radio Frequency Identification technology and ubiquitous computing are revolutionising the manner in which we look at simple objects. Radio Frequency Identification (RFID) allows RFID labeled objects to be identified at a distance without physical contact, and ubiquitous computing provides a virtually connected environment for the objects. RFID labels are frequently referred to as the next generation barcodes. RFID Systems provide increased productivity, efficiency, convenience and many advantages over bar codes for numerous applications, especially global supply chain management. RFID labeling has a number of advantages over conventional bar code systems. The optics based bar code systems could be rendered useless by common everyday environments containing dirt, dust, smoke, grease, condensation and by misorientation and misalignment. Furthermore bar codes are subject to fraudulent duplication and counterfeiting with minimal effort. While there is a googol of information on Radio Frequency Identification systems, most of which arose in the last decade, it is important to identify concepts and operating principles and to present views on modern RFID systems. This chapter will provide an overview, however brief, of modern RFID systems.

Keywords: EPC, RFID System, Anti-collision, UHF, HF, Tag, Reader

1 Introduction

An evolving set of identification technologies continues to play an important role in the provision of identification and other related information about entities both animate and inanimate in various contexts, ranging from Allied air planes in World War II [1] to items stacked on supermarket shelves. The technologies used range from biometric data recognition, optical character recognition, magnetically coded ink (MICR) used on cheques, magnetic strips used on credit cards, Wiegand wire and barium ferrite inserts used in access control badges, speech recognition, smart cards and barcodes [2]. The term “Auto-ID” is used to refer to systems based on the above technologies because automatic identification is almost always a primary function performed by such systems.

Amongst the identification technologies, barcodes, of which a standardised form (Universal Product Code) developed by a consortium of inter-industry trade associations called the Uniform Code Council (UCC), recently renamed as GS1 [3], are available for consumer items [4], and have continued to proliferate throughout the world as a solution to both logistic and inventory control by providing a product identification code. The proliferation of Universal Product Code (UPC) technology was aided by the continued growth in the commerce of goods and the need for efficient logistical support for global trade and transport of goods. A research report compiled by PriceWaterhouseCoopers in 1999 entitled "17 Billion Reasons to Say Thanks" depicts the impact of the UPC on international economies and industry and on the consuming public. Since the inception of the idea by Norman Joseph Woodland in 1949, various forms of the barcode have been developed over the years [5]. One such example is the 2D bar code which can carry a considerable amount of data on a smaller surface area than compared with its predecessor [3]. The latter fact is made clear by comparing Figure 1 and Figure 2 which show that the space required to encode 62 characters using 2D barcodes is much less than that possible with conventional barcodes.

UPC codes are a contactless technology where scanners work in the range of a few centimeters but require a clear vision of the bar code in the scanning field. Bar code labels have to be printed carefully and exactly since line thickness and spaces signify the alphanumeric characters forming the UPC, while the objects labeled with barcodes must be physically maneuvered to align the barcodes with the scanners. In addition, barcodes are not suitable for hostile environments where optical recognition of the barcode may not be possible. Another common failure of barcodes is that they are prone to physical damage as the printed symbols may be smudged or concealed, making the process of scanning difficult or impossible. These issues limit the usefulness, possible application and performance of barcode based "Auto-ID" technologies.

An "Auto-ID" technology devoid of many of the imperfections of barcodes has emerged to rival the simple barcode. This technology is commonly referred to as Radio Frequency Identification (RFID). Radio Frequency Identification systems allow automatic identification using a unique identifier associated with an object or person. RFID technology has been in existence for more than sixty years with the earliest application of RFID commonly cited as the "Identify Friend or Foe" system used in Allied aircraft in the Second World War [1]. Current applications of RFID are a diverse collection of identification and data capture applications. Some typical applications are automatic toll collection, animal identification, proximity cards for secure access, authentication, theft detection, tamper proofing containers and last but not least, supply chain logistics [2, 6].

This chapter investigates advanced RFID systems. What soon becomes apparent with study of advanced RFID technology is the large number of issues

Fig. 1 A conventional barcode with 10 encoded characters.

Fig. 2 An example of a 2D barcode with 62 encoded characters.

leading to disappointing performance, and the imminent threats posed by vulnerabilities of low cost RFID systems. The following section introduces the components of a modern RFID system.

2 Radio Frequency Identification Systems

A simple illustration of the concept of an RFID system is provided in Figure 3. Here, a transmitter of interrogation signals which is contained within an interrogator, communicates via electromagnetic waves with an electronically coded label to elicit from the label a reply signal containing useful data characteristic of the object to which the label is attached. The reply signal is detected by a receiver in the interrogator and made available to a control system.

There is a wide range of operating principles for such a system [6, 7, 8 and 9]. The operating principle and operating frequency are driven principally by the application of the labelling system and the constraints provided by electromagnetic compatibility regulations, environmental noise, and the ability of fields to permeate a scanned region of space or to penetrate intervening materials.

Over the years a number of RFID systems have evolved with advances in material science, microelectronics, and fabrication technology. Irrespective of the underlying technology and the type of labels around which an RFID system is built, (that is, microelectronic labels, surface acoustic wave labels, labels using multiple resonances to encode data and so on [7]) all modern RFID system infrastructures can be categorised into three primary components given below.

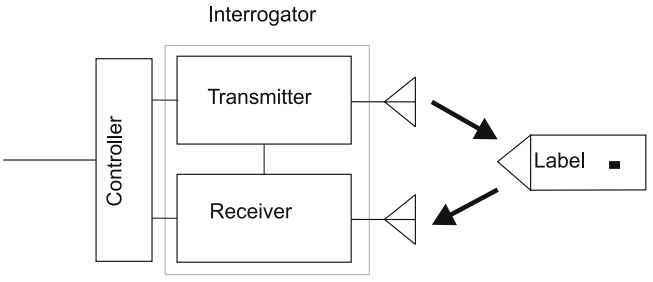


Fig. 3 An illustration of an RFID system.

1. RFID labels (transponder)
2. RFID label readers or interrogators (transceiver)
3. Backend network (electronic databases)

The material presented in this chapter will consider RFID systems based on using microelectronic devices for RFID labels.

3 RFID Labels

Generally, a microelectronic RFID label consists of a small microchip with some data storage and limited logical functionality, and an antenna. The antenna allows the label to couple to an electromagnetic (EM) field to obtain power, or to communicate with the reader, or to do both. Figure 4 provides an illustration of the components of an HF and a UHF RFID label.

RFID labels can be distinguished based on their frequency of operation as listed below and described in Table 1.

- Low Frequency (LF)
- High Frequency (HF)
- Ultra High Frequency (UHF)
- Microwave

Labels may also be categorised based on powering techniques used to operate the microelectronic circuitry as listed below and compared in Table 2.

- Passive
- Semi-Passive
- Active

In a primary category of passive systems, the most common operating principle is that of RF backscatter or load modulation [6, 7] in which a powering signal or communication carrier supplies power or command signals via an HF or UHF link. However, the circuits within the label operate at the carrier frequency or at

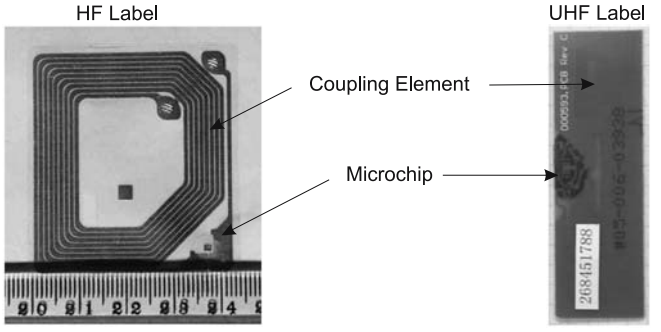


Fig. 4 Components of an RFID Label.

Table 1 Characteristics of tags based on their frequency of operation.

	LF	HF	UHF	Microwave
Frequency Ranges Used	125–134 kHz	13.56 MHz	860 MHz – 960 MHz	2.45 GHz 5.8 GHz
Data-Transfer Rates	Low	Medium	High	Very high
Advantages	Tags are less affected by metals and liquids	High data-transfer rates, cheaper than LF tags	Higher data-transfer rates, cheaper to manufacture. Less sensitive to environmental detuning.	Very high data-transfer rates.
Disadvantages	Low data-transfer rate Tags are expensive because of the copper required for larger antennas.	Affected by metals and liquids	Poor near metals and liquids	Poor near metals and liquids. Line of sight required for long distance communication.
Typical Applications	Access control, animal tracking	Contactless smart-cards, Libraries, and Access cards. Item level tracking.	Item level and pallet level tracking of goods in a supply chain	Electronic toll collection

a lower frequency, and reply via sidebands generated by modulation within the label, or by modulation of a portion of the powering carrier. This approach combines the benefits of relatively good propagation of signals at HF and UHF and the low power operation of microcircuits at RF or lower. Powering at UHF is employed when a longer interrogation range (several metres) is required, and HF powering is employed when electromagnetic fields, which exhibit good material penetration and sharp spatial field confinement are required, or sometimes when a very low cost RFID system implementation is desired.

The data stored on an RFID label can be used to uniquely identify the object, person or animal associated with the label. Unique identification requires a unique identifier defined according to some standard. Among the various identifiers used in different industries, the EPC [10, 11], which is a unique product identification code format, is being standardised for use in RFID applications. An EPC typically

Table 2 Comparison of passive, semi-passive and active labels.

	Passive labels	Semi-Passive labels	Active labels
Powering	Incident RF signal	Battery	Battery
Transmission	Backscatter	Backscatter	Powered from a battery
Typical read range	3 m – 5 m (UHF)	10 m – 20 m	1000 m

contains information that identifies the manufacturer of the item to which a tag is attached, the type of item and the serial number of the item. This information is also referred to as the label ID.

It should be noted that an EPC is one form of a unique numbering scheme proposed for RFID applications and is a result of an ongoing effort to standardise the data format on labels. The widespread adoption of EPC is a result of its suitability for object identification in the global supply chain. It has been demonstrated that, through supply chain visibility, manufacturers, distributors, retailers and other intermediaries can reduce inventories and ensure product availability, resulting in tremendous cost savings and increases in efficiencies. This is currently the largest market for RFID technology and the potential cost savings and efficiencies are the drivers that are propelling RFID technology into the future. Thus, throughout this chapter, the author will primarily focus on examples related to the global supply chain, while a closer look at the implications of RFID technology to supply chain applications is given in Chapter 4.

3.1 Label to Interrogator Communication

Labels perturb the field created by an interrogator to achieve near or far field, label to reader communication. However there are similarities and differences in the way communication is achieved in both the far and the near field by a label antenna.

In near field systems, a reader senses the reactive power flowing in the label. A label is able to control the reactive power flowing through itself by varying the load across its antenna [6, 7].

However in the far field, labels cause some of the incident RF energy to reflect back or scatter. This is referred to as backscatter. Similarly to the manner in which RFID labels in the near field vary the load across the label antenna to vary the reactive power, a label in the far field also varies the load across its antenna to modulate the amplitude and phase of the backscattered wave. In the far field, variation of a label's load impedance causes an intentional mismatch between the label antenna and its load. Backscattering a label response in this manner is akin to the label antenna radiating its own weak signal [6, 7, and 9].

Variations in the load across a label antenna can be achieved by either switching on and off a resistor or a capacitor across the label antenna. The particular mechanism used gives rise to ohmic load modulation and capacitive load modulation, respectively [6, 7 and 9].

In the near field, the real power received by a label is enhanced by using a high quality factor in the label antenna resonance [8], while the reactive power which flows in the label is further amplified above the real power by that same quality factor. Thus the quality factor of the label antenna increases the effect of the changes in reactive power flowing in the label.

However, in the far field, the power backscattered is not aided by the quality factor of the antenna where the radiation resistance of the antenna is much larger than the antenna losses. Such antennas are generally considered lossless and the best that can be achieved is to reflect all the incident power back to the interrogator

while modulating the amplitude or phase of the reflected wave in a time varying manner. This is not fully achievable since some power must be used to power the microcircuits of the label antenna. The latter concepts and coupling relations in the near and far fields are more comprehensively considered in [12], wherein it is shown that excitation fields are best described in the near field by reactive power density per unit volume and in the far field by power flow per unit area in the Poynting vector. The corresponding measures of tag coupling are coupling volume for the near field and effective area for the far field. However, the coupling volume concept can be shown to be the more general in that it may be used in both the near and for fields.

3.2 EPC Concept

The concept involves identifying objects through a uniquely formatted number kept on each label as a data field, with associated data, stored in a backend system database. The unique object identifier must have a global scope that is capable of identifying all objects uniquely and act as a pointer to information stored about the object somewhere over the network. The Electronic Product Code (EPC) is a scheme designed for universal object identification with the associated standards developed by EPCglobal Inc [13]. A binary representation of the EPC is shown in Figure 5 [13]. The Header identifies the EPC format used by the tag; 96-bit, 64-bit or 256-bit while the General Manager Number identifies an organisational entity. These numbers need to be unique and can be assigned by a standards body such as EPCglobal Inc. The Object Class is used by a General Manager to identify a specific product class. The Serial Number is a unique number within each Object Class.

MSB LSB

Header	General Manager number	Object Class	Serial number

Fig. 5 Bit level representation of an EPC general type identifier format.

It is important for serial numbers to be unique for objects labelled by a particular organisation within a product class. However, different objects may reuse the same serial numbers, as the difference in product codes will ensure unique identification of the product. Hence, the triplet of General Manager Number, Object Class, and Serial Number uniquely identifies an object.

3.3 Label Hierarchies

A classification of RFID labels proposed by the former Auto-ID Center is based on a functional hierarchy. The functional classes have emerged as a result of discussions between the Auto-ID Centre and sponsor companies. The formulation of a label class hierarchy permits the development of a range of labels to suit different application requirements while setting realistic expectations for functionality.

This is an important aspect of the development of this technology, as the environmental, cost, functional, operating range, security and privacy requirements will vary depending on the application employing RFID technology. Figure 6 outlines the proposed class hierarchy. The Class 1 labels with simple read-only EPC data form the backbone of the evolving label class hierarchy. The following section provides a brief overview of the functional classes.

In the scheme outlined below, the higher classes exhibit an inheritance of characteristics of the lower classes; in particular they are all based on the EPC concept. Clearly each class consists of a set of diverse functionalities. For instance, Class II labels are expected to possess a multitude of functionality on a low cost passive label and this requires careful attention in defining its specification.

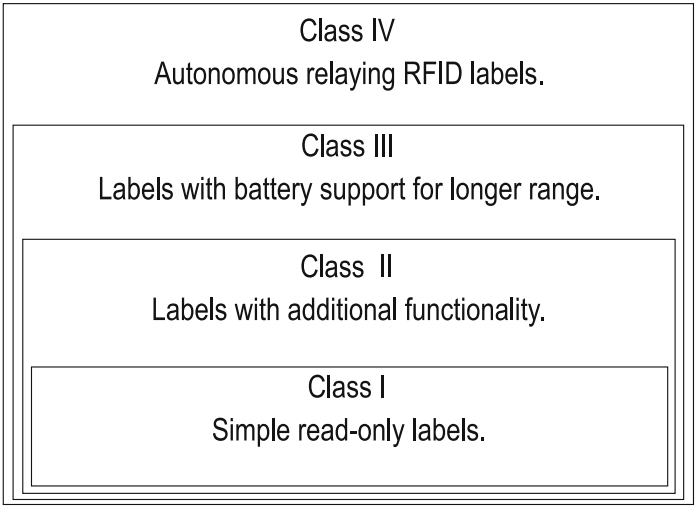


Fig. 6 An outline of the label class hierarchy.

3.3.1 Class I

Class I labels provide minimal required functionality. In addition to a read only write once EPC code, a label falling into the Class 1 category carries a CRC for transmission verification, and a password that is used in the process of label destruction. As a guarantee of customer privacy, labels can be electronically destroyed before goods are released to the purchaser at their final point of sale. Class I labels are low cost RFID labels that are typically passive.

The latest standard being ratified by EPCglobal for UHF RFID communication in the 860 MHz – 930 MHz band known as Class 1 Generation 2 (C1G2), is developed to standardise the air interface for Class I labels [14].

3.3.2 Class II

EPC labels with additional functionality, such as read-write ability, provision for data security, privacy, and theft detection ability. Class II labels are likely to be passive with limited read range.

3.3.3 Class III

Class III labels contain battery support for long-range communications. These labels may be active or semi passive and may support broadband communication.

3.3.4 Class IV

These are active labels capable of peer-to-peer communication with other Class IV labels utilising the same frequency. Class IV labels are increasingly being viewed as autonomously networking labels that are able to form ad-hoc wireless networks.

3.4 A Classless RFID Label Society

The dilemma of diverse functionalities may be resolved by means other than a rigid hierarchy, as outlined above, by writing standards that allow manufacturers to define and extend functionalities without having to revise or continually rewrite standards.

The RFID label will continue to evolve and be more than just an identifier. The fusion of sensor technology with RFID will create a vibrant and diverse future for RFID with labels capable of a variety of different functionalities in addition to those related to security and with various product specific functions. Given such a pool of as yet undefined capabilities for an RFID label we should not limit ourselves to an inheritance based hierarchy that is inflexible.

Instead it is possible to create a classless society where the functionality of a label can be related to its EPC or a further EPC and defined on a database. It is then possible to cache that database in readers, given that it is relatively small. With this philosophy it is possible for manufacturers to be creative and define methods for accessing those tag capabilities.

If, however, distinction is required, as it is human nature to classify and categorise all that exists, it is possible to have Class I labels, Class II Simplex labels (labels with read write memory) and Class II complex labels (labels processing read write memory and unlimited functionality, in theory).

4 Interrogators

The readers communicate with the labels using an RF interface. Either a strong energy storage field in the vicinity of the reader antenna, or radiating EM waves, establish the RF interface. The communication between a reader and a label

process may involve interrogating the label to obtain data, writing data to the label or transmitting commands to the label so as to affect its behaviour. The readers consist of their own source of power, processing capability and an antenna. The general design principle in EPC based RFID systems is to off load silicon complexity of the label to backend systems and to the reader in order that the cost of the labels may be kept to a minimum. Thus, readers have ample computing power and are capable of acting as proxies to perform computationally intensive tasks for RFID labels. However, increasingly in pursuit of low cost readers, reader architectures are being simplified to manage cost and complex tasks are being pushed further back to the backend systems. A more detailed look at interrogators can be found in [15].

5 Anti-Collision

An important aspect of an RFID system is being able to read multiple labels in a relatively short period of time as in any realistic system there is likely to be more than one label in the field of a reader. Implementing such plurality of reading requires a means of overcoming collisions between label replies (refer to Figure 7). Here the term ‘collision’ is used to imply the resulting signal interference from two or more tags replying simultaneously. A collision generally results in a failed communication. There are a number of anti-collision algorithms used to implement such plurality of reading and prevent the occurrence of collisions. These algorithms are embedded in multiple tag reading protocols.

The anti-collision needs of RFID systems are similar to multiple access communication conflict resolution or detection methods used in various computer networks such as Carrier Sense Multiple Access (CSMA) [16]. However RFID anti-collision methods are constrained by the limited computational power, memory and power constraints of an RFID label. Anti-collision methods used in RFID must consider the wireless and ad hoc nature of RFID networks along with the necessity to recover from sudden power loss, as is the case with passive RFID systems.

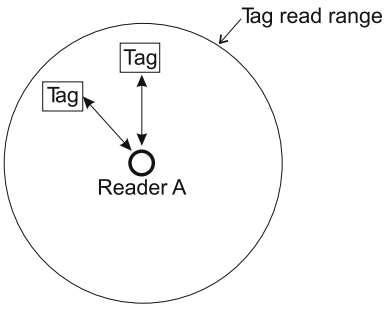


Fig. 7 Tag reply collision.

There are a wide variety of anti-collision algorithms. The vast majority of existing anti-collision methods are time-domain methods which are either deterministic or probabilistic. The most widely used deterministic methods are based on the binary tree walking protocol while most probabilistic methods are based on Slotted Aloha [17]. These schemes are intended to reduce the occurrence of multiple simultaneous responses from tags to a reader query.

In addition to the feature which reduces the frequency of collisions, the capacity to detect collisions is a powerful addition to an anti-collision algorithm. Most commonly used techniques rely on line coding schemes. When simultaneously transmitted signals, even of different strengths, coded by line coding schemes, interfere they will always be demodulated by an interrogator attempting to decode a demodulated signal from an RFID label. However an interrogator attempting to decode those signals may fail due to the decoding process producing an invalid symbol as a result of the type of line coding scheme used by the tag to interrogator communication. Level coding schemes such as NRZ and RZ coding schemes are example schemes where collisions will not cause a resulting invalid symbol at the interrogator during the decoding process, while the Manchester coding technique and other transition coding techniques allow collision detection by preventing the proper decoding of the received signal at the interrogator [6] by the resulting invalid symbols that may occur as a result of a collision.

With two messages of unsynchronised bit periods, the level codes will decode to some valid symbol (even though they are not the symbols in either message), but transition codes might lose or develop transitions in the middle of presumed bit intervals and produce invalid symbols (dissimilar to those present in either message). Nevertheless it should be noted here that when a weak tag to interrogator signal collides with a strong tag to interrogator signal, no line coding scheme will allow the detection of a collision. Instead the strong signal will tend to mask the weak signal.

While CRC codes appended to communication between readers and tags are not part of the anti-collision algorithm or a collision detection method, due to the properties of CRC codes, it is possible to detect a possible collision using CRCs irrespective of the properties of the line coding scheme. It is important to stress the word ‘possible’, since a weak tag reply and thus a poor signal to noise ratio in the received reply from a tag may also cause an invalid CRC. In the event a tag reply to an interrogator is correctly decoded (but the reply is not that intended from either one of the tags attempting to communicate simultaneously), the CRC calculated on the reply may not match that part of the message interpreted at the CRC. This is because the CRC itself may have been incorrectly decoded, or in the event the CRC is that which was sent by one of the tags, it may not match that calculated by the interrogator for the received message.

In addition to the collision between label responses, applications requiring readers to operate in close proximity to other readers can cause the interrogation signals from one reader to interfere with signals from other readers [17]. There are several instances where readers can be the source of a collision.

A simple scenario is when an interrogation signal from a nearby reader interferes with a weak reply from a tag outside the read range of the interfering reader during an interrogation of that tag by another reader in the read range of the tag.

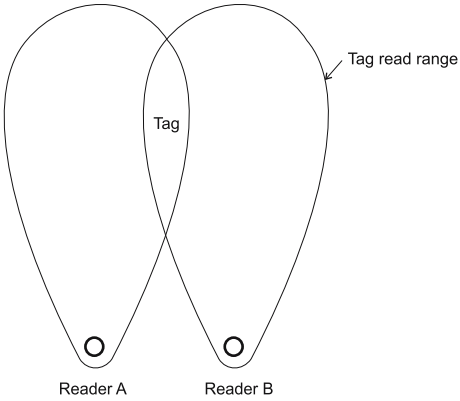


Fig. 8 A practical interrogator arrangement where carrier sensing can fail.

Such a problem can be mitigated by using Frequency Division Multiple Access (FDMA), as has been demonstrated by the C1G2 protocol [14] which calls for the spectral separation of tag replies from the reader transmission frequency, while readers themselves operate at a number of different channels in the allocated band.

Another situation may arise resulting in tag confusion when multiple readers simultaneously attempt to read the contents of a tag within their tag read range. Such a collision can cause the misinterpretation of the reader communication resulting in an incorrect reply or a complete failure to reply because the command can not be decoded. In such a situation, a carrier sensing scheme may be utilised to prevent concurrent access to a tag by nearby readers [15, 16].

However, using carrier sensing becomes a difficult issue if the reader antennas are directional, as it is often the case with practical implementations. Hence a reader may not detect the RF field of another active reader because both reader antennas are not in the RF field created, as a result of beam forming to increase antenna gain. While it is possible for a tag to be in read range of both readers as illustrated in Figure 8, the RF field of the readers may not be detected by either reader antennas. When such a phenomenon occurs, the commands or queries from different readers collide and the resulting signal will cause the misinterpretation of the intended reader communication. Since tags themselves are made simple, to control costs, they can not be expected to participate and aid in a collision avoidance scheme.

In addition to the phenomenon discussed above, tag collisions, reader collisions, noise from other readers and tags, and echo of tag replies and reader replies has lead to the observation of phantom tags. In this situation, a reader reports an EPC of a tag that does not exist in its tag reading range. However the C1G2 protocol was postulated with the idea of eliminating the phenomenon of ghost reads by the following mechanisms:

- Tags must respond to a reader within a short time frame, so that responses that do not fall into this time frame are eliminated.
- Tags transmit a preamble with every transmission. A preamble is a predefined signal sequence that the reader expects from a valid tag signal. This reduces the likelihood of a noise corrupted tag signal being interpreted as a valid tag by a reader
- The reader performs a validation check of the EPC, to confirm if the reply contains a valid preamble. This is achieved by checking that the data received conforms to a valid EPC format as defined in the tag data specification [13].
- The reader then performs a third validation check to ensure the number of bits received in the response is the same as that reported by the tag. Any mismatch results in the reader discarding the response.
- Finally the reader performs a CRC check on the data received to check for bit errors.

6 Conclusion

There are complex hardware devices (labels and readers) and software that bundle together to form an RFID infrastructure. Wide scale adoption of this technology requires an overall reduction in the cost of these devices. Thus, there is a great deal of interest in low cost RFID technology, which essentially involves reducing the cost of RFID labels.

There is a variety of RFID devices designed for various applications. The Auto-ID Center took a modular approach to accommodate this diverse application base and technological possibilities. The EPC concept and the Class structure are a result of this modular thinking.

Class I and Class II labels fall into the low cost end of the RFID label hierarchy. The microcircuits of these classes are extremely resource scarce with only hundreds of bits of storage and only thousands of gates available for logic functions. These circuits are designed with strict limits on power consumption and cost.

This chapter has provided a brief introduction to a very broad subject area but provides a number of references where further information can be obtained. While the RF identification aspects are commonly discussed in the literature, there is generally very little information regarding the backend infrastructure vital to the formulation of a modern RFID system. Such discussion can be found in the next chapter.

References

1 Royal Air Force website. Available from: http://www.raf.mod.uk/history/line1940.html
2 Ames, R.: Perspectives on radiofrequency identification. Van Nostrand Reinhold, New York (1990)
3 GS1 Home page. Available from: http://www.gs1us.org/

4 Haberman, A.L.: Twenty-five years behind bars. Harvard University Wertheim Publications (2001)
5 Woodland, J., Silver, B.: US Patent No. 2,612,994, Oct. (1952)
6 Finkenzeller, K.: RFID Handbook: Radio Frequency Identification Fundamentals and Applications. John Wiley & Sons, New York (1999)
7 Cole, P.H., Hall, D.M., Loukine, M., Werner, C.D.: Fundamental constraints on RF tagging systems. In: Proceedings of the fourth annual wireless symposium and exhibition, Santa Clara (1995) 294–303
8 Eshraghian, K., Cole, P.H., Roy, A.K.: Electromagnetic coupling in subharmonic transponders. In: Journal of Electrical and Electronic Engineering, Vol. 2 (1982) 28–35
9 T. A., Scharfeld, "An analysis of the fundamental constraints on low cost passive radiofrequency identification system design", Masters thesis, Massachusetts Institute of Technology, August 2001
10 Brock, D.L.: The Electronic Product Code – A naming scheme for physical objects. In: Technical Report MIT-AUTOID-WH-002, Auto-ID Center, January (2001). Available from: http://www.autoidlabs.org/researcharchive/
11 EPCglobal Inc. home page. Available from: http://www.epcglobalinc.org.
12 Ranasinghe D.: New Directions in Advanced RFID Systems PhD Thesis, University of Adelaide (2007)
13 EPCglobal Inc: EPC Generation 1 Tag Data Standard Version 1.1 Rev. 1.27, May (2005)
14 EPCglobal Inc.: Class 1 Generation 2 UHF Air Interface Protocol Standard Version 1.0.9: "Gen 2" (2005). Available from: http://www.epcglobalinc.org/standards/
15 Jamali, B.: High performance RFID systems, PhD Thesis, University of Adelaide (2005)
16 Roshan, B., Leary, J.: 802.11 Wireless LAN Fundamentals, Cisco press (2003)
17 Engels, D.W., Sarma, S.E.: The reader collision problem. In: IEEE International Conf. on Systems, Man and Cybernetics, Vol. 3 (2002) 6–13

Chapter 4
EPC Network Architecture

Damith C. Ranasinghe[1], Mark Harrison[2], and Peter H. Cole[1]

[1] Auto-ID Labs, School of Electrical and Electronic Engineering,
The University of Adelaide, SA 5005, Australia. {damith, cole}@eleceng.adelaide.edu.au
[2] Auto-ID Labs, Institute for Manufacturing, Cambridge University
Engineering Department, Mill Lane, Cambridge, CB2 1RX, United Kingdom
{mark.harrison}@cantab.net

Abstract. The concept of a "Networked Physical World" originated from the Auto-ID Center, now called the Auto-ID Labs [1]. Such a "World" can be realised with the combination of an automatic identification technology and a ubiquitous computer network that will glue the physical world together. Low cost RFID (Radio Frequency Identification) technology can automate identification of physical objects by providing an interface to link a vast number of objects to the digital domain. Thus, RFID as the enabling technology has paved the way forward for the creation of a "Networked Physical World". The ability to form a ubiquitous item identification network has a wide range of applications including automation of manufacturing and supply chain management. The previous chapter provided a brief overview of RFID systems. This chapter describes the backend system components formulating a distributed ubiquitous item identification network enabled by the development of automatic identification provided by RFID technology, and examines the flow of tag data, once obtained by an interrogator. The implementation of such an architecture using a web services based model, as well as the impact of the network on supply chain applications, is also investigated.

Keywords: RFID, EPC, EPCIS, ONS.

1 Introduction

The roots of the architecture to build a ubiquitous item identification network originated at the former Auto-ID Center, now called the Auto-ID Labs [1]. Further development of that network and the process of standardization of issues related to that network, are currently managed by a number of working groups at EPCglobal Inc [2].

The Auto-ID Center vision was to create a "Smart World" by building an intelligent infrastructure linking objects, information, and people through a ubiquitous computer network. The creation of the intelligent infrastructure demanded the

ability to identify objects automatically and uniquely with the backbone of infrastructure provided by a ubiquitous computing system leveraging the Internet for global connectivity. The components forming the intelligent infrastructure are commonly referred to as an EPC Network where the term EPC (Electronic Product Code) is the unique object identification scheme employed by the system. This new infrastructure enables object-centric computing that will allow universal coordination of physical resources through remote monitoring and control by both humans and machines.

The availability of a Networked Physical World system connecting physical objects to the Internet will have an immediate and profound impact on supply chain management, home automation and manufacturing automation. However, such infrastructure needs to be cost effective to allow its feasibility and large scale adoption.

The EPC Network is assembled from many building blocks representing a number of fundamental technologies and standards. The EPC Network architecture has continued to evolve with technology since its inception. Current developments in the EPC Network are based on a layered service oriented architecture [3].

2 N-tier Layered Service Oriented Architecture

Contrary to the component based EPC Network architecture developed initially by the Auto-ID Center, the more modern version is based on a layered service oriented architecture [4] with an emphasis on defining interfaces. The interfaces define the required standard functionalities and optional functionalities as well as the methods by which these can be accessed, rather than defining components and their associated functionalities. The benefits of such architecture are many; primarily the modularity of the software and the standardized interfaces allows various components to be purchased from different suppliers. Thus, in addition to complexity reduced by modularity, there is also a strong business case for supporting the current architecture of the EPC network. The definition of standard interfaces will ease the process of compliance certification with standards and allow the definition of data models and methods at an abstract level, usually using UML (Unified Modeling Language) diagrams. Such interfaces can be easily described in XML (Extensible Markup Language) or WSDL (Web Services Description Language) so that concrete implementations of these can be both cross-platform and neutral regarding choice of programming language or implementation details (such as whether a relations database, XML database or object-centric database is used).

The N-tier service oriented architecture approach [4] fits naturally with an object oriented modeling of the architecture because objects encapsulate information and state while offering functionalities through their interfaces. The modules also have a loose coupling due to the independence of different modules. This reduction in dependency implies that the system is easier to manage and enhance.

The EPC Network architecture is built around a loosely coupled and interoperable system of modules. Many of the concepts for this architecture come from the conceptual architecture called service-oriented architecture (SOA) [4].

Web services are one method of implementing the SOA over standardized protocols and interfaces. There is a strong tendency and a technological trend driving the EPC network architecture towards a web services based SOA.

The following section presents the system architecture and technologies of the ubiquitous item identification network (EPC Network) and provides a seamless architecture for extending the local area EPC Networks to a wide area infrastructure while the later sections include an application of such a network. The chapter is concluded with an investigation of the impact of the EPC Network on its most revolutionary application; supply chain management.

3 EPC Network

The EPC Network can be described as an intelligent ubiquitous infrastructure that automatically and seamlessly links physical objects to the global Internet. The network of physical objects is achieved by integrating a tag, or an RFID tag, into each object. The system networks objects seamlessly by communicating with these tags using interrogators at suitably placed locations, for example: RFID portals; handheld readers; and potentially, eventually for some tags, continuously throughout the environment by a network of readers. Readers collect data from tagged objects. The RFID tagged objects communicate an EPC code to a reader and thus identify themselves as a unique entity. The data originating from the network of readers is passed to backend systems that control and collect data while providing service layer functionalities.

An illustration of the components constituting the EPC Network is shown in Figure 1 where the arrows indicate the flow of data from tags to the network support system and the flow of command and data back to the readers and tags.

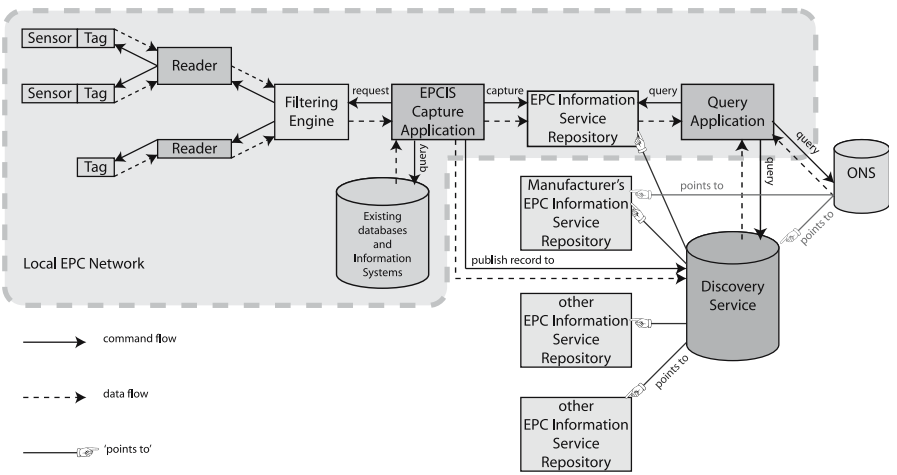


Fig. 1 An Overview of the EPC Network. The part that is local to a single organization is shown within the shaded region with the dashed border.

EPC Networks are significantly different from more traditional computer networks in the sense that the flow of data and information is from many nodes (RFID tags) at the edge of the network towards a number of servers that collect and process this data. In RFID networks, readers detect certain events or readers query RFID tags to obtain event data and forward the resulting information to backend applications or servers. The application systems then respond to these events and application processes orchestrate corresponding actions such as ordering additional products, sending theft alerts, raising alarms regarding harmful chemicals or replacing fragile components before failure.

The EPC Network can be separated into six primary modules, some physical, some logical: (1) RFID tags (also known as 'labels' or 'inlays'), (2) RFID tag readers (also known as interrogators), (3) EPC, (4) the filtering middleware that supports the Application Level Events (ALE) interface, (5) Object Name Service (ONS), and (6) EPC Information Service (EPCIS). Figure 2 shows the structure of a typical EPC Network where readers communicate with tags and then capture that information to be passed up the information chain to be processed by the service layers and made available for various applications.

The EPC Network shown in Figure 1 is a local area EPC Network akin to a LAN. This model captures the architecture of the system at a local site, company or organization, or a private network. Although data collected within an organization from a local EPC network may be centralized towards a local EPCIS of that company, the architecture is designed to support decentralized distributed information management.

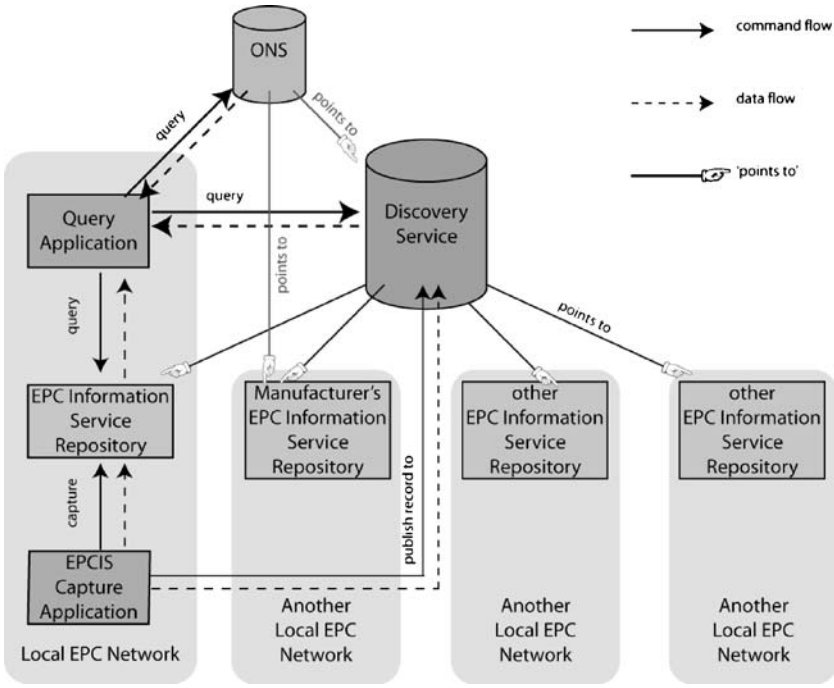


Fig. 2 An overview of a wide area EPC Network.

The EPC Network architecture achieves the latter goal by linking local EPC Networks together through the already well established backbone of the Internet. The resulting wide area EPC Network achieves a global flow of information and data, while at the same time extending the reach and the usefulness of the network. The wide area network architecture enables the exchange of information between organizations at the level of individually identifiable unique objects and supports the extraction of, for instance, 'lifecycle' information for an object which may be fragmented across multiple organizations rather than one single centralized database. Figure 2 illustrates such architecture where a global public ONS system together with Discovery Services may be used to connect public local area EPC Networks. The subject of ONS and Discovery Services are considered in Section 6.

There are ongoing collaborative efforts to standardize the interfaces linking the modules outlined in Figure 2 by the various actions groups of EPCglobal, and the Auto-ID labs, to achieve interoperability and to allow hardware and software vendors to be able to compete in a fair and open market in the supply of technology and equipment to establish EPC Networks. Figure 3 identifies the interface layers being standardized by EPCglobal [2]. At the time of compiling this chapter all of the interfaces identified in Figure 3 have been developed and ratified as standards by EPCglobal, with the exception of standardized interfaces to Discovery Services. All ratified EPCglobal standards are freely available in electronic format from [2].

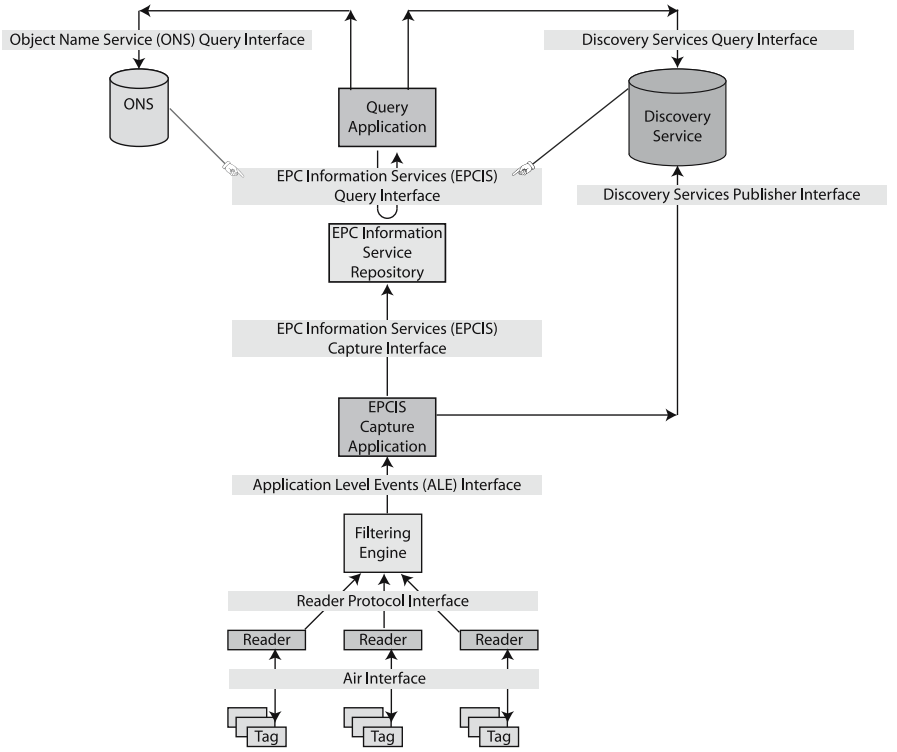


Fig. 3 EPC Network architecture showing standardized interfaces.

4 RFID Components

The RFID aspects of the EPC Network consist of RFID tags, readers and the unique identifier (EPC). These modules of the network have been discussed in Chapter 3. The following is an overview of the RFID components of the EPC Network while a more complete and thorough coverage of the topic may be obtained from Chapter 3.

RFID tags, when coupled to a reader network, form the link between physical objects and the virtual world in the EPC Network. RFID tags have a small radio antenna that transmits information over a short range to an RFID tag reader [5, 6]. RFID technology may use both powered and non-powered means to activate the electronic tags. Powered devices use batteries to actively transmit data from the tags to more distant readers. Electronic highway toll systems are good examples of active RFID tags. Passive RFID devices literally harvest energy from the electromagnetic field of an active reader to both power the tag and transmit the data and therefore do not require any additional built-in power source, such as a battery. Modern RFID tags are fabricated using standard CMOS technology and are interrogated by the process of RF backscatter [5, 6, 7]. In the most cost effective and popular technology, the tags are passive and in consequence the ranges of operation are limited (few meters) [7, 8]. Passive systems are well suited for use in the EPC Network due to their low cost.

The architecture of the network does not place a restriction on the tags that can be employed, as tags with substantially enhanced functionality will extend the depth of the application layer. Such functionality can be provided with active RFID tags. The most common objective of an active RFID tag is to obtain a long read range using battery-assisted backscatter. An active backscattering tag will modulate the powering carrier wave or a subcarrier to establish a communication link with the reader while using the battery to power the logic circuits of the tag [5, 7].

However other types of active tags, the independent reply generating tags, may not use backscatter but instead use a battery (such as a paper battery) for powering and transmitting requirements. This distinction is more apparent in the range of operation of the tag. Under RFID systems operating under the US regulations for the ISM (Industrial, Scientific and Medical) band of 902–926 MHz (allowed transmit power in this band is 4W EIRP), a backscattered reply can be correctly decoded in the range of several tens of meters while an independent reply generating tag will work in the range of several hundred meters.

A simple illustration of the concept of a Radio Frequency Identification (RFID) system used in the EPC Network is shown in Figure 4. Currently the dominating air interface protocol used in consumer goods packaging supply chain applications is the Class I Generation 2 air interface protocol ratified by EPCglobal. It has been accepted an ISO standard: (ISO 18000-6 Type C) along with the existing set of RFID air interface protocols defined in the ISO standards (ISO 18000).

Communication between a reader and a tag (via a radiofrequency interface) may involve interrogating the tag to obtain data, writing data to the tag or transmitting commands to the tag so as to affect its behavior. The readers possess their own source of power, processing capability and an antenna. A ubiquitous reader network will allows continuous tracking and identification of physical objects.

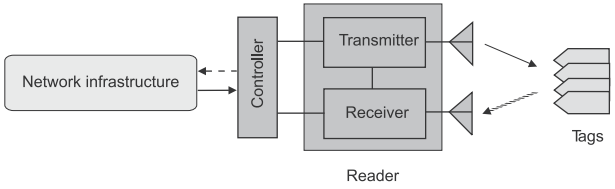


Fig. 4 Overview of an RFID system. The transmitter, receiver, and the controller form the RFID reader. Here a transmitter of interrogation signals which is contained within a reader communicates via electromagnetic waves with an electronically coded tag to elicit from the tag a reply signal containing useful data characteristic of the object to which the tag is attached. The reply signal is detected by a receiver in the interrogator and made available to a control system.

Antenna arrays can be fabricated and integrated in floor tiles, carpeting, shelf structures, cabinets and appliances. Similarly to cellular phone grids, the reader network may provide seamless and continuous communication to RFID tags. A data collection and control system must support the reader network to enable efficient use of the continuous, or at least very frequent, object communications. Additionally, in order to access and identify these objects, a scheme is required to uniquely name and identify objects.

The unique object identifier must have a global scope that is capable of identifying all objects uniquely and act as a pointer to information stored about the object and the functionalities of the tag somewhere over the network. The Electronic Product Code (EPC) is a scheme designed for universal object identification with the associated standards developed by EPCglobal Inc. A binary representation of the EPC is shown in Figure 5 [9].

The unique identifier format defined as the EPC is a flexible data structure, which already supports embedding of a number of legacy identifier schemes already in widespread use today. These include the following GS1 identifiers: Serialized Global Trade Item Number (GTIN), Serial Shipping Container Code (SSCC), Global Location Number (GLN), Global Reusable Asset Identifiers (GRAI) and Global Individual Asset Identifiers (GIAI), together with identifier schemes developed by the US DoD for logistics purposes. However, the flexible framework provided by the EPC permits organizations to integrate their own standards based numbering format.

The EPC Tag Data Standard (TDS) [9] defines the structure of the EPC for each identifier scheme and provides the encoding/decoding rules, while EPC Tag Data Translation Standard (TDTS) [23] provides this in a machine-readable format, to support automated translation and validation tools at any stage in the EPC Network architecture shown in Figure 3. The format provided in Figure 5 is only one possibility that might be suitable for a sensor network.

MSB LSB

Header	General Manager number	Object Class	Serial number

Fig. 5 Bit level representation of an EPC general identifier (GID) format.

4.1 Extending RFID to Sensing Applications

The mechanism used by RFID readers to obtain information stored in the EEPROM (Electrically Erasable Programmable Read-Only Memory) memory of an RFID tag can be applied directly for the collection of data obtained by sensors. This will require extending a simple RFID chip interface to log sensor derived data [10, 11] in its memory.

However there are a number of engineering challenges that need to be overcome. Passive RFID systems do not have an onboard power source and thus the sensors on passive tags can not operate while the tag is outside the range of a reader. Passive RFID tags are also power constrained systems that function by deriving their power from reader interrogation signals, thus the available energy for a sensor is limited and is dependent upon the proximity to the reader.

Nevertheless this is not a serious impediment to the development efforts as active RFID tags provide a suitable and cost effective alternative. In addition it is possible for a reader network to frequently power a distributed network of RFID sensors or to power when required, to regularly obtain sensor derived data. However in the future energy scavenging systems on board passive RFID tags may power the sensors and provide sufficient power to store sensor derived data in a EEPROM memory [12, 13].

5 Filtering middleware and Application Level Events (ALE)

Filtering middleware layer provides real time processing of RFID tag event data. The EPCglobal Application Level Events (ALE) standard [14] provides a standard mechanism for applications to request filtered event data. Conceptually filtering middleware occupies, as shown in Figure 3, the space between a reader (or multiple readers) and event repositories such as an EPCIS repository. Networked filtering systems form a framework to manage and react to events generated by tag reads by interrogators.

The filtering middleware receives requests from applications, processes data from the reader(s) and returns unique tag identifiers, and possibly other data from sensors (although processing of sensor data is not part of the ALE standard interface version 1.0), either back to the requesting applications or to other systems specified in the requests. The middleware reports EPC data as ‘events’. The events reported through the ALE interface are statements of ‘what’ EPC(s) was seen, ‘where’ and ‘when’.

The filtering middleware has several fundamental functions integrated into its design, some of which are data filtering of received tag (and possibly sensor data), aggregation and counting of tag data and temporary accumulation of data over time periods, although long-term retention of events is not expected in filtering middleware. These fundamental functions are required to handle the potentially large quantities of data that RFID systems are capable of generating through continuous

interrogation of tags. For instance, the ALE interface enables local applications to state the significance of specific data obtained from RFID tags (for example a record of EPCs observed or temperature variations over a time period) and to report accumulated data using a standard format defined by an XML schema (an existing XML schema definition can be found in [14]). The filtering middleware framework may also implement an application specific XML schema (such as that more suited towards a specific sensor application) or a number of such schemas to allow the capture and reporting of physical world events and measurements.

The filtering middleware implements the ALE interface, through which client applications communicate, and uses the reader protocol interface to communicate with readers. Filtering middleware may be composed of multiple filtering middleware services, each with their own functionality. The services can be visualized as modules in the filtering middleware. These modules can be combined to perform certain functions for specific applications. Hence one or more applications may make method calls to the filtering middleware resulting in an operation being performed (collection and return of temperature readings from a sensor) and the return of results. Other than filtering middleware services interacting with each other to perform certain tasks, filtering middleware services can also interact with services such as the EPC Information Service (EPCIS) to provide services for the framework of global applications (the EPCIS will be considered in detail in Section 7). Figure 6 shows a conceptual architecture of the filtering middleware system.

Event management is a primary service provided by filtering middleware services. A common event management function is filtering, which is particularly useful in situations where there is heavy data traffic. For example, readers may read data coming in from multiple RFID tags repeatedly. Not all the data from all the tags may be of interest to an application. Filtering of that data can eliminate information that is either redundant (multiple reads of the same data), or that is not required (tags read but not of interest to that application), from reaching an application. Furthermore, ALE events can be constructed from differential sets of

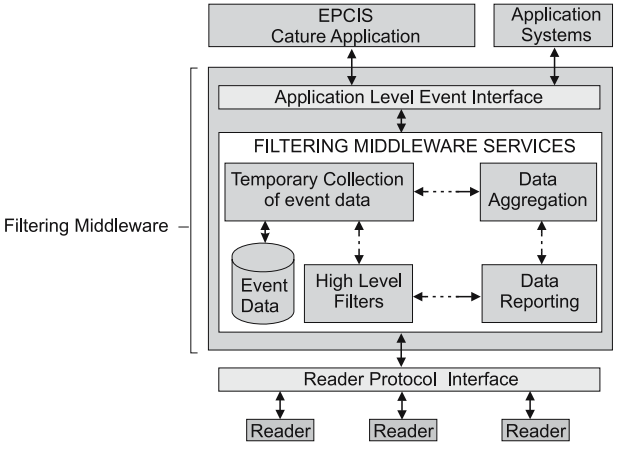


Fig. 6 Architecture of a filtering middleware and its interaction with EPC Network components, EPCIS and readers.

IDs, generating an event when a particular ID is first seen (arrival) or when it is no longer seen after a period of time (departure).

5.1 EPC Data Encapsulation and Reporting

The EPC serves as a reference to information. However, the storage, transport and description of that information requires a structured and universal vessel that can be easily understood, stored and transported across the Internet. Previously the Auto-ID Center defined the Physical Mark-up Language (PML) [21] to encode captured object information. However, recent developments have retreated from such a rigidly defined schema to the characterization of two instances: `ECSpec` (Event Cycle Specification) and `ECReports` instances using a standard XML depiction [14]. Thus requests to the filtering middleware are sent as `ECSpec` objects while data from the filtering middleware is returned as an `ECReports` objects. The event data reported may be filtered according to a client's filtering specification, for example by company ID, object type or logistic unit (such as item, case or pallet).

The core XML schemas for these objects are defined with extensions and rules to accommodate application or manufacture specific XML schema (such as that suited for a specific sensor application) or a number of such schemas to allow the capture and reporting of physical world events and measurements.

XML schema defined do not interpret the data that they handle nor do they promise a universal means for encoding structured information. The XML schema definition is rigid, simple and all the elements can be understood easily because of the use of long descriptive tag names which increase human readability and help to avoid mistakes in the interpretation and the understanding of data, and how that data is to be handled.

6 Object Name Service

The functionality provided by the ONS system is similar to the services provided by the Domain Name System (DNS); however instead of translating host names to their underlying IP addresses for user applications, ONS translates an EPC into URL(s). The Object Name Service (ONS) in an EPC Network identifies a list of authoritative service endpoints associated with the EPC and does not contain actual data related to an EPC. By authoritative, we mean that the entity that retains control over the information about the EPC placed on the ONS is the same entity that assigned the EPC to the item (this implies that the ONS points to the manufacturer's information services). It should be noted that ONS version 1.0 does not provide different records for each individual EPC but rather, records for an object class, product type or SKU (Stock Keeping Unit). ONS normally does not provide dynamic links to multiple providers of information across the supply chain or lifecycle of an individual object; that is the role of Discovery Services –

although to date, EPCglobal have not yet chartered a work group to define standard interfaces for Discovery Services. These service endpoints can then be accessed over a network [15].

The ONS functions like a “reverse phone directory” since the ONS uses a number (EPC) to retrieve the location of EPC data from its databases. The ONS is based on existing DNS systems and thus queries to, and responses from, ONS adhere to those specified in the DNS standards (RFC 1034: Domain names, Concepts and facilities). In fact, ONS uses a particular type of DNS record, called Naming Authority Pointer (NAPTR) records, (defined in IETF RFC 2915), to provide for future flexibility, since NAPTR records support the use of regular expression pattern matching; although this is not currently used in ONS, it means that the ONS results can in future be interpreted as a pattern match, resulting in a URL address which contains the serial number as part of the URL, without needing to add an ONS record for each serial number of a particular product type. This can be observed in the ONS resolution process outlined in Table 1. However, unlike the DNS, ONS is authoritative.

Figure 7 shows the overview of an ONS system where an EPC encoded in an RFID tag is read by an RFID reader, where obtaining information associated with the EPC involves the resolution of the EPC through a query of the local ONS server to obtain a location of an appropriate application layer service (such as that provided by an EPCIS). In the event that the local ONS server is unable to satisfy the requests it is forwarded to a global ONS server infrastructure for resolution (the details of an EPC resolution process are outlined in Table 1)

The root of the ONS server infrastructure is the ONS Root providing a query interface for a GTIN using a GS1 company prefix. EPCglobal Inc. is responsible for the administration of the Root ONS, although they have subcontracted the operations to Verisign Corp.

The ONS is currently implemented using DNS technology. However, unlike the familiar use of DNS to translate a domain name or hostname into an IP address (and vice versa), the role of ONS is to return one or more authoritative URL addresses of services for a given EPC class, when queried with a hostname derived from that EPC class. The URL address may directly indicate the server providing a service, such as a webserver providing product-specific web pages.

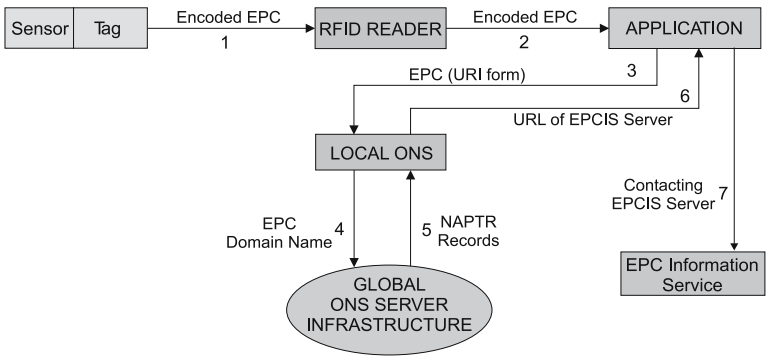


Fig. 7 An overview of an ONS system functionality.

Table 1. Outlines a description of the object name resolution process illustrated in Figure 8.

	Description
1	A reader interrogates a tag and obtains the EPC in binary form. Example: Assuming that the EPC is of a 96 bit general identity type [7] and the following binary string is returned. This bit string represents an EPC encoded in the form identified in Figure 5 where the first 8 bits form the header, the next 28 bits the general manager number, followed by a 24 bit object class and the last 36 bits are the serial number. (`00110101 000000000000000000000010 000000000000000000011000 000000000000000000000000110010000`)
2	The EPC obtained (as a binary number) is passed to the local network application processes. Example: (`00110101 000000000000000000000010 000000000000000000011000 000000000000000000000000110010000`)
3	The EPC is then converted into URI form to provide a means by which application software is able to manipulate the EPC codes independent of any tag level encoding scheme, thus allowing software systems to treat the EPC data in a uniform manner, irrespective of how they are formed or obtained. All URIs are represented as Uniform Reference Names (URNs) using the standard format defined in RFC2141 using the URN Namespace epc. Example: For an EPC general identifier the URI representation is as follows: `urn:epc:id.gid.GeneralManagerNumber.ObjectClass.SerialNumber` In the above representation the *GeneralManagerNumber, ObjectClass* and the *SerialNumber* refer to the fields identified in Figure 5 [7]. [`urn:epc:id:gid:2.24.400`]
4	URI is converted into domain name form so that a query in the form of a DNS query for a NAPTR record for that domain can be issued. In the event that the local ONS can not respond, the request is sent to the global ONS server infrastructure. Remove `urn:epc` Example: [`id:gid:2.24.400`] Remove serial number, since resolution down to the serial number level is not undertaken by the ONS and the resolution process stops at the Object Class level. This is a practical implementation as resolution to such granular level adds both complexities, cost and raises questions regarding the scalability of the architecture. Example: [id:gid:2.24] Invert the order of the fields of the string, replacing ‘:’ with ‘.’ Example: [`24.2.gid.id`] Append “`.onsepc.com`” Example: [`24.2.gid.id.onsepc.com`]
5	Perform a host lookup for type 35 / NAPTR records. Example: [`host –t NAPTR 24.2.gid.id.onsepc.com`] The ONS server infrastructure will return a set of possible URLs that point to one or more services that are provided for the *ObjectClass* and *GeneralManagerNumber* in the ONS query.

Example:
In the example the list of URLs returned relate to the *GeneralManagerNumber* 2 and the *ObjectClass* 24.
[`http://bar.com/epcis.php`;
`http://advark.com/sensor_is.asp`;
`http://foo.com/epc_is.wsdl`]

6 The correct URL is picked and extracted from NAPTR records by the application depending on the application type and need; the ONS records include a service type field to assist applications with this selection.
Example NAPTR record:

Order	Pref	Flags	Service	Regexp	Replacement
`0`	`0`	`u`	`EPC+epcis`	`!^.*$!http://www.foo.com/epc_is.wsdl!`	`.`

7 The application system contacts the desired service.
Example:
In the example the application system can examine the wsdl file to obtain the desired service and then using the information provided therein contact that service.
[`http://www.foo.com/simpleEventQuery`] – Returns a list of `EPCISEvent` instances [22].

Alternatively, the URL may indicate the address of a web service description file (e.g. a WSDL file), which in turn contains further redirections to specific service end-points, e.g. for an EPCIS web service. At no point is the ONS used to find IP addresses; ordinary DNS is used to resolve any domain names appearing within the service address URLs returned by ONS.

A challenging aspect of the resolution process is the ability to select the required URL since a list of URLs corresponding to a particular EPC may be returned by the ONS server (as shown in step 5). The format of the choices

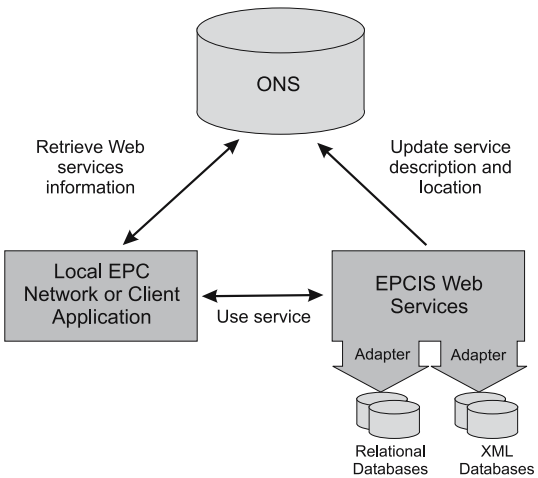


Fig. 8 Interaction between EPCIS, ONS and external applications in a based web services based implementation of the EPCIS.

returned by ONS is defined in the Naming Authority Pointer (NAPTR) record. The complete definition of NAPTR can be found in IETF RFC 2915. In essence, NAPTR is a collection of information that points to locations on the World Wide Web when only a URI is provided. The NAPTR is formatted as shown below.

Order	Pref	Flags	Service	Regexp	Replacement

In a NAPTR, the URL is located in the [Regexp] field while [Order], [Pref] (Preference), and [Flags] are used to state the preference order of a list of URLs. [Service] is used to specify the type of service that is offered at each URI, such as an EPCIS. The [Replacement] field is not used for EPC Network purposes while the [Flag] field is set to 'u' to indicate that the [Regexp] field contains a URI [15].

It should be noted here that the ONS does not resolve queries down to the level of fully serialized EPCs. The depth of the query stops at the Object Class level (product type) of the EPC. Thus queries directed at the serial number level must be resolved by the service locations obtained from the ONS query. A proposal for such services, called EPC Discovery Services, is outlined and discussed in Section 7. Although Discovery Services are outlined as part of the Network Architecture Framework [3], no formal work group has been chartered to develop standardized interfaces to Discovery Services.

7 EPC Information Services

EPC Information Services (EPCIS) are the gateways between any requester of information and the databases containing that information. They primarily respond to queries from authorized entities that are expressed in a standard format; however the internal storage of that data within the databases may be in any format or standard. The EPCIS is the "interpreter" communicating between database(s) and application(s) and provides a standardized interface to the rest of the EPC Network for accessing EPC related information and transactions. However, it is also possible to have an EPCIS implementation without any repository where the services offered by the EPCIS offer streams of current events without storing them.

EPCIS interfaces are defined in [22] and can be implemented using web services technology. Web service interfaces allow applications in the wider area network to utilize services provided by local EPC Information Services using a remote method invocation (RMI) – also known as remote procedure call (RPC) – paradigm (refer to Figure 8). Such an architecture has the advantage of enabling communications between different operating systems and also between systems written in different programming languages because it leverages standardized XML messaging frameworks, such as that provided by SOAP (Simple Object Access Protocol). A description of the available services is defined in terms of a WSDL (Web Services Description Language) file. Hence an application requiring information is able to access a WSDL file which has a description of the available service methods, the required input and output parameters to the methods and information to invoke those methods.

EPCIS provides a model for the integration of RFID networks across the globe.

However it is important that EPCIS provides a secure communication layer so that local EPC Networks can retain the ability to control access to information. WS-Security [16] is a candidate proposal for enhancing web services security that describes enhancements to SOAP messaging to provide message integrity and message confidentiality while proposed architectural extensions to the existing WS-Security profile [17] could provide access control as well as a federated security model for EPCIS.

Information about a particular EPC may be spread across a number of local networks (in the event of a supply chain application an object will pass through a number of physical locations, for instance manufacturers, distributors and retailers). The ONS does not resolve to the serial number level of the EPC and the DNS technology upon which the ONS is based also is not suited to providing the fine grained resolution down to serial number levels. Resolution down to serial EPC level (to a specific object) is handled by the EPC Discovery Service (EPC-DS).

EPC-DS is best described as a "search engine" for EPC related data [18]. EPC-DS provides a method for custodians of a particular RFID tag data to update the EPC-DS to indicate that that they hold data related to an EPC. The register may contain a list of EPCIS URLs where such information may be obtained [18, 22]. However unlike the ONS, the Discovery Service provides links to non-authoritative providers of information about an EPC, including other organizations that have handled it during its lifecycle or supply chain.

Discovery Services is an addition to ONS, since it is Discovery Services that will provide the serial-level links to multiple local area EPC networks (or more specifically to multiple EPCIS addresses of companies that provide more detailed information about an individual object). The root ONS primarily points just to the manufacturer's information service and possibly to a Discovery Service for that EPC.

8 An EPC Network Application

Figure 9 shows an example of a simple supply chain model. The manufacturer is linked to raw material suppliers, production factories, distributors and some direct

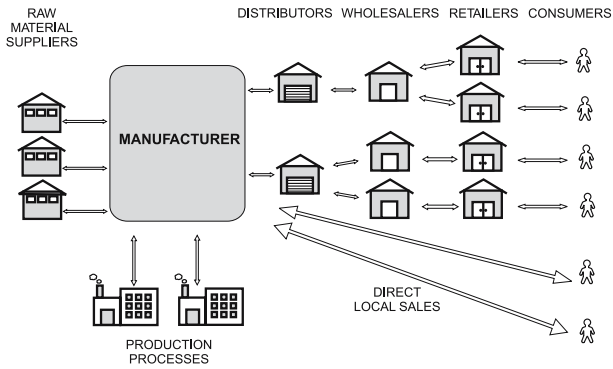


Fig. 9 A simple supply chain model.

sales customers. The flow of goods is from the manufacturer to distributors and then to wholesalers. The goods then flow from wholesalers to retailers, and then to consumers. In reality, multiple linkages or relationships are present, and the distribution system is much more complex than that shown in Figure 9. However, the simple model in Figure 9 is adequate to demonstrate a use case scenario.

Each party in this simple model, except for the consumers, is linked to the flow of goods through a wide area EPC network infrastructure. In a supply chain utilizing RFID technology all transactions including individual consumer purchases can be automated. In such an application the architecture of the EPC Network at each party is illustrated in Figure 3.

8.1 Supply Chain Management

RFID technology allows transactions to be automatically recorded and inventory levels to be updated in real time. When a customer purchases a product from a retailer, the RFID tag attached to the product will be scanned. The local system of the retailer will be informed of the current status of the product (i.e. product sold). If the stock level of the product has reached a critical level, business logic systems at the local system can automatically send a purchase order to the wholesaler. The wholesaler, upon receiving the order request from that retailer, will automatically check through its inventory for stock availability. If the stock level is sufficient, the goods will then be dispatched to the retailer. If the stock level is low, either before or after the goods are dispatched, the wholesaler can place an order to the distributor automatically. A similar process will occur between the manufacturer and the distributor and the manufacturer and its raw material suppliers. However, it should be noted here that the EPC network services do not allow an external party to influence the collection of data without a prior contractual agreement.

The real time visibility into the supply chain allows the dynamic transfer of supply and demand information along the supply chain and thus prevent what is termed the “bull whip effect” with the consequence of considerable savings to businesses and improved services to consumers [19]. Visibility provided by RFID technology allows the organizations along the supply chain to adjust rapidly to meet market conditions without the burden of large inventories to meet a forecast demand that may or may not occur.

Automating inventory control and smart supermarket shelves [20] will also ensure product availability and dramatically reduce customers lost to businesses through product unavailability.

The EPC Network provides the ability to capture an instance of the supply chain in real time. Enterprise applications exploiting knowledge-based architectures can utilize the events stream collected from an EPC network to adjust business processes to minimize cost and increase efficiency while perhaps notifying managers when critical events occur.

Below are a few scenarios depicting how the EPC Network can enhance and bring about further improvements to supply chain logistics.

8.2 Solutions to Grey-Market Activity and Counterfeiting

Products falling into grey-markets have been one of the main issues that are of much concern in some industries, such as the pharmaceutical industry. One scenario where grey-market goods appear is in the re-importation into a particular country of products which were previously exported from that country, at prices far below those ruling in the domestic market of the exporting nation. These export prices may be below the domestic prices for various reasons, including subsidies. Some operators illegally import products back to the country of origin, where substantial profits can be made by selling the products at the higher domestic price.

Product authenticity is also of concern to consumers. For example, in the pharmaceutical industry, imitations of most high value drugs are illegally manufactured by various manufacturers. Many of these products have the same potency as the authentic product, but some provide no benefit, or worse, cause harm. The ability of the EPC network to track and trace a product throughout the supply chain can prevent the entry of illegal counterfeits.

Figure 10 shows one possible data flow in an EPC enterprise system to fight grey-market distribution. When a product is sold and the EPC is recorded at the point of sale, the local EPC Network will send a request to the Discovery Services of the manufacturer to establish the existence of a valid chain of custody (step 1 in Figure 10) by verifying the trace history of the product (path given by 2-3-4 in Figure 10). If a valid supply path exists for that product, the product is validated. Grey-market goods will not pass this examination and an alert can be sent to the appropriate authority in such a scenario.

For product authentication, a party such as the retailer will only need to send a request to the manufacturer. The manufacturer can check the EPC received with its database. The manufacturer will be able to validate that the product was indeed manufactured and is currently available at a specific location in the supply chain. The retailer will also be able to check for multiple sale records to ensure that the item tag was not obtained from a previously sold item. Thus the ability to track and trace provides an electronic history of the item throughout its life cycle through the supply chain and thus helps prevent counterfeiting and the re-sale of

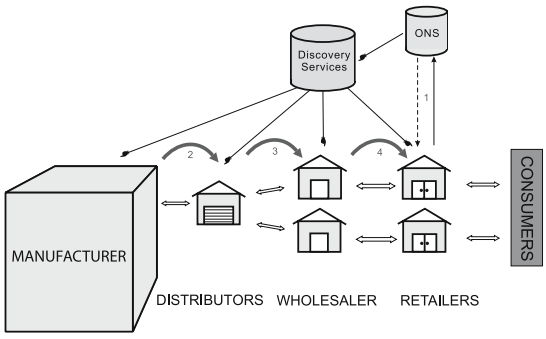


Fig. 10 Counterfeit goods detection.

goods in grey markets. The term “electronic pedigree” has been coined to refer to the electronic history of an item’s life throughout the supply chain. An electronic pedigree usually also involves digitally signed records confirming changes of custody and/or transfer of ownership at each handover stage in the supply chain.

However such a simple solution may not be sufficient when dealing with counterfeit tags introduced into the supply chain and other appropriate solutions need to be contemplated. Such issues related to security are considered in more detail in the Networked Based Solutions and Cryptographic Solutions Sections of this book.

8.3 Product Recall and Other improvements

Apart from the obvious advantages of supply chain visibility and the reduction in labor needed to check and scan shipments to confirm that the correct goods were received, there are many other advantages arising from the use of EPC Networks. These include, amongst many others, easy and efficient product recall and the convenient identification of returned or rejected goods, at all levels of the supply chain, from retailer back to manufacturer.

The most significant of the above applications is product recall. It is possible to flag each EPC of a product destined for recall, and such products reaching each location of the supply chain can be detected and removed from the supply chain even before they reach the retailers.

9 Conclusion

This chapter has introduced and elaborated on the technology and on the concepts of the EPC Network. An application of the network to supply chain management was also illustrated with a brief outline of how the services provided can be implemented, leveraging the existing technologies provided by web services tools and standards.

A set of version 1.0 standards has now been ratified by EPCglobal covering the majority of the architecture (Reader Protocol, Reader Management, Application Level Events, EPC Information Services, Object Name Service). Extensions and updated versions are also under development. The functionality of the EPC Network can be summarized as providing the linkages between all physical objects with RFID tags, the management of the vast volume of data generated by readers and the provision of a universal query system for accessing and sharing information that describes objects over the Internet for access by remote services.

References

1 Auto-ID Labs. Available from: http://www.autoidlabs.org
2 EPCglobal Inc. Available from: http://www.epcglobalinc.org
3 EPCglobal Inc.: Network Frame Architecture (2005). Available from: http://www.epcglobalinc.org/standards
4 Erl, T.: Service-Oriented Architecture: Concepts, Technology, and Design. Prentice Hall, (2004)
5 Finkenzeller, K.: RFID Handbook: Radio Frequency Identification Fundamentals and Applications. John Wiley & Sons, New York (1999)
6 Cole, P.H.: Coupling and quality factors in RFID. In: Design, Characterisation and Packaging for MEMS and Microelectronics. In: Paul D.F. (eds.): Proceedings of SPIE, Vol. 4593 (2001) 1−11
7 Cole, P.H., Hall, D. M.: Integral backscattering transponders for low cost rf id applications. In : Fourth Annual Wireless Symposium and Exhibition, Santa Clara (1996) 328−336
8 Cole, P. H., Hall, D. M., Loukine, M., Werner, C. D.: Fundamental Constraints on rf Tagging Systems. In: Third Annual Wireless Symposium and Exhibition, Santa Clara, (1995) 294−303
9 EPCglobal Inc: EPC Tag Data Standard Version 1.3 Rev. 1.3 March (2006). Available from: http://www.epcglobalinc.org/standards/ (05/2007)
10 Cernosek, R. W., Chin, B. A., Barbaree, J. M., Vodyanoy, V., Conner, D. E., Hsieh, Y-H. P.: A Rapid Biosensing System for Detecting Food-Borne Pathogens. In: Proceeding of the Sensors Expo (2001) 113−116
11 Wentworth, C.: Radio Frequency Identification Sensors. In: 7th World Multiconference on Systemics, Cybernetics and Informatics (2003)
12 William, C. B., Yates, R. B.: Analysis of a micro-electric generator for Microsystems. In: Transducers '95/Eurosensors IX (1995) 369−372
13 Glynne-Jones, P., Beeby, S. P., White, N. M.: Towards a piezoelectric vibration-powered microgenerator. In: IEEE Proceedings of Science, Measurement and Technology, Vol. 148(2) (2001) 68−72
14 EPCglobal Inc.: ALE specification Version 1.0, Sept. (2005). Available from: http://www.epcglobalinc.org/standards/ (05/2007)
15 EPCglobal Inc.: Object Name Service (ONS) v1.0, Oct. (2005). Available from: http://www.epcglobalinc.org/standards/ (05/2007)
16 Atkinson, B., et al.: Web Services Security (WS-Security) Specification version 1.0.05 (2005). Available from:http://www-128.ibm.com/developerworks/ webservices/library/ws-secure/, (06.2005)
17 IBM Corporation and Microsoft Corporation: Security in a Web Services World: A proposed Architecture and Roadmap. In: White paper (2002). Available from: http://msdn.microsoft.com/library/default.asp?url = /library/en-us/dnwssecur/html/securitywhitepaper.asp (04.2005)
18 Harrison, M.: EPC Information Service (EPCIS). In: Auto−ID Labs Workshop, Zurich (2004)
19 Lee, H. L, Padmanabhan, V., Whang, S.: Information distortion in a supply chain: the bullwhip effect. In: Management Science, Vol. 43(4) (1997) 546−558
20 Chappell, G., Durdan,D., Gilbert, G., Ginsburg, L., Smith J., Tobolski, J.: Auto-ID in the Box: The value of Auto-ID Technology in Retail Stores. In: Auto-ID Centre White Paper (2005). Available from:http://www.autoidlabs.org/whitepapers/ acn-autoid-bc006.pdf (01.2006)

21 Floerkemeier, C., Anarkat, D., Osinski, T., Harrison, M.: PML core specification 1.0. In: Technical report, Auto-ID Center, Sept. (2003)
22 EPCglobal Inc.: EPCIS specification Version 1.0, April (2007). Available from: http://www.epcglobalinc.org/standards/ (05/2007)
23 EPCglobal Inc: EPC Tag Data Translation Standard Version 1.0, Jan. (2006). Available from: http://www.epcglobalinc.org/standards/ (05/2007)

Chapter 5
A Security Primer

Manfred Jantscher[1], Raja Ghosal[1], Alfio Grasso[1], and Peter H. Cole[1]

[1] The School of Electrical and Electronic Engineering,The University of Adelaide, Adelaide SA 5005, Australia.{manfred, rghosal, alf, cole}@eleceng.adelaide.edu.au

Abstract. This paper presents an overview of modern cryptography. As the title of this paper suggests this paper is a primer, and provides background information that can assist other researchers in further study, and in developing security mechanisms suitable for inclusion on RFID tags.

Keywords: Security, Authentication, RFID.

1 Introduction

Information technology security or cryptography, nowadays, is a well established area of computer science. Modern applications of cryptography allow us to transfer confidential data over insecure channels, take care of our bank transactions on transfer line using of-the-shelf computers, and sign obligatory contracts over the Internet. These applications are based on cryptographic building blocks, so-called primitives that provide certain desirable services like encrypting and decrypting messages.

Unfortunately, often it is not possible to straightforwardly implement cryptographic primitives in RFID systems because tags normally are very restricted in available power and, because they have to be cheap, chip area plays a crucial role. A demand for new, lightweight cryptographic primitives arises offering security services tailored to the requirements of RFID systems while considering their resource constraints.

This paper is structured as follows: Section 2 discusses information technology security in general presenting a state-of-the-art overview on cryptography. After dealing with desirable security services, common attacks, and security models, Section 3 covers cryptographic primitives divided into the three main groups unkeyed, secret-key and public-key primitives. Finally, a conclusion summarizes the main features of the work.

2 Information Technology Security

In the age of information processing, Internet and digital communication obviously there is a strong need for information technology security. Exchange of confidential messages, online-transfer of money and access to information services are just a few examples of procedures that rely on the security of computer systems and networks. Therefore, information systems and information processed by information systems have to be protected. Cryptography is the science that deals with protection of information [1]. In this section we cover the basics about information technology security and cryptography, and its appropriateness to RFID systems and in particular implementation of cryptographic primitives on an RFID Tag. Possible cryptographic solutions may be applied to the anti-counterfeiting applications for a secure supply chain. The reader may be adverted to [2] for definition of terms used in this section. This may be especially helpful if terms are used prior to their actual definition which is sometimes unavoidable.

2.1 Model

Although cryptography is much more than just encryption, encryption could be seen as the primary goal. Thus, to start to describe information technology security, the following simple model of Figure 1 is provided.

Alice and Bob are two parties who, despite or perhaps because of their somewhat pneumatic appearance, trust each other. They want to communicate using the channel, in their case a river, which they know to be insecure. In more real applications they can be human beings equipped with computing equipment, and communicating with electrical signals. In the above example, Alice sends a message which holds information designated for Bob. Eve is an unknown third party who also has access to the channel. The model could be interpreted in two different ways, either as transmission in space, where Alice and Bob sit at different places, or as transmission in time, where e.g. Alice stores a message onto the hard disk of a computer and Bob recalls it at a later time. Given this model, there are a number of concerns. Can Alice be sure that Bob is the only one who can read her message? Can Bob be sure that the message was sent by Alice and if it was, can he be

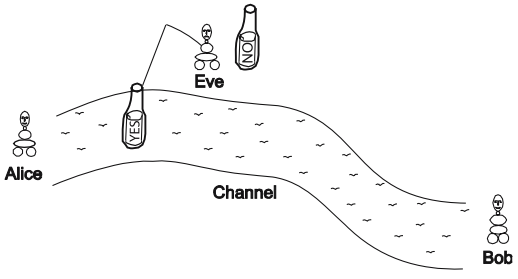


Fig. 1 Encryption model.

sure that it wasn't altered? Isn't Bob able to deny that he has received the message and isn't Alice able to deny that she has sent the message? These questions lead us to the services that are required for information technology security.

2.2 Security Services

As described in the preceding section, there are a number of security concerns when thinking about computer systems and digital communication. Therefore, a number of services have to be provided in order to enhance information technology security. The following paragraphs are dedicated to the most important of them [3], [4].

- **Confidentiality** ensures that only authorized parties are able to understand information.
- **Authentication** refers to the ability for a party to be sure the received information is from the source it claims to be from.
- **Integrity** assures that a message was not altered on the way to its recipient. Hence, it provides the recipient with the certainty to know what it has received is what was sent by the trusted origin of the message.
- **Non-repudiation** ensures that neither the sender is able to deny that it has sent the message nor the receiver is able to deny that it has received it. Therefore, it provides a proof of transmission and reception of a message.
- **Access Control** refers to the ability to restrict and control access to a system.
- **Availability** provides means to ensure that a system is available whenever needed. Thus, availability services guard systems from attacks against loss or reduction in availability.

Most of these services might be achieved by applying an appropriate cryptographic tool, like an encryption algorithm or a cryptographic hash function, or a series of those. If a series of tools has to be applied in a well-defined way, this way is referred to as a cryptographic protocol. Before we will discuss the main types of cryptographic tools we would like to spend a few words about attacks and security models.

2.3 Attacks

Modern cryptography tools (primitives) face a variety of attacks they have to withstand. Before classifying these attacks a basic principle of state-of-the-art cryptography has to be explained. According to Kerckhoffs' principle, stated in the nineteenth century by Auguste Kerckhoffs, the security of a cryptosystem must not depend on the secrecy of data independent details about the system [3]. Data independent details of a cryptographic system are the algorithm and its implementation. Therefore, attacks on modern primitives often aim at the recovery of plaintexts from ciphertexts or even worse on the recovery of secret keys [5].

Attacks may be classified into passive and active attacks [4]. Passive attacks denote monitoring of channels or side channels, but not alteration of messages. Obviously monitoring of a channel includes directly listening to data being transferred. Monitoring of side channels, is listening to effects that come along with the activities on the channel like electromagnetic emanation or current consumption [6]. Passive attacks on encryption schemes may be further subdivided into the following kinds of attacks [4]:

- **Ciphertext-only attack**: The attacker tries to recover plaintext or the secret key just by analysing the corresponding ciphertext.
- **Known-plaintext attack**: By analysing a given block of plaintext and the corresponding ciphertext the attacker tries to extract useful information for the recovery of plaintext encrypted in different ciphertexts or the secret key.
- **Chosen-plaintext attack**: Given the attacker is able to choose plaintexts and generate corresponding ciphertexts it tries to extract useful information in order to recover new plaintexts from new ciphertexts or even may try to extract the secret key.
- **Adaptive chosen-plaintext attack**: The attacker tries to recover plaintext or the secret key by subsequently applying chosen-plaintext attacks where the choice of the plaintext for later attacks depends on the outcome of prior attacks.
- **Chosen-ciphertext attack**: This attack is based on the assumption that the attacker, for a limited amount of time, is able to access means to decrypt ciphertexts. Therefore, we assume the means to be a black box which is able to decrypt a limited number of ciphertexts. The attacker chooses ciphertexts and analyses the corresponding plaintexts in order to gain useful information for the later purpose of recovering plaintexts from ciphertexts or of recovering the secret key without the availability of the black box.
- **Adaptive chosen-ciphertext attack**: As with the adaptive chosen-plaintext attack the attacker subsequently applies chosen-ciphertext attacks, with ciphertexts for later attacks depending on the outcome of prior attacks.

Although we mentioned the attacks above are aimed at encryption schemes, most of them are also applicable for other primitives like cryptographic hash functions or digital signature schemes.

Active attacks as opposed to passive attacks are based on alteration of data transmitted over a channel or alteration of computation in a device. The later is also referred to as fault attack [6]. Note that fault attacks and side-channel attacks are sometimes denoted as implementation attacks since they do not aim at the cryptographic algorithm but its specific implementation.

As mentioned above, security services might be based on primitives or on cryptographic protocols. Cryptographic protocols face yet another type of attack – so-called protocol attacks. The following list names the most important of them [4]:

- **Known-key attack**: Based on the knowledge of prior keys the attacker might try to obtain new keys.
- **Replay attack**: The attacker who is able to record a series of messages exchanged by the trusted parties replays part of it or the complete series at a later time.

- **Impersonation attack**: In this case, the attacker somehow assumes the identity of one of the trusted parties.
- **Dictionary attack**: This is a well known attack usually applied on password schemes. The adversary somehow manages to test a very large number of probable passwords with the intention guessing the right one.

2.4 Security Models

In order to describe the security of cryptographic primitives there are so-called security models. The following paragraphs describe the most important of them [1].

- **Unconditional security** also known as perfect secrecy is the non-plus-ultra security model. It assumes unrestricted computational power of the adversary. Therefore, for a cryptographic primitive to fall into this category there must not be an algorithm for breaking it, irrespective of the computational power available. An example of a simple primitive offering unconditional security is the one-time pad. In order to generate the ciphertext a plaintext is XOR-ed with a unique secret key of the same length as the plaintext. Because of the possible large key sizes such systems are impractical for conventional message encryption. However, there may be applications in systems with small information sizes like RFID.
- **Computational security** assumes polynomial computational power of the adversary. Therefore, a cryptographic primitive is assumed to be computationally secure if there is no algorithm known to break it within polynomial time. Modern primitives are supposed to fall into this category.
- **Practical security** also refers to the computational power of the adversary. However, as opposed to computational security there are no relative bounds. Instead, for a primitive falling into this category there must not be a breaking algorithm which requires less than N operations. The number of operations N is chosen sufficiently high. Modern cryptographic primitives typically offer practical security.
- **Provable security** means that it is possible to show that the complexity of breaking a primitive is equivalent to solving a well known supposedly hard mathematical problem like the integer factorization problem. Typical cryptographic primitives based on public keys fall into this category.

3 Cryptographic Primitives

So far, we discussed why information technology security is important, what potential threats are and how the security of cryptographic tools can be classified. What was not covered yet are cryptographic tools themselves, the primitives that provide us with the required services discussed above.

Table. 1. Some useful cryptographic primitives.

Some useful cryptographic primitives	
Un-keyed primitives	Hash functions One-way functions Random sequences
Secret-key primitives	Secret-key ciphers Message identification codes Identification primitives
Public-key primitives	Public-key ciphers Digital signatures

Cryptographic primitives may be separated into three large groups: un-keyed primitives, secret-key primitives and public-key primitives. Each of these groups may further be subdivided into primitives that serve different purposes. Table 1 shows a classification of some useful cryptographic primitives [4]. Each of these groups is further discussed the following sections.

Important differences between cryptographic primitives include the level of security, basic functionality, methods of operation, performance and the ease of implementation. Basic functionality deals with the objectives discussed in Section 2.2 whereas the methods of operation regard to specific ways primitives can be used, e.g. for encryption or decryption. Ease of implementation refers to both software and hardware environments [4].

3.1 Un-keyed Primitives

Un-keyed primitives, as the name betrays, are cryptographic tools that are not based on any keys at all. Examples for un-keyed primitives are cryptographic hash functions, one-way functions and random number generators. Taking into account Kerckhoffs' principle, because they are not based on keys, they do not fulfil security objectives on their own but often are part of a security system or a cryptographic protocol.

3.1.1 Hash Functions

"A hash function is a computationally efficient function mapping binary strings of arbitrary length to binary strings of some fixed length, called hash-values."[4]

The above definition is very basic. For hash functions used as cryptographic tools a set of requirements has to be fulfilled. First of all, it has to be a one-way function, which means, given a message m it should be easy to calculate the hash-value $h(m)$ but it should be computationally infeasible to do the inverse operation namely to find a message m given the hash value $h(m)$ so that $m = h(m)$. Second, good hash functions should be collision resistant. Theoretically this would mean that there are no two messages *m1* and *m2* so that $h(m1) = (m2)$. However, since there are an infinite number of inputs but a finite number of outputs collisions are unavoidable. Therefore, in

practice collision resistance means it should be computationally infeasible to find collisions. The third important requirement says that the hash-function should be a random mapping which in practice means it should be computationally infeasible to distinguish a hash-function from a random mapping [7].

Basically there are two types of attacks on hash functions, collision and pre-image attacks. Collision attacks in parallel try to find two different messages *m1* and *m2* that lead to the same hash value $h(m1) = h(m2)$. If we consider digital signatures (see Section 3.4.2) where normally hash values of messages are signed for efficiency reasons this leads to a serious security threat. By deliberately designing two messages with the same hash value a digital signature for one message automatically is valid for the second message. Pre-image attacks try to construct a message m that leads to a given hash value $h(m)$.

Hash functions are very important primitives in practice. They are used whenever fixed length values are required instead of arbitrary length messages. They may also be used to generate various pseudorandom keys from a single input.

Un-keyed hash functions are often called modification detection codes (MDC) because of their ability to detect whether a message has been altered. Well known MDCs used in practice are MD5 and SHA-1. MD5 has an output length of 128 bits. The best known attack recently was described by Vlastimil Klima and is able to find MD5 collisions in about one minute using a state-of-the-art notebook computer (Intel Pentium 1.6 GHz) [8]. SHA-1 has an output length of 160 bits. The best known attack, illustrated on Bruce Schneier's Weblog on behalf of the author Xiaoyun Wang, shows a time complexity of 2^{63} which is even better than a brute force attack which would lead to a time complexity of 2^{80} for an output length of 160 bits [9]. Although, so far only collision attacks but no pre-image attacks are known, MD5 and SHA-1 cannot be seen as practically secure any longer. Hence there is a need for new hash functions. The US government standards agency NIST (National Institute of Standards and Technology) in 2001 published a new group of SHA algorithms collectively known as SHA-2. This group includes algorithms with output lengths of 256, 384 and 512 bits. [7]. So far no practical attacks are known for SHA-2.

3.1.2 One-Way Functions

As already mentioned with hash functions, one-way functions are mappings $f: X \rightarrow Y$ which are easy to compute but hard to invert. Expressed in a more formal way this means a polynomial time algorithm exists for computation but no probabilistic polynomial time algorithm for inversion succeeds with better than negligible probability [10].

There are special one-way functions called trapdoor functions with the additional property that given special information, called the trapdoor information, it is possible to calculate the inversion.

One-way functions and trapdoor functions are very important primitives in cryptography and a lot of other primitives are based on them. Examples would be hash functions which are based on one-way functions or public-key cryptography which is based on trapdoor functions.

In more detail, a trapdoor function is defined as a function that is easy to compute in one direction, yet believed to be difficult to compute in the opposite direction (i.e. finding the inverse) without special information, called the "trapdoor". We define "difficult" or "infeasible" in this section to mean computationally intractable, i.e. not possible to perform in a reasonable amount of time, e.g. one year, based on current state of technology. Trapdoor functions are widely used in cryptography.

To provide an illustration of a trapdoor function we will use as a context the RSA encryption system.

In RSA [4] the encryption operation performed is $c = m^e \bmod (n)$ (were n is the product of two large primes p and q) with the encryption key e and the modulus n (and the encryption formula) being disclosed, (and the primes p and q not being disclosed). We observe here for use in the discussion below that finding the primes p and q from their publicly disclosed product is believed to be infeasible when the primes are large. Here the ciphertext c is regarded as a function of the plaintext m.

The RSA problem, i.e. finding the inverse, is defined as taking the e^{th} roots modulo a composite number n. It is regarded as infeasible to solve except in the circumstances described below.

The inverse of the encryption, i.e. finding the plaintext m from the ciphertext c, can be performed if we have some additional information called trapdoor information. It is done in practice in the RSA system by using a decryption key d and the formula $m = c^d \pmod{n}$, but finding the decryption key d from the publicly disclosed e and n is believed to be difficult to the point of being infeasible.

The decryption key d (and the decryption formula) could be regarded as the trapdoor. The two large primes p and q could alternatively be regarded as the trapdoor. The path to find d from them is tortuous, and will not be described here, but is feasible to traverse.

It should be mentioned that a rigorous justification of the existence of one-way functions is an open problem in theoretical computer science.

3.1.3 Random Number Generators

The third important un-keyed primitive in cryptography are random numbers or random number generators respectively. Many keyed primitives or even cryptographic protocols are based on random number sequences. Examples of use are keys used in public-key cryptography, session keys often used in secret-key cryptography or random sequences (called nonces) in cryptographic protocols.

Before looking at sources of random numbers we should discuss what is randomness? In general, in order to be able to talk about randomness a context has to be defined [4]. For example, we cannot talk about a random number 7 in general but 7 could be a number randomly selected or generated out of a container holding the numbers 1 to 10 which is the context in this case. Furthermore randomness is based on uniform distribution and independence [3]. Uniform distribution means that there is equal probability for each of the numbers in our container to be selected, i.e. the probability to randomly select or generate 7 out of the container 1 to 10 is assumed to be 1/10. Independence refers to sequences of random numbers.

In order to talk about a sequence of random numbers there must not be any coherence in the sequence of these numbers, i.e. in the example with a container holding the numbers from 1 to 10, after randomly selecting 7, probabilities of selecting any of the numbers 1 to 10 should be the same.

The sources of ideal random numbers as discussed above, if there are ideal random numbers at all, are based on physical means. Examples could be gas discharge tubes or leaky capacitors. However, often they tend to be costly or slow in the generation of random numbers. Another interesting source are so-called physical unclonable functions (PUF). Based on manufacturing variations of ICs they might be used for secret key generation in electronic devices [11]. However, nowadays most of the random number generators used for cryptography are based on software algorithms. Since pure software algorithms are deterministic the sequences generated are not really statistically random. Therefore, algorithms based random number generators are called pseudo-random number generators and the sequences generated are called pseudo-random number sequences. The sequences generated are based on short random sequences called seeds. Most of the time, the generation algorithms are known but the seed is unknown. Sequences generated by good pseudo-random number generators will pass many tests of randomness and therefore they are applicable for cryptographic purposes.

3.2 Secret-Key Primitives

Secret-key cryptography denotes information technology security systems that are based on keys secretly shared between trusted parties. In the literature also the word symmetric-key cryptography can be found meaning the same systems. As Figure 2 shows, there are various applications of secret-key cryptography. The following sections will highlight the most important of them.

3.2.1 Secret-Key Ciphers

Ciphers deal with encryption and decryption of information. In the world of secret-key cryptography basically there are two types of ciphers in use today, block ciphers and stream ciphers. Both are based on the same model which we will explain first with the aid of Figure 2.

Assuming the basic encryption model already described in Section 2.1 (Figure 1) of this work, Alice wants to send a message *m* to Bob secretly, i.e. Eve should not be able to understand what she sees on the channel. Therefore, Alice and Bob agree on a secret key-pair (only known to them) for encryption and decryption before the

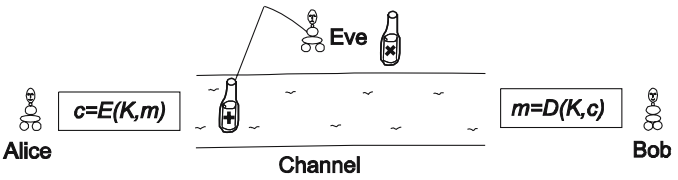


Fig. 2 Secret-key encryption.

actual start of the transmission. An important property of secret-key primitives is that the encryption key can be easily derived from the decryption key and the other way round. However, in nearly all modern secret-key encryption primitives the encryption and decryption key are the same and therefore are denoted as key K. Before sending the message m, Alice encrypts it applying an encryption function E and using the secret key K. The result of this operation is called ciphertext c which is transmitted over the channel. Therefore, $c = E(K, m)$. The size of the ciphertext is at least the size of the plaintext. This follows from the Pigeonhole Principle [12] which, converted to this scenario, states that if the output size of an encryption is smaller than its input size there must be necessarily 2 various inputs that lead to the same output which is not acceptable with encryption. On the other side, the size of ciphertexts could be bigger than the size of plaintexts if the plaintexts are padded before encryption for whatever reason. Bob, by applying the appropriate decryption function D and using the secret key K, recovers the original message m. Therefore, $m = D(K, c)$. Eve, who also received the ciphertext, is unable to understand it because she is missing the secret information, i.e. the key, to decrypt it. Assuming a good secret-key encryption primitive, it is computationally infeasible to recover the message from the ciphertext without knowing the key. What Bob will be able make of any bogus information that Eve may insert into the channel cannot be stated. A further question that is not covered in this scenario is how to exchange the key secretly. Unless Alice and Bob are unable to meet personally and need to exchange the key over a communication channel this may lead to a "Chicken and Egg problem" well known as the key distribution problem. However, there are smart solutions to this problem which may be gleaned in [4].

Note that although both concepts, encryption based on blocks and streams, also appear in public-key cryptography, the literature uses the names block cipher and stream cipher to denote secret-key encryption concepts.

Block ciphers form a special subset of secret-key ciphers where the message to be encrypted is divided into fixed length blocks. Each of these blocks, which are elements of the set of plaintexts, is transformed into an element of the set of ciphertexts. This happens under the influence of the secret key [1]. There is no change of size of the blocks during the transformation, i.e. ciphertext-blocks are of the same size as plaintext-blocks. The encryption is reversible, which means given the secret key it is also possible to recover plaintext blocks from ciphertext blocks. Block ciphers may be applied directly to messages with lengths equal to the size of a single block. If the size of the message is smaller than the block size padding is applied, i.e. additional symbols are added at the end of the message to fit the block size. Several padding techniques exist that allow distinguishing between message and padding. There are so-called modes of operation for messages longer than the block size [7].

Examples for block ciphers in use today are DES (data encryption standard), Triple DES (3DES) and AES (advanced encryption standard) [5]. DES was invented nearly 30 years ago and was one of the most widely used secret-key encryption algorithms. However, it is based on very short 64 bit blocks and 56 bit keys which, because of the increasing computational power available, seem not to be appropriate any more. There have been successful exhaustive key search attacks on DES already. Therefore, Triple DES (3DES) has been invented where DES is

applied three times with different keys. Therefore, it has a key size of 168 bits but inherits the disadvantages of DES like the small block size of 64 bits [7]. Triple DES is sometimes used because it offers more security compared with DES and for legacy reasons (Triple DES can be executed on DES hardware). However, in the meantime AES was invented to replace DES and 3DES. It is based on a block size of 128 bits and a selectable key size of 128, 192 or 256 bits. So far, there have been no successful attacks to the AES algorithm apart from side channel attacks which are implementation attacks rather than attacks on the algorithm [13].

Stream ciphers can be understood as block ciphers with block size one and the additional feature that the encryption transformation changes with every symbol processed. An advantage of stream ciphers is that they may have limited or no error propagation which means they may be able to deal with flipped bits or even with missing or inserted bits depending on their specific implementation. Therefore, often they are a good choice if errors are highly probable in transmissions.

A very simple example should clarify the operation of stream ciphers. The so-called Vernam Cipher is a stream cipher for binary streams [4]. The inputs are a binary message stream $m_1m_2m_3\ldots m_t$ and a so-called binary key stream $k_1k_2k_3\ldots k_t$. The binary key stream is a random binary sequence of the appropriate length, that is the length of the message stream. For encryption each binary symbol of the message stream is XOR-ed with the corresponding binary symbol of the key stream, that is $c_i = m_i$ XOR k_i. Obviously, to decrypt the ciphertext, the process has to be repeated, i.e. $m_i = c_i$ XOR k_i. The Vernam Cipher is exactly what we called one-time pad in Section 3.2.3 provides unconditional security (assumed the key stream is a truly random sequence with the same length as the message) but is hardly used in practice because of the large key size. However, stream ciphers used in practice are based on the one-time pad with the difference that the key stream is generated from a short random sequence by a deterministic algorithm. Thus, only the short random sequence has to be exchanged by the trusted parties.

As discussed above, practical stream ciphers require generators to create pseudo-random key-streams based on keys short enough to be exchanged conveniently. Often so-called linear feedback shift registers (LFSR) are used for this purpose. They exist of a series of delay blocks which are most often initialized with the secret key. Triggered by a clock signal the contents of the delay blocks are shifted. The input of the first delay block is a XOR of a subset of the delay blocks. Figure 3 shows an example of an LFSR.

Advantages of LFSRs are that they are easy to implement in hardware and software and they can produce sequences of large periods with reasonably good statistical properties [4]. The longest unique sequences are generated by so-called maximum length LFSRs. The period length of such LFSRs equals the maximum number of states that can be represented by a certain number of bits minus the zero

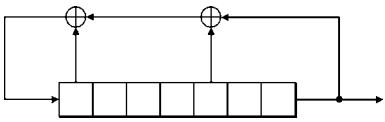


Fig. 3 Linear Feedback Shift Register (LFSR) [14].

state. Hence, a 3 bit maximum length LFSR has an output period length of $2^3 - 1 = 7$ unless it is initialized with 3 zeros.

However, with time mathematical methods have been developed to analyse LFSRs and they are not considered secure any longer [14]. Nevertheless, they form the building blocks of more secure key-stream generators in use today called nonlinear feedback shift registers (NLFSR).

NLFSRs generally consist of LFSRs as building blocks used in combination with methodologies that destroy the linearity of the output of simple LFSRs. In [4] three methods are discussed. So-called **nonlinear combination generators** apply a nonlinear function that combines the outputs of two or more LFSRs. An example for a nonlinear combination generator is the "Geffe generator" consisting of three maximum length LFSRs of pairwise relatively prime period lengths that are combined applying the function x_1x_2 XOR x_2x_3 XOR x_3. **Nonlinear filter generators** consist of just one maximum length LFSR which output is generated by a nonlinear combination of several stages of the LFSR. The generators discussed so far are clocked regularly, i.e. at each time step each LFSR is clocked. So-called **clock-controlled generators** introduce nonlinearity by clocking downstream LFSRs based on the state of upstream LFSRs. An example is the "alternating step generator" which consists of three LFSRs. The output of this generator is the XOR result of the output of two LFSRs which are clocked depending on the output of the third LFSR. If the output of the third LFSR is 1, one of the other two LFSRs is clocked, if it is 0 the other one is clocked.

Although there is no real standard for stream ciphers so far, RC4 is most widely used and can be seen as de-facto standard [15]. As with most stream ciphers, it is based on one-time pad with a pseudo-random key-stream generator. Other than using LFSRs the key-stream generator is designed to be easily implemented in software. RC4 is in very heavy use today. It is the cipher used in WEP and WPA and may be optionally selected to be used in SSL (secure socket layer) and SSH (secure shell). Nevertheless, there are known attacks with the best of them being able to distinguish a random sequence from the pseudo-random sequence generated by RC4 given about one gigabyte of output data [16].

When comparing stream ciphers with block ciphers, there are clear practical advantages of stream ciphers. They are much easier to implement in software and hardware, generally they are faster, they do not require large memories to store blocks and they can deal with errors in the way that there is no error propagation. However, relatively few fully specified stream ciphers are published in the literature, and as opposed to block ciphers there are no standardised stream ciphers so far [4], [15].

The secret-key primitives described so far ensure that Eve is unable to understand, what is transmitted over the channel. Ciphers do not provide security against alteration of messages.

3.2.2 Message Authentication Codes

In order to be able to detect changes of messages transmitted over insecure channels so-called message authentication codes (MAC) or cryptographic checksums are used [3]. They can be understood as hash functions on data that includes a secret key.

Figure 4 shows the principle of MACs. Note that the example does not consider encryption of the message. Before sending the message, Alice generates a message authentication code mac applying a MAC-function which processes the message m and a secret key K. Then, Alice transmits both, the message and the message authentication code. Bob receives them and also generates the message authentication code using the same MAC-function which processes the received message and the shared secret key. Thereafter, Bob compares the MAC he generated with the MAC he received along with the message. If they match, because of the shared secret key, Bob knows that the message was not altered and he knows that the message was sent by Alice. In the Figure 4, Eve, who also receives the message and the MAC, can understand and alter the message but there is no way to change it prior to forwarding it to Bob without Bob knowing that something went wrong. Additionally, Eve cannot create messages and send them to Bob as if she were Alice [7]. Therefore, message authentication codes provide data integrity and data origin authentication [4].

In practice there exist several types of MAC-functions. They might be based on block ciphers, like the DES-CBC MAC, be based on stream ciphers, be constructed from un-keyed hash functions applying a secret key, like the MD5 MAC, or they might be based on one-time pad cipher [17]. Construction of MAC-functions from un-keyed hash functions means that the original algorithm, like MD5, is altered to incorporate a secret key into the compression function [4].

In practice there exist several types of MAC-functions. They might be based on block ciphers, like the DES-CBC MAC, be based on stream ciphers, be constructed from un-keyed hash functions applying a secret key, like the MD5 MAC, or they might be based on one-time pad cipher [17]. Construction of MAC-functions from un-keyed hash functions means that the original algorithm, like MD5, is altered to incorporate a secret key into the compression function [4].

3.2.3 Identification Primitives

One of the most important classes of primitives in today's computer systems are identification techniques which are sometimes also called entity authentication or identity verification techniques [4]. The purpose of entity authentication techniques is to allow one party to gain assurance about the identity of another party. Thus, entity authentication tries to prevent impersonation attempts. When compared with message authentication techniques, entity authentication techniques

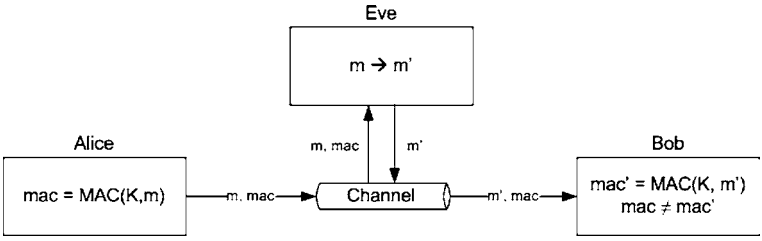


Fig. 4 Message Authentication Code (MAC). Modified from Figure 3.4 of Ferguson N. and Schneier B.: Practical Cryptography. Wiley Publishing, Indianapolis (2003).

typically involve no meaningful message but normally they are based on a real-time process, i.e. both parties are active at the same time).

Identification techniques may be divided into three categories. They may be based on something known by the identifying party, like a password or a secret key. They may be based on something possessed by the identifying party, like a magnetic stripe card or a smart card. Or, they may be based on something inherent to the identifying party, like voice, fingerprints or handwritten signature.

Typical applications of entity authentication include access control to resources, like information systems, sometimes accompanied with the need to track resource usage, e.g. for billing purposes.

Fixed passwords are a very simply scheme of entity authentication based on something known. They are shared secrets between users and an information technology system. For identification typically the system asks the user for a user-id and the appropriate password. If the data entered by the user matches the data stored in the system the identification was successful and the user is granted access. Different fixed password schemes may store the passwords either in plaintext, or in encrypted form. A major security problem of fixed passwords is the so-called replay attack, when an attacker records the password transmitted over a channel and replays it at a later time to be granted access to a system. For that reason fixed passwords often are refereed to as weak authentication [4]. Countermeasures include encryption of the channel or even better the use of one-time passwords.

There are different kinds of **one-time password** schemes. Two parties may either share a list of passwords using one after another, or they may sequentially update their password, or they may generate one-time passwords with the help of one-time functions [4].

The problem with both of the above described password schemes, fixed and one-time, is that the actual secrets are released by the party which tries to identify itself. Therefore, whenever an active attacker is able to gain access to the secret (e.g. the list of passwords in a one-time password scheme) it is subsequently able to impersonate the party to which the secret belongs to. **Challenge-response identification** schemes address this vulnerability. Rather than releasing the actual secret they generate a response which is based on the secret and a time-variant challenge. That is why challenge-response schemes are also known as strong authentication mechanisms [4]. The time-variant challenge is provided by the verifying party every time an unknown party wants to be identified. A challenge includes a so-called nonce being the time-variant parameter. Nonces could be, for example, pseudorandom numbers, sequence numbers, or time stamps. The use of nonces prevents replay attacks as described with fixed passwords. It should be mentioned that although challenge-response schemes often are based on secret-key cryptography there are also implementations based on public-key primitives.

3.4 Public-Key Primitives

In addition to secret-key primitives, so-called public-key primitives form the second large group of keyed cryptographic tools [4]. Public-key primitives are based

on key pairs instead of single shared keys. Each key pair is made up of a public key and a private key that are linked together mathematically. As their names betray, one of the keys has to be kept secret by the owner and the other one is shared publicly. Because public-key primitives are based on two different keys, they are often called asymmetric-key primitives. The main motivation to have public-key primitives is that with secret-key primitives each pair of trusted parties has to share one secret key [7]. Since this is very complex, if there is a larger number of parties involved, public-key primitives provide an appropriate solution because only one key has to be shared publicly for each party. On the other side, public-key primitives are less efficient and that is why there is still a need for secret-key primitives. The following sections describe the two most important applications of public-key cryptographic tools, public-key ciphers and public-key signature schemes.

3.4.1 Public-Key Ciphers

As already mentioned with secret-key primitives, ciphers deal with the encryption and decryption of messages. Figure 5 shows the basic principle of public-key encryption [7].

The key pair used in this example is the secret key of Bob (SBob) and the public key of Bob (PBob). As already mentioned earlier these keys are linked together mathematically in order to be able to use the public key for encryption and the private key for decryption. Therefore, a major premise for public-key cryptography is that it should be computationally infeasible to derive the secret key given a public key. Now, if Alice wants to send a message m to Bob secretly she encrypts it using the encryption function E under the influence of Bobs public key PBob. The resulting ciphertext c can be transferred over the insecure channel because only Bob is in possession of the secret key SBob which is necessary for decryption. Hence, in order to be able to read the received encrypted message c Bob decrypts it using the decryption function D under the influence of the secret key SBob. The example also shows Eve, the passive attacker, who is able to receive the public key PBob and the ciphertext c but cannot extract any useful information thereof.

It may now seem that the problem of secretly exchanging keys in secret-key cryptography is solved because public keys can be transferred over insecure channels. In fact, many practical systems are based on a mixture of secret- and asymmetric encryption. Often asymmetric-key encryption is used to agree on a shared

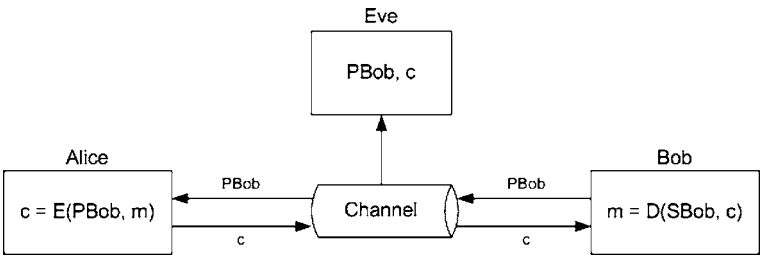


Fig. 5 Public-key encryption.

secret short term key (session key) which further on is used in secret-key encryption to exchange the actual information. The term session key is based on the fact that it is used for a limited time (the session) only. This approach combines the advantages of both, the publicly shared key of public-key cryptography and the efficiency of secret-key cryptography [4]. However, there is a remaining issue with public-key cryptography. Since public keys normally are exchanged over insecure channels an active attacker would be able to impersonate another party by providing a public key which seems to belong to this other party but actually is part of a key pair of which the attacker is in possession of the private key. Consequently, the attacker would be able to decrypt messages which were intended for the impersonated party. For that reason public keys often are exchanged using a so-called public key infrastructure (PKI) which solves the impersonation problem by issuing certificates. Certificates basically store identities and corresponding public keys. Each certificate is digitally signed (see next section) by a trusted third party who consequently prevents impersonation attacks as described above. Ferguson and Schneier ([7]) provide further information about PKIs for the interested reader.

Probably the most important public-key technique is the **RSA** cryptosystem [4], [5]. It was invented in 1978 by R. Rivest, A. Shamir and L. Adleman and is based on the well known integer factorization problem. The idea is to multiply two sufficiently large prime numbers p and q to obtain $n = p \times q$. Since it is believed (not proven) to be a hard mathematical problem to factorize n into its factors p and q, n is part of the public key and p and q are parts of the private key.

No practical attacks are known on the RSA cryptosystem provided it is based on sufficiently large keys (size of n). In respect to the computational power available, current references recommended to have keys of at least 2048 bits tending to 4096 or even 8192 bits for future applications.

The security of public-key cryptosystem is based on keys made up of very large integers. Usage of large keys, however, leads to less efficient execution of algorithms. Especially when considering mobile devices which are limited in computing power and energy efficient cryptosystems play an important role. Therefore, much attention is paid to **elliptic curve cryptography** (ECC) which offers public-key techniques that are much more efficient than traditional algorithms. ECC-calculations are based on points on elliptic curves. Since elliptic curves used in cryptography are defined in terms of modular arithmetic they only contain a limited number of points. The main operation used in ECC is called scalar point multiplication which means deriving a point P which satisfies the equation $P = kQ$ for a given point Q and a given integer k. This is a one-way function, i.e. the security of ECC is based on the so-called elliptic curve discrete logarithm problem (ECDLP), namely finding a k in $P = kQ$ for a given P and Q [1]. Solving the ECDLP is considered to be computationally infeasible if k is sufficiently large. Since the best known algorithm for solving the ECDLP shows exponential complexity key sizes of more than 224 bits are regarded to be sufficient large taking into account the computational power available at the moment [18].

3.4.2 Digital Signatures

Digital signatures are an essential cryptographic primitive in providing authentication, authorization and non-repudiation services. Signing primitives used in digital signatures provide a method to bind the identity of the signatory entity to the message to be transmitted.

While there are many digital signature algorithms, all of them are based on public key algorithms. That is there is some secret information that only the signing entity has knowledge of and there is some public information that allows any other entity to verify the signature. In this context the process of signing is called encrypting with the private key while the process of verification is called decrypting with a public key. Unfortunately the usage of terminology in signature schemes can be confusing when considered with that of public key ciphers.

Figure 6 shows the principle of a digital signature scheme [7]. We assume that Alice has already generated her key pair, i.e. SAlice, her private key, and PAlice, her public key. Whenever Alice wants to sign a message m she applies the signing algorithm S under the influence of her private key to this message with the result of a signature s. Subsequently, she distributes the message and the corresponding signature over the insecure channel. Bob, who wants to verify her message, applies the verification algorithm V under the influence of Alice's public key to the received message and the signature. The outcome of the verification could be either 'valid' or 'invalid'. 'Valid' would mean that the message was signed by Alice and it was not altered during its transmission. In our example Eve altered the message. Therefore, the result of Bob's verification is 'invalid'. Due to the fact that public-key techniques generally are less efficient, in practice often hash values of messages are signed rather than the actual messages.

One simple method of implementing a digital signature scheme is by using the RSA public key cipher where the encryption and decryption functions are both inverse operations of the other. This property is unique to the RSA cipher. Other examples are DSS (digital signature standard), ElGamal (named after its inventor Elgamal), and algorithms based on elliptic curve techniques. Obviously, a fact that all these algorithms share in common is that they are based on trapdoor functions with the trapdoor information being the private key [5].

The **ElGamal** digital signature scheme is partly based on the discrete exponential function and partly based on the Diffie-Hellman key agreement. The discrete exponential function was already covered in Section 3.1.2 of this work. Since the description of the ElGamal scheme goes beyond the scope of this work

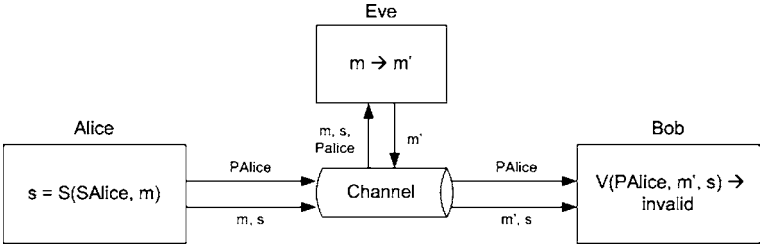


Fig. 6 Digital signature scheme.

the interested reader is adverted to [4] and [5] which provide detailed information about this topic and also cover the digital signature standard (**DSS**) which is another public-key technique used for digital signing and which is very similar to the ElGamal scheme.

3.5 Comparison of Secret-Key Primitives with Public-Key Primitives

Most of the advantages and disadvantages of secret-key primitives and public-key primitives were already discussed in the paragraphs above. This section is intended to summarize them and to provide a few additional notes.

The main advantages of secret-key primitives are that they are based on relatively short keys and that there are often efficient hardware implementations for them. Hence, they are applicable for processing large amounts of data. However, secret-key primitives come along with the disadvantage that a large number of keys have to be managed since each pair of trusted parties has to share an individual secret key. Additionally, it is considered as good practice to change the secret key regularly. This is to keep the amount of data being processed using a single key low in order to minimize the amount of input for potential attacks.

Public-key primitives, on the other side, have the advantage of easy key management. That is, only one key (the own private key) needs to be kept secret and public keys can be distributed over insecure channels. But most public-key techniques are based on very large key sizes which lead to less efficient execution of algorithms.

Considering the advantages and disadvantages of the large groups of primitives, today's cryptographic systems often are based on a mixture of both. These systems apply public-key cryptography to agree on so-called session keys which are shared secrets that are used for a limited time (the duration of a session) only. Subsequently, session keys are used in secret-key cryptography for processing the actual information. Therefore, the advantages of easy key management and efficient data processing have been combined in those mixed systems.

Elliptic curve cryptography, in future, may partially replace mixed systems since they combine the advantages of easy key management and fast data processing in one public-key scheme [19].

4 Conclusions

This paper provides an overview of state-of-the-art cryptography. Starting with information technology security it covers desirable security services, attacks and security models in general. Divided into the three groups un-keyed, secret-key and public-key, modern cryptographic primitives are presented.

Although standard cryptographic primitives offer aid to secure low cost RFID systems, resource constraints impede us from implementing most of the ordinary

cryptographic tools. RFID tags are very restricted in available operating power and chip size and unfortunately most of the cryptographic primitives require both. Hence, there is a need for new lightweight cryptographic primitives to be used in RFID technology.

Such lightweight primitives should be based on the well established knowledge on cryptography but tailor the services to the requirements and constraints of RFID.

References

1 Oswald, E.: IT security lecture notes. Institute for Applied Information Processing and Communications, Graz University of Technology, Austria (2005)
2 Grasso A. and Cole P., Definition of Terms Used by the Auto-ID Labs in the Anti-Counterfeiting White Paper Series. Available from http://autoidlabs.eleceng.adelaide.edu.au/static/Definition%20of%20Terms.pdf (7.06.2007)
3 Stallings, W.: Network and internetwork security: principles and practise. Prentice-Hall, New Jersey (1995)
4 Menezes, A., van Oorschot, P., Vanstone, S.: Handbook of Applied Cryptography. 2nd edn. CRC Press, Boca Raton (1997)
5 Delfs, H., Knebl, H.: Introduction to Cryptography: Principles and Applications. Springer-Verlag, Berlin, Heidelberg, New York (2002)
6 Oswald, E.: Introduction to Information Security lecture notes. Institute for Applied Information Processing and Communications, Graz University of Technology, Austria (2004)
7 Ferguson, N., Schneier, B.: Practical Cryptography. Wiley Publishing, Indianapolis (2003)
8 Klima, V.: Tunnels in hash functions: MD5 collisions within a minute. In: Cryptology ePrint Archive. Available from: http://eprint.iacr.org/2006/ 105.pdf (4.05.2006)
9 Schneier, B.: New cryptanalytic results against SHA-1. In: Weblog: Schneier on Security (2005). Available from: http://www.schneier.com/blog/archives/2005/08/ new_cryptanalyt. html (4.05.2006)
10 Goldreich, O: Foundations of Cryptography: Basic Tools. Cambridge University Press, Cambridge (2001)
11 Ranasinghe, D., Lim, D., Devadas, S., Abbott, D., Cole, P.: Random numbers from metastability and thermal noise. In: IEE Electronic Letters, Vol. 41, Iss. 16 (2005) 13–14
12 Grimaldi, R.P.: Discrete and Combinatorial Mathematics: An Applied Introduction. 4th ed. (1998) 244–248
13 AES Lounge: AES security. Available from: http://www.iaik.tu-graz.ac.at/research/krypto/AES/index.php#security (6.03.2006)
14 RSA Laboratories: What is a linear feedback shift register? Available from: http://www.rsasecurity.com/rsalabs/node.asp?id=2175 (4.04.2006)
15 RSA Laboratories: What is a stream cipher? Available from: http://www.rsasecurity.com/rsalabs/node.asp?id=2174 (4.04.2006)
16 Fluhrer, S.R., McGrew, D.A.: Statistical analysis of the alleged RC4 keystream generator. In FSE (2000) 19–30

17 RSA Laboratories: What are Message Authentication Codes? Available from: http://www.rsasecurity.com/rsalabs/node.asp?id=2177 (4.04.2006)
18 Chang, S., Eberle, H., Gupta, V., Gura, N.: Elliptic curve cryptography – how it works. Sun Microsystems Laboratories (2004). Available from: http://research.sun.com/projects/crypto/ (9.04.2006)
19 Gura, N., Shantz, S., Eberle, H., et al.: An end-to end systems approach to elliptic curve cryptography. Sun Microsystems Laboratories (2002). Available from: http://research.sun.com/projects/crypto (9.04.2006)

Part II
Security and Privacy Current Status

Chapter 6

Addressing Insecurities and Violations of Privacy

Damith C. Ranasinghe[1], and Peter H. Cole[1]

[1] Auto-ID Lab, The School of Electrical and Electronic Engineering, The University of Adelaide, Adelaide SA 5005, Australia. {damith, cole}@eleceng.adelaide.edu.au

Abstract. RFID systems, and indeed other forms of wireless technologies, are now a pervasive form of computing. In the context of security and privacy, the most threatening (to privacy) and vulnerable (to insecurity) are the 'low cost RFID systems'. The problems are further aggravated by the fact that it is this form of RFID that is set to proliferate through various consumer goods supply chains throughout the world. This is occurring through the actions of multinational companies like Wal-Mart, Tesco, Metro UPS and of powerful government organizations such as the United States DOD (Department Of Defence) and FDA (Food and Drug Administration). This paper examines the vulnerabilities of current low cost RFID systems and explores the security and privacy threats posed as a result of those vulnerabilities. The paper will also formulate a framework for defining the problem space constructed around low cost RFID systems, and consider the challenges faced in engineering solutions to overcome the defencelessness of low cost RFID systems. Security issues beyond and including interrogators will not be considered as such concerns may be easily resolved using existing technology and knowledge, and because interrogators are powerful devices where complex encryption and decryption operations may be performed using either the embedded systems, DSPs, or using hardware implementation of encryption engines on a FPGA device onboard a reader.

Keywords: Security, Privacy, Low Cost RFID

1 Introduction

"RFID" is increasingly used as a common term to encompass a number of different implementations of RFID technology such as VeriChip [1] and SpeedPass payment tokens; however the focus of this paper will be on low cost RFID systems as identified in Section 2, with its primary application being the tagging of cases and pallets in supply chain applications.

One of the inhibitors to wide-scale adoption of RFID technology is the cost of a label. Thus low cost RFID refers to an RFID system based on inexpensive RFID tags. It is imperative that the cost of RFID labels is reduced if RFID technology is to gain any significant market penetration. For example, the current cost of a gate of silicon logic is about one thousandth of a cent [2]. Thus, a company producing 100 billion units of a product per year would loose $1 million in profits due to the addition of a single logic gate to a label. Therefore, a great deal of attention is naturally focused on low cost RFID.

The proposed Class I and Class II labels by EPCglobal represent the low cost end of RFID labels. A characterization of a low cost RFID system with its cost structure is provided in Section 2 and such a system will be analyzed to highlight its vulnerabilities in the following sections with details of how to such weaknesses can be overcome.

2 Characteristics of a Low Cost RFID System

The most dominant form of low cost RFID technology set to spread through out the consumer goods supply chain is that advocated by EPCglobal as Class I and Class II. The low cost RFID labels involving Class I and Class II labels are based on passive RFID technology. Due to their potential for prolific use in the future, most discussions regarding low cost RFID inevitably consider various aspects of such labels. The following sections provide an overview of low cost label manufacturing costs and IC components in an RFID label and focus on describing low cost RFID systems based around Class I and Class II labels.

2.1 A Low Cost Tag

Current fabrications of Class I labels consist of around 7,000 to 15,000 logic gates while Class II labels may have several thousand more gates. An RFID micro-circuit can be subdivided into three primary sections: RF front-end, Memory circuitry, and Finite State machine (label logic circuitry). Figure 1 is an illustration of a typical low cost RFID transponder (that is, a passive label). The block diagram of an HF chip and a UHF chip varies little, the primary difference being the way in which the local oscillator clock is derived. In a UHF chip there is a dedicated low power oscillator, while in an HF chip, the clock signal is derived from the received carrier by dividing down the carrier (at 13.56 MHz) in stages.

2.1.1 RF Front-end

RF front-end consists of antenna pads for attaching the terminal of the antenna to the label IC. The antenna input passes through circuits for ESD (electrostatic discharge) protection. The ASK (Amplitude Shift Keying) demodulation circuits extract the modulation dips from the received signal while the Rectifier rectifies the

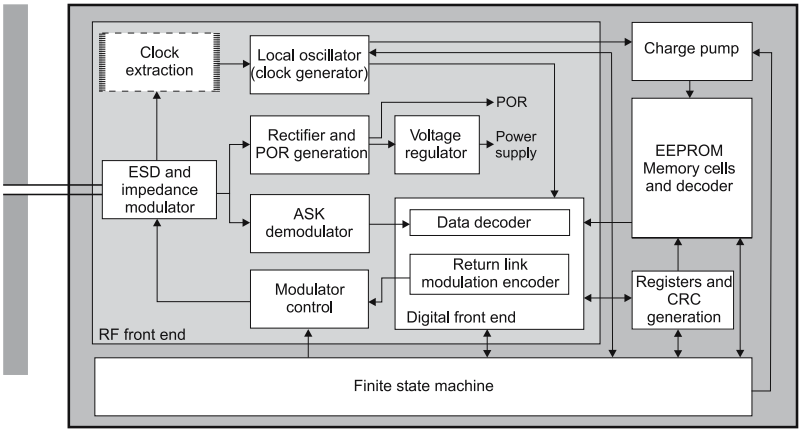


Fig. 1 A block diagram of a passive UHF/HF RFID label (Damith C. Ranasinghe and Peter H. Cole, "Confronting Security and Privacy Threats in Modern RFID Systems", 40th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, California, 29 Oct. –1 Nov. 2006. © 2006 IEEE).

received signal to generate power which must be regulated using a voltage regulator to avoid voltage surges due to variations in RF field intensities.

Passive RFID chips contain a relatively large capacitor following a rectifier for storing charge to power the circuit in the absence of a battery. It is important to note here that the capacitor occupies a relatively large portion of the silicon area and RFID chips consuming larger amounts of power will need higher capacity capacitors and thus will cost more.

2.1.2 Memory Circuitry

Low cost tags have limited memory that is either write once or a read-write memory. Class 1 labels have only read only memory while Class II labels may have some read-write memory. Read-write memory, at the time of writing, is implemented using EEPROM and thus requires a large voltage before information can be written to memory. Thus a charge pump, consisting of a series of capacitors, is required to achieve a voltage of about 17 V for writing to the tag's memory.

The CRC circuits are used in validating the CRC in the received data and commands from an interrogator. The CRC generation unit is also used in the computation of the CRC for data sent from the tag to an interrogator before being encoded for modulation by the Return link modulation encoder.

In the implementation of an EPC, tag the E^2PROM will store the EPC number of the tag and the rest of the memory (generally in the order of a few kilobytes) is available to the users. A tag's memory resources account for a significant portion of the tag cost.

2.1.3 Finite State Machine (Logic Circuitry)

The logic on board the chip will define the label functionality. Primarily, chip logic will execute reader commands and implement an anti-collision scheme that allows the reading of multiple labels by a reader. These logic circuits are highly specialized and optimized for their tasks. Furthermore, the logic circuits also control read and write access to the EEPROM memory circuits.

A block diagram of a low cost RFID tag is given in Figure 1 along with a description of the various functionalities of the tag components. The following Sections provide quantitative characteristics of low cost RFID systems and reasonable assumptions that need to be taken into consideration when solutions for security and privacy issues are developed.

Computation capability of a low cost tag is limited to a state machine with hard wired logic functionality. The only arithmetic operation performed by current low cost RFID tags is the calculation of a CRC for checking errors in received data and the computation of a CRC prior to transmitting data. Thus, for a low cost tag any additional hardware required to implement security needs to be designed and fabricated incurring additional cost.

2.2 Tag Cost

Tag cost is generally based on the evaluation of the area of silicon required for a physical implementation. While this includes the analogue front end of the tag, a reduction in costs has been achieved through the miniaturization of digital functional blocks and not through devices such as capacitors, inductors or resistors. Hence, keeping tag costs low requires focusing on limiting the number of gates on a tag even though the bulk of the tag cost is associated with the analogue components whose costs are difficult to reduce due the nature of passive components.

2.2.1 Manufacturing Costs

There are a number of key stages involved in the manufacture of RFID labels after the design of the IC. An outline of the stages is given in Figure 2 below. Today, the cheapest RFID labels are passive and cost around 10 US cents in large quantities [3]. Presently RFID read only chips have design sizes ranging from 0.16 mm^2 [3] to 0.25 mm^2 [4] IC foot prints.

Further improvements to IC manufacturing processes will bring the cost of microcircuits even lower. This invariably involves producing more microcircuits per silicon wafer. However, reducing die sizes to very small levels can incur added costs due to the increase in cost of handling smaller die.

A more practical avenue for reducing costs is the use of obsolete IC manufacturing processes and filling up such fabrication pipelines with RFID IC chips. This is a worthwhile consideration as people migrate to smaller and smaller micron processes and larger, highly tuned micron processes such as 0.5 micron become available at a fraction of the cost due to depreciated fabrication equipment and the availability of smaller processes. This will reduce the cost of the IC component of

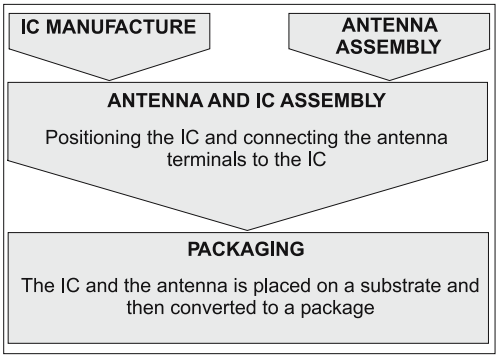


Fig. 2 RFID Label Manufacturing processes (Damith C. Ranasinghe, Daniel W. Engels, Peter H. Cole "Low cost RFID systems: confronting security and privacy", 2005 Auto-ID Labs White Paper Journal Volume 1, © 2005 Auto-ID Labs).

the chip considerably. The older processes have the added advantage of having few or no reliability concerns while being able to provide stable yields.

2.3 Tag Power Consumption

A tag's power consumption will vary depending on whether the tag is just being interrogated or whether the tag is required to perform a write operation. Tag power consumption is also influenced by other factors such as the data transmission rate, the feature size of the fabrication process used, as well as the effort spent in designing low power CMOS circuitry. A tag performing a read operation will require about 5 μW – 10 μW, while a tag attempting to perform a write operation to its E^2PROM will require about 50 μW or more.

2.4 Physical Protection (Tamper Proofing)

Low cost tags do not utilize anti-tampering technology due to cost constraints and therefore the contents of a labels memory or the layout of logic circuits are not protected from physical access. Hence the long-term security of label contents cannot be guaranteed.

2.5 Standards

There are a variety of standards encompassing all aspects of RFID systems. The ISO 18000 is a multi-part standard that defines the air interface standard of a number of different frequencies from LF, HF to UHF. However for UHF, tags the most prevalent standard is that ratified by EPCglobal, called the Class I Generation II air interface protocol [5]. However the recent amendment to ISO 18000-6 to include

Type C has given rise to a protocol specification almost equivalent to EPCglobal's Class I Generation II.

Accordingly, the labels within reading range have a means of revealing their presence, but not their data, when interrogated by a reader. The labels then reply with a non-identifying signal to an interrogation by using a randomly generated number as described in C1G2 air interface protocol [5].

However, for HF tags, there is no such prevalent standard, although EPCglobal is currently developing a HF specification to complement its UHF air interface protocol. Possibly the most prevalent HF tag protocol specification is the ISO 18000-3 Mode 1, most commonly used by tag deployments in various libraries around the world. The existing standards most commonly in use for HF tags, other than the ISO 18000, are listed below.

- ISO 14443 (types A and B). Devices operating under this standard are proximity RFID devices with a reading range of a few centimeters
- ISO 15693 is a recent addition for "vicinity card" RFID devices, where the operating range of the devices can be close to 1 meter (the operating mode I of ISO part 3 specification is based on ISO 15693)

2.6 System Operational Requirements

RFID systems are required to meet various minimum performance criteria to justify their benefits to the end user community. Two such important and related performance parameters are the number of label reads per second and data transmission speeds. Performance criteria of an RFID system demand a minimum label reading speed in excess of 200 labels per second. In accordance with C1G2 protocol, a maximum tag to reader data transmission rate of 640 kbps and a reader to tag data transmission rate of 126 kbps based on equi-probable binary ones and zeros in the transmission can be calculated.

2.7 Communication Range

Considering the current electromagnetic compatibility (EMC) regulations, the operating range of low cost labels is limited to a few meters for those operating in the UHF spectrum and a few centimeters for those operating under the FCC regulations for the HF spectrum [45] (HF systems operating under current European regulations for the HF spectrum can have an operating range well in excess of 1 meter as discussed in Section 2.8).

2.8 Frequency of Operation and Regulations

Important considerations affecting all EM related issues, especially the powering of RFID labels, are the regulations that govern the operating frequency, power,

and bandwidth in different regions of the world. There are a number of regulatory organizations and different EMC regulations around the world. Australian regulators are likely to follow the footsteps of their US counterparts; this is highlighted by the experimental license granted to GS1 (Australia) which stipulates a reader radiated power of 4 W EIRP from 920 MHz to 926 MHz by the Australian Communication Authority. Hence, treatment given here for EMC regulations will not focus on Australian regulations. It is important to note here that the EMC regulations are enforced in the far field.

Most RFID systems operate in the Industrial, Scientific and Medical (ISM) bands designated by the ITU [6]. The ISM radio bands were originally reserved internationally for non-commercial use of RF electromagnetic fields for industrial, scientific and medical purposes. The most commonly used HF ISM band in Europe and America is centered at 13.56 MHz and the UHF band in the US is 902–928 MHz [7].

Figure 3 shows the revised European regulations at 13.56 MHz and the revised FCC regulations [7, 8]. FCC regulations for the HF spectrum allow only minimal 5 mW radiation when using an antenna of gain 1.76 dBi. Hence devices operating under these regulations only have a very small reading range in the order of a few centimeters. However European regulations depicted in Figure 3 allow radiating 320 mW of power with an antenna gain of 1.76 dBi. Using larger interrogator antennas and large label antennas have shown that reading ranges under European regulations [8] can be increased to approach the mid field distance (that is, around the 3 metre range).

Near and far fields scale differently with distance and, in particular, the near field energy density per unit volume decreases as the inverse sixth power of distance from the antenna [9 and 10]. The result is that close to the antenna, substantial energy densities may be obtained, but these diminish very quickly as distance increases. The limits on the radiated power generally ensure that the previously

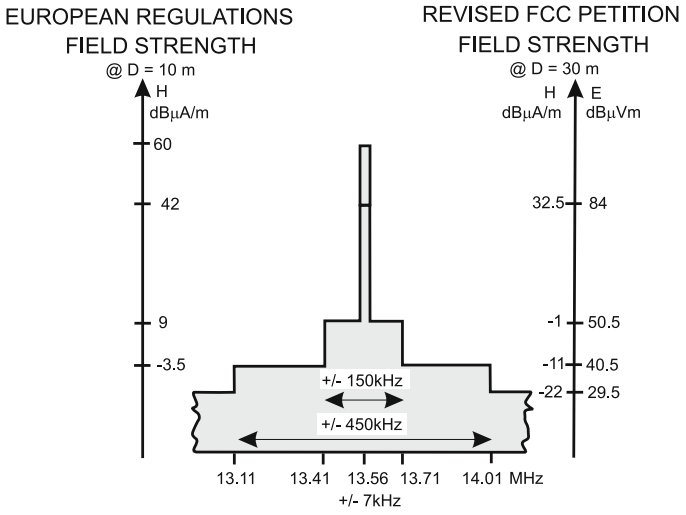


Fig. 3 Revised HF electromagnetic compatibility regulations.

mentioned inverse sixth power of the reactive power density sufficiently reduces the label energizing signal to a level below an acceptable level for practical operation before the boundary of the far field. Thus, under current regulations, operation of HF systems is almost entirely confined to the near field and short distances. Conversely, at UHF frequencies, the boundary between the near field and the far field is in the vicinity of the antenna; thus, the operation of UHF systems is almost entirely in the far field region.

Each frequency band provides its own set of advantages and disadvantages. The 13.56 MHz band has a 14 KHz bandwidth. This places a limitation on the bandwidth of the reader to label communication since the central portion of the spectrum shown in Figure 3.2 regulates the operation of RFID equipment in the HF region.

The 902–928 MHz band, under US regulations, allows multiple reader to label communication choices with much higher communication bandwidths and hence data rates. The regulations allowing the longest communication range require the reader to change its communication frequency every 400 milliseconds. The reader may hop between any numbers of channels; however the maximum bandwidth of a channel cannot exceed 500 kHz [7]. This technique is referred to as 'frequency hopping'. Table 1 below highlights the range of frequencies in use in the UHF region around the world.

2.9 Security Provided by Class I and Class II labels

The most dominant form of low cost RFID technology set to spread throughout the consumer goods supply chain is that advocated by EPCglobal as Class I and Class II. The low cost RFID labels involving Class I and Class II labels are based on passive RFID technology [11]. Due to their potential for prolific use in the future, most discussions regarding low cost RFID inevitably consider various aspects of these labels.

Led by EPCglobal, the RFID community in its efforts towards standardization has produced a list of end user requirements for Class I and II labels that has flowed into the current C1G2 protocol standards and is outlined in the following

Table 1 UHF RFID frequency allocations.

Region	Frequency range (MHz)	Bandwidth (MHz)
Europe	865 – 868	3
USA	902 – 928	26
Japan	952 – 954	2
China	920.25 – 924.75 (@ 2 W ERP)	4.5
	840.25 – 844.75 (@ 2 W ERP)	4.5
Australia	918 – 926	8
	920 – 926 (experimental band till 12th July 2007 with 4 W EIRP EMC regulation limit)	6

sections. One aim of that list has been to address the privacy and security risks posed by RFID Class I and Class II labels containing an EPC. The security requirements are an appropriate guideline when considering the level of security and privacy that can be expected and required from Class I and Class II RFID labels. The following is an outline of the security features that can be expected from the previously mentioned classes of labels.

2.9.1 Security Features of Class I Generation 2 Labels

Class I labels, identified in [11], have only a read only or a write-once memory and are incapable of participating in a complex security mechanism. Hence Class I labels are required to provide "Kill" capability and a password to control access to the kill command, so that consumers have the choice of completely disabling an RFID label at the time an RFID labeled item is purchased.

"Killing" a label involves the destruction of the label thus rendering it inoperable [5] by perhaps setting off a fuse or disconnecting the antenna. Unfortunately, the destruction of the label denies the user the significant benefits that could have been obtained from a "smart object". As a solution, an alternative idea to killing entertained previously involved the removal of the unique serial number of the EPC code in articles that allows the label owners to be tracked, albeit with difficulty in practice. This does not remove all the privacy concerns as tracking is still possible by associating a "constellation" of a label group with an individual. This implies that a particular taste in clothes and shoes may allow an individual's location privacy or anonymity to be violated. However "killing" a label will eliminate privacy concerns and prevent access by unauthorized readers when combined with a password to control access to the kill command.

While throttling is not specified as part of the C1G2 standard, it is reasonable to assume the employment of a delay based throttling mechanism on tags to prevent the guessing of kill or access passwords [12]. The concept behind delay based throttling is that, on the occasions a tag is given an invalid password, the tag enters a sleep state where it will not accept another kill attempt for a specified amount of time. This method can significantly increase the time required by an attacker attempting to kill a tag. In a situation such as a retail environment, the delay factor can be an adequate deterrent to such brute force attacks.

Class I labels should also have the ability to lock EPC data so as to provide one-time, permanent lock of EPC data on the label, so that EPC data cannot be changed by an unauthorized interrogator once it has been written. Interrogators are also prevented from transmitting complete EPC data except when data needs to be written to an RFID label, so that the EPC information may not be eavesdropped upon from a distance without being discovered.

2.9.2 Security Features Expected from Class II Labels

Other requirements were identified as necessary for higher class labels since these labels will have greater functionality and thus more hardware. Higher class labels are required to provide a secure forward link for communication with an RFID label while providing access control to label functionalities.

2.10 Backend System Services: Track and Trace Capability

RFID labels are given a unique identification number: for Class I and Class II labels, the unique identifier is an EPC. Using information technology services offered by backend systems, such as the EPC Network services, it is possible to dynamically generate a profile of the RFID label to create an electronic history of the label as it passes through various stages of the supply chain. The scheme was discussed in detail in [13]. The electronic history, called an electronic pedigree, can serve to thwart cloning attacks.

3 Vulnerabilities of Low Cost RFID Systems

Low cost RFID technology described in Section 0 has the potential to promote a sound business case because of its potential to save costs and improve processes, while providing certain security benefits. However, low cost RFID systems generate significant security risks, mainly due to their cost constrained implementations and the insecure communication channels over which tags and readers communicate. The security risks that arise as a result are due to a number of reasons outlined below.

- Communication between a tag and a reader takes place over an insecure channel
- Tags are accessible by any reader implementing the air interface protocol
- Tags are not tamper proof and allow a channel for physical access to tag contents and circuitry (as a result, tags cannot be expected to secure information for long periods).
- IC designs are constrained by cost and are thus minimalist implementations
- Air Interface protocols are designed to reduce tag complexity
- Design flaws in reader implementations due to cost constraints

The reasons above form the basis from which various weaknesses have arisen in low cost RFID systems. The resulting vulnerabilities are examined in detail in the following sections.

3.1 Eavesdropping and Scanning

Transmissions from a reader and a tag take place over a clear communication channel which may be observed by a third party. Low cost labels with minimum functionality are only able to identify themselves by transmitting a unique identifier, and these labels can be read by any reader adhering to the air interface protocol used by an RFID tag. Hence a third party may monitor a conversation between a label and a reader to obtain sensitive information. Illicitly obtained information in this manner may be used to create fraudulent labels, unauthorized readers, or used to discover secret information stored on labels (such as a tag password).

Similarly, competitors of an organization (such as a rival supermarket) may, over time, scan another organizations inventory labeled with RFID labels or eavesdrop on the organization's own valid operations to obtain valuable information, such as sales data, to ascertain the performance of its competitors (an act commonly referred to as corporate espionage). Publications such as [14] have attempted to define various eavesdropping ranges based on the reading ranges of tags. Similar descriptions of the eavesdropping distances possible are stated below so that vulnerabilities of eavesdropping can be better understood. However, it should be stated here that, while it is useful to define terms to explain ideas, the fact that a third party can eavesdrop on a conversation between a tag and reader from a distance still remains a fundamental vulnerability.

Figure 4 illustrates a simple model for a passive RFID communication channel. It is possible to consider the distances at which a third party can listen to a conversation between a tag and an interrogator to formulate the following general classification of eavesdropping distances. Figure 5 gives an illustration of the latter distinctions discussed and explained below.

Operating range: Tag operating range is either defined by product specification based on user requirements or it may be based on a certain standard. The operation range of a tag will also be application dependent as tag reading distances are affected by various environmental factors. The operating range of tags is then the maximum distance at which any given tag will read to a given reliability, when illuminated by a reader operating under the electromagnetic compatibility regulations of that region for a particular frequency band of operation.

Backward channel eavesdropping range: The backward channel refers to the communications sent from a tag to a reader. In a low cost system where the tags are passive, this reply is weaker in signal strength than a reader transmission as it is achieved by reflecting some of the incident RF energy at the label antenna.

The backward channel range is generally much greater than the operating range of the tag since a third party is capable of using a narrow beam antenna with an RF receiver of higher sensitivity and because the third party does not have to use the same antenna for powering and receiving (unlike most low cost systems which use a monostatic antenna configuration; that is, a single antenna for powering, transmitting and receiving).

Forward channel eavesdropping range: The forward channel refers to the transmissions from a reader to a tag, and the forward channel eavesdropping range is the maximum distance at which a third party with a high gain antenna and a highly sensitive RF receiver can correctly record a tag transmission.

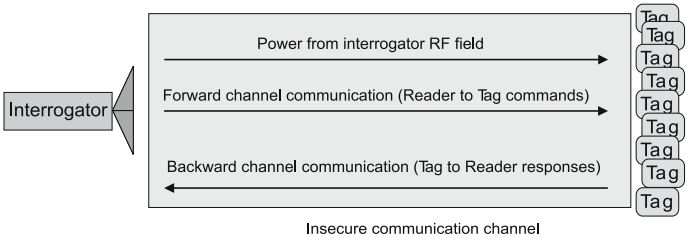


Fig. 4 A passive RFID communication channel model.

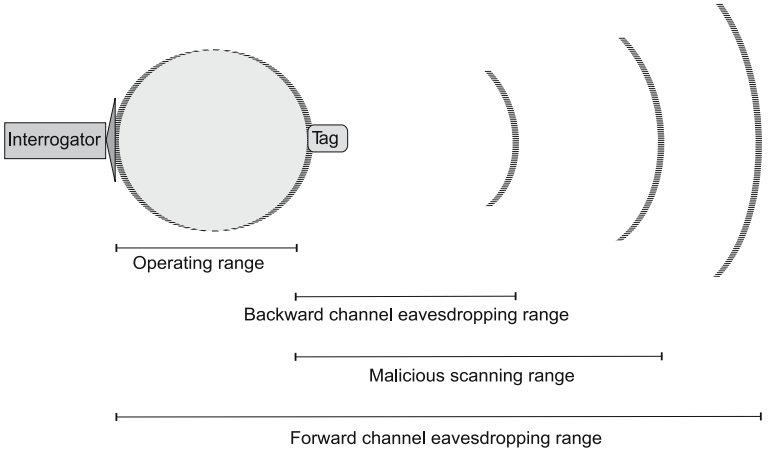


Fig. 5 Eavesdropping range classification (Damith C. Ranasinghe and Peter H. Cole, "Confronting Security and Privacy Threats in Modern RFID Systems", 40th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, California, 29 Oct. –1 Nov. 2006. © 2006 IEEE).

Malicious scanning range: This read range is derived by considering an adversary with no regard for electromagnetic compliance or standards and whose only intention is to both power and read the tag at any cost or to eavesdrop on a conversation between a tag and a reader at any cost. Combined with the prospect of a highly sensitive RF receiver, a narrow beam antenna and the willingness to break electromagnetic regulations, malicious scanning range will have reading distances in excess of that possible for backward channel eavesdropping.

There are generally two forms of eavesdropping possible with low cost RFID systems; passive eavesdropping and scanning (active eavesdropping). The following sections will discuss the previously mentioned forms in an RFID context.

3.1.1 Passive Eavesdropping

As the names suggests, passive eavesdropping relates to the observation and, or, recording of communication between a reader and a tag by an unintended recipient. Passive eavesdropping may be performed by a third party in the operating range, the backward channel eavesdropping range or the forward channel eavesdropping range.

3.1.2 Scanning (Active eavesdropping)

In this situation, a third party or an adversary is actively attempting to read the contents of a tag without the authority of the tag owner. In a scanning scenario with respect to a low cost RFID system, an adversary is using a rogue reader to power the tag and communicate with the tag without raising the suspicions of the tag owner. An active eavesdropper will have a working range within the malicious scanning range outlined in Figure 5.

3.2 Cloning

Devices designed to impersonate tags or readers (imitating the behavior of a genuine label or a reader) present a serious threat to an RFID system. Impersonation will add a new dimension to thieving as attackers are able to write EPC data onto devices that function like RFID tags. A direct consequence of cloning is the possibility for counterfeiting, where a genuine article tagged with an RFID label may be reproduced as a cheap counterfeit and tagged with a clone of the authentic RFID label. The 'track and trace' concept outlined in Section 2.10 is one possible solution to detecting cloning in a supply chain application.

At the time of writing, there is no mechanism for a reader to verify that it is communicating with a genuine RFID label and not a fraudulent label. Thus a thief may replace a tag of a valid item with a fake tag or replace the tag of an expensive item with that of a fake tag with data obtained from a cheaper item. Hence the lack of a means for authentication allows an adversary to fool a security system into perceiving that the item is still present or fool automated checkout counters into charging for a cheaper item. Such fake labels may also be used to create imitation items. There is presently no mechanism for a reader to authenticate itself to a label or a label to authenticate itself to a reader. Thus labels and readers are constantly in a vulnerable environment where the integrity of messages is doubtful and there are no means for establishing the legitimacy of a reader by a label or the legitimacy of a label by a reader.

Clearly, more expensive RFID system implementations are also not immune from cloning as shown by a recent cloning attack published in [15] where a cloned tag was used in the purchase of fuel at a service station and to start an automobile locked with an RFID based car immobilizer. A similar example of cloning of proximity cards is given in [16] while the possibility of cloning the VeriChip [1] in a discussion of its possible use to tag employees was outlined in [17].

The EPC Class I tags have no mechanism for preventing cloning as the tags are simple bit storage devices that transmit a string of bits on request from any valid reader. All that is required by an adversary is to scan a tag and copy its EPC number onto another tag or another device that is capable of impersonating a tag. The EPC Network architecture aims to remedy the problem by creating an electronic history of the product's life through the supply chain by way of an electronic pedigree. While the access to an electronic pedigree of a product by way of a secure database only solves the problem of confirming the existence of an illegal clone, it may not be possible to distinguish the original tag from its illegal duplicate(s).

3.3 Man-in-the-Middle

An RFID system is constantly under threat from man-in-the-middle attacks resulting from eavesdropping on reader and tag transmissions. A third party may monitor a conversation between a label and a reader, record and use or alter parts of the conversation and retransmit messages to illicitly obtain information from RFID devices or to command RFID devices to the detriment of the system.

Retransmission of such recorded information may be used to query RFID labels or fool RFID readers. Such an attack has the ability to fool personal access control systems and contactless payment systems based on RFID technology [18].

For instance the EPC C1G2 protocol uses a "kill" command [5], protected by a password, to disable the label so that it cannot be read. It is possible for a third party to record the conversation between a reader and a tag that performs a kill operation and use that information to kill other tags provided that they are protected with an identical kill password.

3.4 Denial of Service

An adversary may initiate a denial of service (DoS) attack to bypass or avoid security systems. A DoS attack is easily carried out by placing a large number of fake labels for identification by a reader. Persons also have the ability to disrupt an RFID system implementation by destroying or corrupting a large batch of labels. Labels are also vulnerable to protocol attacks. DoS attacks may also be performed by exploiting weaknesses in the air interface protocol or weaknesses in the implementation of a tag's finite state machine. A simple scenario of such an exploit may involve labels being repeatedly asked to perform an operation, thus making them unavailable to an authorized reader.

In addition, tags may be prevented from being read by using the simple concept of a Faraday cage or by jamming the RFID interrogator signals, for instance by intentionally creating noise in the frequency band in use. For critical applications, a DoS attack may have devastating effects.

3.4.1 Code Injection

RFID interrogation signals can be disrupted or blocked, or RFID readers can be attacked using RFID tags designed to manipulate weaknesses in the air interface protocol or the implementation of the reader to create a denial of service attack during an RFID interrogation process by creating situations of system unavailability. The possibility of RFID viruses has been highlighted in [19] where a more sophisticated tag may exploit interrogator or protocol vulnerabilities to affect a number of systems by using a reader to cause a system failure by way of a code insertion attack caused by creating a buffer overflow in a reader's memory stack using carefully constructed SQL instructions disguised on the tag as the data that gets transmitted to an RFID reader. This vulnerability is only present if the middleware is intentionally made vulnerable by accepting any data transmitted by the tag without checking for validity of the format of the data sent. Also, modern SQL servers are guarded against malformed SQL instructions.

3.4.2 Communication Layer Weaknesses

The recently ratified EPCglobal C1G2 air interface protocol [5] has a number of security features based on the use of tag specific passwords. Probably the most

important feature that is protected is the *KILL* command by using a kill password. There is also a means for access control on the tag using an access password.

A recent publication in [20] has shown how the kill password of a tag can be deduced by the careful analysis of the tag power consumption to a series of well constructed test passwords. This highlights a particular vulnerability of low cost tags to power analysis attacks and the vulnerabilities of storing long term secret information on a tag. However, it is possible to prevent such an attack as power analysis attacks have been well studied in the context of smart card devices. The RFID ICs in the future will need to be designed to avoid such an attack but this will take place at added cost to an RFID tag.

While power analysis attacks may be prevented in the future, the fact that each RFID tag has at least two unique passwords will create both potential security and logistical nightmares if the problem of careful key management is not considered. This problem will be aggravated in the future as item-level tagging begins to proliferate through the global supply chains and PoS (point of sale) devices may need real time access to passwords as consumers purchasing goods may want their tags deactivated at the point of sale. Hence the problem of careful key management needs to be considered in the context of low cost RFID systems where the potential for key discovery is highlighted by the global aspects of supply chains. It is not difficult to imagine a scenario in the future where a list of kill passwords anonymously appears on a public web site.

The recently ratified C1G2 protocol also relies on the tag generating a random number to be used as an input to an exclusive-or operation. The risks associated with inefficient or inadequate random number generation in RFID tags (that is, a high correlation between the random numbers, in a pseudo-random number sequence) is emphasized in [21]. The consequences are two-fold for tags using the C1G2 protocol. Primarily, the lack of randomness may cause particular tags to respond with an identical time slot during the execution of the slot selection process. Thus an attacker may be able to track a tag depending on the time slot it selects in a seemingly random manner. Since the security of the information sent to a tag relies on the randomness of the number that the tag generates, a lack of randomness may allow an adversary to easily decrypt information transmitted, once the attacker successfully decrypts an encrypted message, or discover the seed used in the pseudorandom generator.

Though with great technical difficulty, [21] also points out the possibility of identifying a tag using a "radio fingerprint". For instance, manufacturing variations that may cause physical glitches in the signals may be used to distinctly identify tags. In such a situation, even strong cryptography protocols and primitives would be ineffective.

3.5 Physical Attacks

In addition, the labels themselves are exposed to physical attacks due to the absence of tamper proofing as dictated by the cost limitations of low cost tags. Physical attacks are possible irrespective of whether measures are in place to protect

labels. However, the ability to gain useful information from a protected label is a much more difficult problem. A physical attack on an RFID label or a reader may yield an adversary secret information, such as passwords (in C1G2 labels), providing security to an RFID system. The importance of physical attacks is more prominent in cases where RFID tags are used as a means of authentication. The problem is compounded when a physical attack leads to the construction of a clone.

An insight into physical attacks can be gleaned from an increasing body of work in the area of smart cards. A complete overview of possible physical attacks and countermeasures is outlined in [22] while specific lower cost physical attacks are presented in [23].

The majority of physical attacks possible on devices in general can be bundled into two broad categories based on the means used for accessing the device. These attacks are relevant to RFID devices, especially since they have no tamper protection to safeguard label contents.

3.5.1 Non-Invasive Attacks

These attacks are as a result of timing analysis, power analysis, analysis of certain glitches (radio fingerprinting), and exploitation of data remnance. Non-invasive attacks are low cost and require little expertise to execute. While non-invasive attacks are generally thwarted by increasing chip complexity in most devices, it is not the case with RFID chips with minimalist implementations that may have design flaws as a result of human errors or insufficient error checking. Non invasive attacks are particularly dangerous as there is no physical evidence and the owner of the tag may not be aware that such an attack has taken place.

3.5.2 Invasive Attacks

In addition, an adversary may simply reverse engineer labels to create fraudulent labels for cloning or DoS attacks or use probing techniques to obtain information stored in memory (microprobing and Focus Ion Beam editing) or alter information stored in memory (using a laser cutter microscope [23]). A recent exploitation by reverse engineering of a more costly implementation of an RFID device with added security to carry out a fraudulent payment was published in [15]. Use of microprobe needles to read out the memory contents of a smart card is published in [24].

Attacks such as optical probing and fault injection attacks where the chip is removed from its packaging with the passivation layer still unbroken, are also invasive attacks but these attacks are may be further qualified as semi-invasive attacks.

3.6 Privacy Violations

The mass utilization of RFID labeled items creates an imminent and potentially widespread threat to consumer privacy. The privacy issues raised by RFID labels

have been receiving a wider audience as a result of the popular press. The mass movement by civil libertarians has seen RFID trials cancelled [25] (despite misunderstandings of the company's intentions [26]) and negative press coverage for other manufacturers causing delays in RFID test trials [27]. Press coverage on privacy issues has also managed to tarnish the image of RFID by personifying the technology as satanic and associating nicknames such as "spy-chips" [28, 29] to infuse fear among consumers.

It is possible to imagine various scenarios of privacy violations and most of those are already existing concerns from technologies such as credit cards, browser cookies, mobile phones and Bluetooth devices. However, RFID, due to artifacts resulting from its cost constraints, presence of a unique identifier readable by anyone, and the encoding of product information on the unique numbering scheme such as the EPC, creates two possible scenarios; profiling and, tracking and surveillance, where the privacy of people as well as corporations may be infringed. These scenarios are discussed in the following sections.

3.6.1 Profiling

There are clear possibilities for unauthorized interrogators to read label contents from unprotected RFID labels due the lack of a mechanism for authentication and the fact that low cost RFID labels as well as interrogators broadcast unique item identifiers such as the EPC. Even if labels are protected, a traffic analysis attack (or predictable label responses) may be used. Hence an individual with a number of labeled items may be scanned by a third party to identify individual possessions or "taste", and specific EPC numbers on products may then be associated with an individual.

The data obtained can be misused to violate an individual's wishes to remain anonymous. For instance, persons carrying religious material or material related to a certain political affiliation, may no longer be able to privately pursue their beliefs or interests and in addition to their reading material potentially becoming public knowledge, their beliefs and opinions may be used in acts of persecution, jealousy or hatred. At the same time, data collected and associated to individuals can be valuable to market researchers or even thieves in search of wealthy victims. The personal information collected regarding individual preferences will act as a powerful tool for marketing products as more targeted marketing to individual tastes and affordability becomes possible by scanning the RFID tagged possessions of an individual.

It is possible to imagine a variety of plausible ways of using such information. For instance, if Bob purchases a brand named jacket using a credit card, the shop can immediately associate "Bob" with the tag id of the apparel. When Bob enters the store again, the shop has the ability to automatically establish his identity along with a history of his spending habits and tastes. While this information may prove positive for Bob, Alice, who might enter the same store, may be wearing cheap shoes and the shop assistants then have the ability to provide preferential treatment to Bob while perhaps neglecting Alice. Similarly, a thief hiding in the corner of the store may read the tag id of Bob's jacket and conclude from the tag id (by way of careful observation and without having access to any backend

databases) that the tag id is indicative of an expensive apparel, then Bob might become the unfortunate victim of a theft.

3.6.2 Tracking and Surveillance

A further privacy concern resulting from the association of unique identifiers to individuals and the unobtrusive scanning of RFID labelled items carried by an individual is posed by the possibility of tracking, albeit with technical difficulty. Correlating data from readers obtained from multiple locations can reveal the movement, social interactions or financial transactions of an individual once an association is made between a unique tag identifier and a person. In response to such concerns, there have been suggestions to remove the unique identifier in an EPC to prevent a specific EPC from being associated with an individual. Even if such a scheme is implemented, individuals may be tracked through a "constellation" of predictable label responses. Hence, a person's unique taste in items may betray their location, movements, or identity.

4 Addressing Vulnerabilities

Issues resulting from vulnerabilities discussed in Section 3 can be divided into two broad categories of security related issues (exemplified by eavesdropping, cloning, man-in-the-middle, DoS, communication layer weaknesses and physical attacks) and privacy related issues (profiling and, tracking and surveillance). Overcoming these seemingly divergent issues can be achieved by the provision of services to enforce measures to address both the privacy and security related issues. These services can be implemented on low cost RFID systems by identifying existing mechanisms, inventing new mechanisms or by re-engineering existing mechanisms to meet the required security and privacy objectives.

However, there is a notion among the advocates of RFID technology that the general nature that is partly hindering the mass scale deployment of RFID technology − that is, the unreliability of low cost systems mainly due to the reasons given in Table 2 − makes the exploitation of vulnerabilities such as profiling and tracking discussed in the previous section impractical.

Clearly, low cost RFID tags are unreliable. For instance an RFID tag placed on your jacket may work while it is on the shop shelf, but it may stop working once the jacket is worn as your body will affect the properties of the RFID tag antenna.

It is due to the reasons given in Table 2 that some of the vulnerabilities discussed in Section 3, in practice, are far from being feasible. Ironically though, the unreliability of RFID tags has prevented much of the security and privacy violations from being realized, with the possible exception of laboratory experiments or in that realm of possibility. While this is the present reality of low cost RFID technology, it is expected that the cost benefits of RFID technology will eventually propel the research community to solve the technical issues outlined above. Hence the idea of using unreliability to brush aside the possible threats is not a long term solution.

Table 2 Sources of unreliability.

	Description
1	Effects of metal and liquids on the propagation of electromagnetic waves
2	Effects of permeability of materials on tag antennas
3	Interference and noise from other users of the RF band
4	Tag orientation with respect to the reader propagation field
5	Distance of the tag from a reader
6	Electromagnetic compatibility regulations
7	Cost and power constrained implementation of RFID chips

Generally, it is clear that the technology of tomorrow is what is being developed currently. Even though deployments of current RFID technology do not adequately satisfy expectations, despite various mandates for RFID compliance, it is gradually beginning to proliferate [30, 31, 32]. Hence, it is important to address vulnerabilities discussed in Section 3, despite some being implausible, so that the systems deployed today do not become problems of tomorrow.

The following sections consider the measures required for addressing security related issues and privacy related issues identified in Section 3.

5 Addressing Security Issues

Eliminating security related concerns regarding RFID systems illuminated by way of the examples in Section 3 require the enforcement of suitable security measures. Before deciding on a set of security measures, the security objectives that need to be satisfied must be identified. Table 3 lists a necessary set of security objectives that will be required to address the potential security threats.

RFID systems must employ mechanisms to achieve one or more of the above security objectives to alleviate various concerns cited in Section 3. As security cannot be solely accomplished by security mechanisms, it should be mentioned that proper legislation, procedural techniques and enforcement of laws is also required. The following sections describe the security objectives outlined in Table 3

Table 3 List of security objectives.

	Security Objectives
1	Confidentiality
2	Message content security
3	Authentication
4	Access control
5	Availability
6	Integrity

and demonstrate that meeting these security objectives eliminates the security threats posed by inherent weaknesses in low cost RFID systems.

5.1 Confidentiality

The term 'confidentiality' can be used to describe a mechanism to keep information from all but those that are authorized to see it [33].

In an RFID system, the communicated information between a reader and a tag needs to be confidential when sensitive data such as secret keys or other such information, which must not be collected by an eavesdropper, is communicated. The confidentiality of any secret information stored on a tag is also at risk and needs to be secured.

Confidentiality may be achieved by having the communication link between tags and readers encrypted, thus establishing a secure communication link. Confidentiality of tag contents may be achieved by tamper proofing the tag to prevent physical access to tag contents. Currently however, there is no secure means of establishing a secure communication link between a tag, and tamper proofing a tag has cost implications that will hinder the economics of low cost RFID technology.

5.2 Message Content Security

Providing message content security or data integrity involves making certain that the data contained in a communication is not altered by unauthorized or unknown means [33]. Alteration in an RFID context may involve the capture, substitution, deletion or insertion of information and the retransmission of that altered information to a reader or to a tag. Ensuring message content security will prevent man-in-the middle attacks involving the retransmission of altered messages. Present low cost RFID systems have no means of providing message content security.

5.3 Authentication

The simple objective of meeting authentication can be expressed as authenticating the devices involved (the tags and the reader) or, in a supply chain application where the tags are used to label products, as product authentication. In some applications where perhaps the tag is an integral part of the tagged object, authentication of the tag may be adequate to guarantee the authenticity of the object to which it is associated. In other applications where tags are placed as an external label to a high value item, authentication of the tag alone may not be adequate. The objectives of tag and interrogator authentication and, product authentication are discussed below.

5.3.1 Tag and Interrogator Authentication

In an RFID context, authentication simplifies to the corroboration of the identity of a tag or a reader. Authentication is an important RFID security measure for preventing counterfeit manufacture or substitution. It is also important for controlling access to label contents. Use of authentication may also be required in other applications of RFID technology such as baggage reconciliation or secure entry systems. Authentication of a tag is useful in addressing vulnerabilities posed as a result of cloning.

5.3.2 Product Authentication

While authentication described above has the objective of establishing that a tag is legitimate and a reader is authorised, in certain application use case scenarios, authentication of the tag is not sufficient to guarantee the authenticity of the product to which the tag is attached as brand or goods substitution may have taken place. Hence in the case of using a low cost RFID tag to label a product, product authentication refers to the establishment of the authenticity of a product by the corroboration of the identity of a tag and/or the legitimacy of the product by creating an irrefutable link between the product and the tag that can be verified by a third party.

5.4 Access Control

In the context of interaction between RFID interrogators and tags, access control implies a mechanism by which a tag or an interrogator grants access or revokes the right to access some data or perform some operation. Generally, tags will require access control mechanisms to prevent unauthorized access to tag contents.

5.5 Availability

Ensuring availability in RFID systems is an important issue since readers need to be ready to detect tags that may enter their reading range at ad-hoc intervals of time (depending on the application). In an RFID context, availability applies to ensuring that the services offered by a reader to an RFID tag or the services offered by a tag to an RFID reader are available when expected [34]. RFID systems meeting the availability criteria will ensure that there are services in place to thwart or prevent a DoS attack.

5.6 Integrity

Integrity of an RFID system applies to the integrity of the devices, such as the reader and the tags where it implies that a reader or a tag has not been malevolent-

ly changed. A reader receiving data from a tag needs to be able to trust that the information received is correct, while a tag needs to be able to trust that the information it receives from a seemingly authentic reader is trustworthy [34]. Ensuring the integrity of a system is an important consideration in addressing physical attacks.

6 Addressing Violations of Privacy

While it is difficult to define privacy, and a number of different interpretations can be found, it can be most simply stated as the *interests* that a person or persons have in "sustaining a 'personal space' free from interference by other people and organizations" [35]. The ideas captured by *interests* that a person has in an RFID context can be further elaborated as given below in Table 4 [35].

Table 4 An elaboration of privacy.

Privacy Interests	Description
Privacy of personal behavior	As the name suggests, privacy of behavior encompasses all aspects of a person's manner. In reality, this is narrowed down to areas that are sensitive to individual people such as political activities, sexual orientation or religious conduct.
Privacy of personal data	Personal data privacy refers to the more commonly used term, data privacy. In essence, data associated with a person should not be accessible by a third party without the consent of that individual. This applies to cases where the data is collected, or processed by a third party.

It is not possible to describe the number of privacy violations RFID technology can potentially cause, since they are numerous as described in [36]. However, it is sufficient to realize that the root cause of such violations stems from the potential to automatically associate human identification information with object identification information and thus addressing privacy requires certain goals to ensure that the latter association is not possible. Privacy goals outlined in Table 5 are an adequate set of goals for addressing the issue of associating object identification data with human identification data and the related concerns outlined in Section 3.

It is important to note that privacy is a multi dimensional issue involving many areas. The successful implementation of the privacy objectives outlined in Table 5 will not only require security mechanisms but will also require the formulation of public policies, legislation and the enforcement of the law by the relevant law enforcement agencies. The latter statement is especially important in order to ensure privacy of personal data [37, 38].

Public policy is a vital aspect because the security mechanisms used to ensure privacy are most effective when implemented in conjunction with a well-formed policy. There are existing privacy policies that can be applied directly in the context of RFID [39]. However, these may need to be clarified, refined or amended to cover aspects specific to RFID systems. Significant issues that must be dealt with by policy

formulation or amendment in relations to RFID are those generated by the following items.

- Unique Identification of all label items
- Collection of information (who collects data generated from RFID systems, how do you exploit that data, ownership of information obtained from the data)
- Dissemination of that information
- Mass utilization of RFID technology

It is important to note that existing barcode systems have many of the same risks; they can be read by a simple bar code reader, can be destroyed easily and can be cloned. However, there is not the potential for these operations to be performed wirelessly and unobtrusively on an immense scale.

Table 5 List of privacy objectives.

	Privacy Objective
1	Anonymity
2	Untraceability (Location privacy)

While public policy and legislation is an ongoing topic of discussion, it is beyond the scope of this paper to address policy tools and legal tools for addressing security and privacy issues. Nevertheless, technical solutions for addressing previously mentioned issues are considered in the Networked Based Solutions section and Cryptographic Solutions section of this book. The following sections discuss in detail the privacy objectives introduced in Table 5.

6.1 Anonymity

While anonymity can be described in a number of ways, the most appropriate is probably the concealment of the identity of a particular person involved in some process, such as the purchasing of an item, visit to a doctor or a cash transaction [35].

Mitigating the problem of anonymity in an RFID context will involve the prevention of associating an EPC of an item with a particular individual as the EPC can be used to obtain information regarding a particular process, or an object, and that information may be associated with a particular person's identity.

For instance, a person walks into a book store, purchases a book of their choosing and pays for the purchase using a credit card. Immediately, this transaction allows a relationship to be created between the identity of the individual and the EPC of the book. The person may then walk on the street, now it may not possible to conceal their identity with regards to the purchase from a third party scanning the book's RFID tag, provided that the third party has access to the relationship between the object identification information and the human identification information. The same person may carry an expensive medication which could be scanned by thieves or by potential employers to his or her detriment.

6.2 Untraceability (Location Privacy)

Untraceability in an RFID context is aimed at addressing location privacy issues. Location privacy is an issue that has surfaced more recently with the availability of reliable and timely information about the location of people as a result of pervasive computing. It is also an issue associated with mobile users and other users of wireless devices. While this is not an issue specific to RFID [40], it does apply to modern RFID systems that are being developed because of their pervasive nature and their ability to leverage the Internet to form a global network that can receive and transfer data in real time.

There are number of ways of defining untraceability and, in an RFID systems environment, it can be stated as a means by which the ability of other parties to learn or track the location of people or transactions from a current or present location, based on information obtained from one or many RFID tags in the possession of that person(s) or party to that transaction, is prevented.

Hence, providing untraceability in an RFID system requires the provision of a mechanism to prevent other parties from obtaining RFID tag data without the tag owner's consent and/or to prevent the association of an EPC of an item with a particular individual and/or to prevent tags from emitting any kind of a unique identification signal when performing a tag query by an authorized reader. Hence a mechanism is required by which a person can hide his or her true identity from devices that scan our personal RFID tags while still being able to take advantage of the benefits of RFID for the consumer.

7 Cryptography

Achieving the security and privacy objectives outlined in Table 3 and Table 5 respectively, require an enormous anthology of technical and legal tools. While legal tools are not considered in this paper, the required technical tools may be provided through cryptography. The following sections of the paper consider cryptography, the science from which a plethora of technical tools for providing services to achieve the privacy and security objectives identified previously can be obtained.

Cryptography is defined as the study of mathematical techniques related to aspects of information security in [33]. However, cryptography is not the only mechanism by which information security may be provided.

Security and privacy issues concerning RFID may be solvable using a set of security mechanisms derivable from various cryptographic primitives. A security mechanism is a collective term used to refer to a combination of cryptographic primitives and protocols used to provide security. Hence, it is appropriate to briefly consider the subject of cryptography in the following sections to examine the range of cryptographic tools available for various applications, the level of security provided by such primitives and a simple classification of the vulnerabilities of various security mechanisms.

7.1 Cryptographic primitives

Cryptography is an ancient art that has been used throughout human evolution to provide security and to protect the privacy of individuals or organizations. Providing security and privacy for RFID systems will inevitably involve using some cryptographic primitive already in existence, or newly defined, along with suitable protocols that take into account the unique nature of RFID systems. Table 6 gives a classification of a broad range of cryptographic primitives. A more complete description of these primitives can be found in [33, 41, 42].

Most modern cryptosystems, such as the RSA cryptosystem (with a few exceptions such as one-time pads), are based on some mathematically hard problem and the level of security provided by the system will depend on the difficulty of the mathematical problem.

It is important to define the difficulty of a problem before the level of security provided by a cryptographic system can be discussed. A mathematical problem is said to be difficult if the time it takes to solve the problem is immense compared to the size of the inputs to the problem. Modern cryptographic systems are based on mathematical problems where the fastest known algorithm takes exponential time to find a solution. This implies that the time taken to solve the problem increases exponentially as the size of the inputs to the problem increases linearly. Thus the level of security provided by a cryptosystem is often expressed as the number of operations required to break the cryptosystem or the time taken. Generally, the level of security provided by a cipher complements the commercial value of the information protected by the cryptographic system. A discussion on quantifying the security provided by a security mechanism is considered in Section 7.3.

While there are numerous cryptographic systems in use based on various primitives as outlined in Table 6, all such systems are not without their own set of weaknesses. There are specific attacks on any cryptosystem or protocol employed by a security mechanism to provide security. These weaknesses are a result of certain vulnerabilities in the cryptographic scheme or due to certain flaws that may have entered into the protocol employed in the security mechanism. A classification of attacks on cryptographic systems in general can be found in [33, 41, 42] and is summarized below.

Table 6 Classification of cryptographic tools.

Primitives without keys	Symmetric key primitives	Asymmetric key primitives
Hash Functions	Symmetric key ciphers (Stream ciphers and Block ciphers)	Public key ciphers
One way permutations	Hash functions	Digital signatures
Random sequences	Digital signatures	Identification primitives
	Pseudorandom number sequence generators	
	Identification primitives	

7.2 Classification of Attacks

Cryptanalysis is the art of recovering the plaintext of a message without the key by using an algorithm to infer the plaintext from a given ciphertext or by deducing the key so that ciphertext can be decrypted to obtain the plaintext. An attempt made by an adversary at cryptanalysis is termed as an attack [41]. The following sections consider the possible attacks on cryptographic primitives and protocols to defeat a security mechanism.

7.2.1 Attacks on Cryptographic Primitives

There are various forms of attacks possible on cryptographic primitives, with various names. However, the most common forms are outlined in below (refer to [33, 41, 42] for more details).

- Ciphertext-only attacks
- Known plaintext attacks
- Chosen plaintext attacks
- Adaptive chosen-plaintext attacks
- Adaptive chosen ciphertext attacks

7.2.2 Attacks on Protocols

Similar to attacks on cryptographic primitives, there is a vast array of attacks on the protocols used and the number of attacks has grown with the emergence of new protocols. The following is a summary of a prevalent list of possible attacks (refer to [33, 41, 42] for more details).

- Replay attacks
- Known key attacks
- Impersonation attacks
- Dictionary attacks

7.3 Level of Security

Many cryptographic systems have been broken because of increased computational resources, development of faster and better algorithms or problems which are proven to be easier than when they were first conceived. This is the reality of any cryptographic system. However, the concerning issue for modern cryptosystems is not that the system will eventually be broken, but that the range of possible attacks on a security mechanism to breach security and the time taken to break the security system using the best possible attack

It should be noted here that, in general, the security of a system is difficult to quantify. The usage of the term, ‘level of security,’ is generally used to refer to the number of operations required or the amount of time taken to break the security of a given system using state of the art technology and the best available algorithms.

Table 7 Defining levels of security.

Level of Security	Description
Unconditional Security	A cryptographic system is described as being unconditionally secure if the security of the system cannot be broken if an adversary is given unconditional resources. An example of an unconditionally secure encryption system is the one-time pad.
Computational Security	Computational security is given as a measure of the amount of computational work required, using the best available methods, to defeat the security of a system. A system is considered computationally secure if the amount of computer resources or time required to break the system is far more than that available to an adversary considered in the analysis of the system. Computation security is also termed Practical security [78].
Ad-hoc Security	Systems are classified as having had-hoc security when postulations are made using any number of apparently convincing arguments that all possible attacks on the system require a level of resources (computational and time) that are beyond the level of resources available to some hypothetical adversary.
Provable Security	A system is described as having provable security if it can be shown that breaching the security of the system involves evaluating the solution to a problem belonging to a class of problems (such as NP-hard problems) that can not be calculated in polynomial time, such as the integer factorization problem or the discrete logarithm problem. As pointed out in [78] it should be noted here that the proof here is that the given problem is at least as difficult as an existing problem, and not an absolute proof of security.

However, it is possible to evaluate the security provided by certain mechanisms and describe them using a number of classifications, some of which are published in [33, 41, 23]. Terms used to describe the level of security of a system are outlined in Table 7.

8 Low Cost RFID and Cryptography

The plethora of available security primitives are too excessive in terms of cost to be implemented on a cost constrained RFID chip. Low cost labels are also not self-powered and only consist of limited logic functionality, unlike smart card processors. However, they may be more suitable for higher class labels with a greater opening price point. For instance, private key cryptosystems such as AES are not suitable since a commercial implementation of AES typically requires 20,000 – 30,000 gates [43]. This is far more than the number of gates on an entire low cost label. However the SHA-1 specified by the US Department of Commerce is a possible candidate for an encryption rule but hardware implementations of SHA-1 are currently too costly to meet the cost budget of low cost RFID labels [44]. Cryptographic systems and protocols need to fit into a label footprint without dramatically increasing the cost of a label.

Considering cryptographic solutions for RFID requires a careful understanding of low cost RFID, underlying assumptions of the system, limitations and expectations from the end user community. There are particular challenges that need to be considered as a result of the nature of low cost RFID systems. These challenges are discussed in the following section.

8.1 Challenges

Challenging aspects to providing security and privacy for low cost RFID systems using traditional cryptographic mechanisms and existing hardware are outlined in Table 8. Each of the listed constraints needs to be considered before designing a practicable security or privacy measure.

It is evident from the description of low cost RFID systems provided in Section 2, and their associated implementation in supply chain applications as Class I and Class II tags, that the main constraint hindering the adoption of more traditional cryptographic solutions is the scarcity of hardware resources as a result

Table 8 Challenges facing the implementation of strong cryptosystems on low cost RFID (Modified from Damith C. Ranasinghe, Daniel W. Engels, Peter H. Cole "Low cost RFID systems: confronting security and privacy", 2005 Auto-ID Labs White Paper Journal Volume 1, © 2005 Auto-ID Labs).

Challenge	Description
Cost	Minimizing cost implies limited memory and silicon area constraints. Tag costs are expected to be less than 5 US cents. The cost of tags has reduced over the years and the trend is expected to continue thanks to Moore's Law. This has two implications; more hardware intensive cryptographic functions will slowly enter low cost RFID chips and the cost of an RFID will continue to decrease. Unfortunately, analogue devices fabricated on IC's do not scale in the same manner as the digital devices so RF front end on chips will still remain a cost factor.
Regulations	Transmit power restrictions, spectral masks, frequency of operations, available bandwidth, and time available for computations.
Power consumption	Important to minimize the power consumption of the label IC circuit to gain maximum performance. A cryptographic device which is the highest consumer of a passive chip's power will adversely affect its performance as it will reduce a label's read range.
Performance	Label performance and system performance goals (data transmission rates, number of label reads per second, percentage of correct reads). Performance goals also place a limit on the time available for any computations by cryptographic hardware. Cryptographic systems requiring access to backend systems will need to take into consideration network delays associated with a security mechanism as such delays will affect system performance.
Power disruptions	Sudden loss of power is a practical reality and any security mechanism should not leave the chip in a vulnerable state during such an event.

of cost limitations. Nevertheless, cost is not the only limitation. There are many other such restrictions and difficulties that result as a consequence of the nature of electromagnetic waves and the constraints placed by end users and electromagnetic compatibility regulations.

As outlined in Table 8, EM regulations pose restrictions on the isotropic radiated power at stated distances. This implies that there is a maximum limit on the power available at a given label distance from a transmitter. Thus, passive labels with size limited by a particular label class or an application receive power from a stated power flow per unit area. The power available to the label is one factor contributing to the determination of the type of security scheme and the cryptographic hardware used in a label. Cryptographic hardware consuming considerable power (in the range of tens of microwatts) will significantly diminish the label reading distance and degrade the performance of the whole RFID system implementation. Furthermore, a security mechanism employing a memory write will have to account for the additional power required to operate a labels E^2PROM.

The power utilization of any security related hardware should not exceed the typical tag power consumption of 10–15 microwatts required for writing to a passive RFID label, as explained in Section 2.3. Ideally, the power consumed should be a fraction of this value for any security related hardware to be viable as considerable power requirements will constrain the label performance by limiting the operating range of the label. However reducing power consumption of any encryption hardware is a challenging prospect.

Power dissipation in integrated circuits is a function of many factors; the fabrication technology, the layout of the design and the scale of the fabrication process. Static CMOS technology is very attractive in low power devices due to the almost negligible power consumption in steady-state operation. Power dissipation in CMOS circuits is mostly due to the charging and discharging of capacitances during dynamic operation. Power consumption can be reduced by the proper choice of circuit, logical or architectural structure. This might come at the expense of silicon area, which is critical to controlling tag costs.

The power consumption in static CMOS circuits is due to the static (or steady state) power consumption and the dynamic power consumption (power consumption during the switching of logic levels). The dominant power dissipation in CMOS circuits is caused by the switching of logic levels, while the static power consumption due to the leakage of current flow through the reversed-biased diode junctions in the transistors is almost negligible.

Equation (1) illustrates the total power consumption of a device i while (2) expresses the static power dissipation as the leakage current I_{static} and the supply voltage to the device V_{dd}. Equation (3) formulates the power consumption during $f_{0\to1}$ switching operations (logic $0\to1$ and $1\to0$ transition) per second, where C_L represents the sum of the intrinsic capacitance (junction capacitance and other parasitic capacitances) and the extrinsic load capacitance (due to the wires and connecting gate) of the device [45].

It is clear from (3) that higher throughput from a device leads to more frequent signal transitions and results in increased power dissipation. There is always a trade off between the power dissipation and area of silicon used (and hence

costs) as use of parallel architectures can reduce the power consumption by reducing the rate of switching components and the supply voltage required. Reducing the supply voltage alone is not sufficient as this reduces the latency of the circuit and any security related hardware may not be able to meet timing constraints or performance constraints. Use of parallelism allows the trade off of silicon area for power. Design methodologies for reducing power consumption in CMOS logic are an active area of research and the reader is referred to [46] for more details.

$$P_i = P_{i\_static} + P_{i\_switching} \quad W \tag{1}$$

$$P_{i\_static} = I_{static} V_{dd} \quad W \tag{2}$$

$$P_{i\_switching} = C_L V_{dd}^2 f_{0\to1} \quad W \tag{3}$$

However, read range might not be a concern in certain applications and thus it is difficult to set a bound on the required power level, except to state that it should not exceed the power required by the tag during the writing of data to the memory as this is the most power consuming task a low cost tag is likely to perform. In addition, requiring more power would imply that a tag in its current position of being just able to operate (as it can be read by an interrogator) may not be able to complete a security related function, causing that operation to fail. This failure may expose or lead to vulnerabilities in the security mechanism. Hence power consumption is an issue that needs to be carefully considered.

Security mechanisms and communication protocols also need to be carefully designed to avoid leaving the label in a vulnerable state during sudden loss of power or interruptions to communications. It is also important for security mechanisms to take into account the more powerful signal strength of the forward channel (reader to label transmissions), which can be detected hundreds of meters away, compared to the tag to reader communication channel which can be received from no greater than 20 m using highly sensitive receivers.

The sections above have considered the nature of low cost RFID systems, its various vulnerabilities and the unique set of challenges to providing cryptographic solutions to alleviate these vulnerabilities. There are various solutions published in literature to address the weaknesses outlined in Section 3. The following section provides a survey of published solutions for addressing various privacy and security concerns related to low cost RFID systems.

9 A Survey of Solutions

Section 7 considered the subject of cryptography in a general perspective and introduced concepts that will be useful in the discussion of security mechanisms

for RFID. The following sections consider cryptographic primitives, protocols and security schemes proposed for low cost RFID systems.

It is important to note here that, in addition to the possible vulnerabilities discussed in Section 3, there will be specific attacks on any cryptosystem or protocol employed by a security mechanism used to provide security. These attacks are a result of certain weaknesses in the cryptographic scheme or are due to certain flaws that may have entered into the protocol employed in the security mechanism. A description and examples of such attacks can be found in [33, 41].
The following sections detail more recent developments addressing the issue of security and privacy for resource intensive environments.

9.1 Cryptographic Hash Functions

There have been a number of security schemes outlined in [44, 47]. A proposed scheme for controlling access to a label uses the difficulty of inverting a one-way hash function [44]. This mechanism, called the 'hash-lock scheme,' is based on a hash value generated from a random message sent to a tag for locking a tag. Until the tag is unlocked the tag only responds with the hash value stored on the tag called the MetaID. The tag can only be unlocked by an authorized reader by sending the original random message to the label, where it is hashed and compared with that stored on the label's memory.

The primary flaw in this approach lies in the fact that a successful discovery of a MetaID and a label ID pair will allow an adversary to engage in a cloning attack. The hash locking method requires the implementation of a suitable hash function and the appropriate logic to implement the details of a communication protocol. The greatest challenge lies in the successful implementation of a hardware efficient hash algorithm on the label IC. Since any reader can obtain the MetaIDs from labels, this scheme does not solve the problem of location privacy violations. The scheme is also susceptible to man-in-the-middle attacks since an adversary can query a label, obtain its MetaID, retransmit the value to a reader, and later unlock the label with the reader response.

However, the hash based access control can be extended to provide access control to multiple users, or control access to label functionalities such as write access. It is also possible to allow a third party to process labeled items using the MetaIDs and a database lookup scheme without having to unlock the labels. However, it should be noted here that any system that will function using the MetaIDs alone will suffer from the same security flaws as an unprotected label since the MetaID will act as a unique identifier (similar to an EPC on an unprotected label).

Randomised access control is another variation of the above scheme described in [44, 47]; with the difference to the previous scheme being that a tag always replies with a random MetaID. However, the randomized hash lock scheme has similar flaws and difficulties. The emphasis is placed on removing the predictable nature of the label responses to reader interrogations. A detailed description of this scheme can be found in [44].

Readers are still susceptible to replay attacks. An adversary only needs to obtain a label response and the corresponding reader response to create a fake label. An important consideration in this scheme is the number of labels that can be successfully supported, since a large number of labels will cause increasing processing delays at the back end database systems when performing brute force searches of databases to obtain the matching tag ID for a given MetaID. It is also not known whether keyed pseudorandom number functions required to implement the scheme are a more efficient hardware implementation than a symmetric key encryption such as a hash function. The hardware complexity of keyed pseudorandom number functions is still an active area of research.

Analysis of the random hash lock scheme in [44, 47] provided in [48] has also concluded that location privacy is only ensured in the scheme if an adversary cannot tamper with the tag to obtain the static tag identifier (such as the EPC).

The 'K-steps ID matching scheme' in [49] has outlined a method that utilizes representation of N tags on the N leaves of a tree of depth K where traversing the nodes of the tree down to the leaves yields the unique label identifier. This method allows to reduce the algorithmic complexity of searching a backend database from $O(n)$ to $O(\log n)$ where n is the number of tag identifiers stored in the database.
The hash locking method requires the implementation of a suitable hash function and the appropriate logic to implement the details of a communication protocol. The greatest challenge lies in the successful creation of a hardware efficient hash algorithm. However, it has been shown in [50] that implementations of hash lock schemes are generally not cheaper than symmetric key encryption schemes using an AES and a SHA-1 implementation as an example of a symmetric key encryption scheme and a hash scheme, respectively.

9.2 Cellular Automata

The theory of Cellular Automata (CA) [51] developed by Wolfram has been used to develop a number of different cryptographic systems. Cellular Hash (CH) [52] is one such outcome and there is a rich variety of inexpensive encryption mechanisms developed based on the chaotic nature of CA systems [53, 54]. CA may be built out of a feedback shift register and a single pair of gates providing a compact solution for low cost RFID. In addition, CA based hashes scale well as the size of the hash digest increases but CA hashes require many parallel calculations and thus they may impose considerable demands on a tag's available power. However, it is possible to perform CA operations in series but that will be at the expense of RFID system performance.

The CA cryptosystem encountered in CAC [54], while being promising, has been shown to be vulnerable to differential cryptanalysis or has been shown to form an affine group [55]. However, the estimated size of the un-optimized pre-layout area is about 4.25 sq. mm, which is far bigger than a typical RFID silicon design, which is about 0.25 sq. mm. Even if optimization halves the design, the silicon cost is too high for a low cost RFID chip. Nevertheless, improvements and a scaling down of the design may be possible since the analyzed design was for a 128 bit key.

The use of CA generators in the formulation of stream ciphers has also been proven to be insecure. It was shown in [56] that the output of a CA generator is identical to the output of a LFSR and hence CA systems are as insecure as LFSR based systems.

9.3 Linear and Non Linear Feedback Shift Registers

While LFSRs are capable of generating pseudorandom sequences they are insecure due to certain non random properties. Part of the problem is the linearity of the bit sequence which makes them of little use for encryption. Even in the event that the internal structure of the LSFR scheme is kept secret, an attacker only requires $2n$ output bits of the generator to determine the entire output sequence of a LSFR of length n [57, 58].

Use of non linear feed back shift (NLFSR) registers to design a hash by using a complicated feed back function is a possibility, since a shift register implementation does not require complex hardware. However an important consideration should be whether the additional cost of a NLFSR provides an adequate level of security, considering the vulnerabilities of various non linear feedback shift register based schemes in literature [33, 41, 42].

9.4 Message Authentication Codes

The use of Message Authentication Codes (MACs) has been discussed in previous literature. Takaragi [4] and his team of researchers have been the first to make an RFID chip (μ-chip) equipped with a MAC commercially available. The chip manufactured using a 0.18 micron CMOS technology occupies less than 0.25 mm^2 of silicon wafer, placing the IC in the low cost end of the RFID labels.

The MAC implementation adopts a very simple approach. The security of the μ-chip relies on a 128 bit ID stored permanently on the chip at manufacturing time. This ID is a concatenation of a previously encrypted MAC and chip data, the MAC being derived by encrypting a portion of the data using a hash function and a secret key, where the secret key is known to the manufacturer and the client. This mechanism does raise the difficulty level for forgers as the process of eavesdropping and creation of fake labels is made more complex. However it does not provide privacy as the ID code embedded in the chip will breach anonymity and location privacy. There is also the risk of the key, which is common to many labels, becoming known.

9.5 NTRU

The NTRU cipher appeared in 1995 [59]. NTRU is based on the Closest Vector Problem which involves finding the closest vector given a lattice L, and a target vector y [41]. It is similar to the knapsack problem. A lattice is defined as “the set

of intersection points of a regular (but not necessarily orthogonal) n-dimensional grid". This is an NP-Hard problem where there is no known algorithm for solving it in polynomial time [60].

There are several cryptosystems based on this problem [61, 62], but they have not gained in popularity due the excessive size of the keys needed to provide security comparable with other public key cryptosystems. The main advantages of NTRU are that it requires moderate resources and it is a faster operating algorithm. However, it is difficult to make accurate comparisons with other algorithms, as NTRU depends on many parameters that govern its behavior. Early research indicates that NTRU is generally faster, relatively easier to implement both in hardware and software than other public key cryptosystems such as RSA, and needs only a modest size memory [59]. Its simple implementation and limited demand on memory have already proven its relevance in RFID applications [63].

Nonetheless, NTRU is susceptible to brute force attacks and multiple message transmissions [59]. A more detailed treatment of attacks on NTRU can be found in [60]. In addition, NTRU has a relatively large message expansion. Encrypted messages are almost twice the length of the plain text messages. This may not be a pressing concern as RFID messages are not of very long length.

9.6 Tiny Encryption Algorithm

TEA is an encryption algorithm designed for simplicity and ease of implementation. The encryption algorithm is based on the Feistel cipher [33] and a large number of iterations to gain security without compromising simplicity. A description of the algorithm is provided in [64].

TEA can be effortlessly translated into any language as long as 'exclusive or' is an available operation. A hardware implementation of the algorithm is stated to have the same complexity as DES [64]. Despite its simplicity and the ease of implementation, TEA is a relatively recent invention and the level of security or its vulnerability to attacks is still not very clear.

9.7 Scalable Encryption Algorithm

The authors in [65] have noted that resource constrained encryption using symmetric cryptography does not have a long history. They cite TEA above and indicate the vulnerabilities of TEA to linear and differential cryptanalysis attacks.

SEA (Scalable Encryption Algorithm) is a scalable encryption algorithm for small embedded applications [65]. Typical performances of the SEA algorithm on encryption and decryption using a 128 bit key and 1 MHz 8-bit RISC processors can be undertaken in a few milliseconds, using a few hundred bytes of ROM [65].

9.8 Re-encryption

In [43] an unorthodox re-encryption mechanism is proposed for providing privacy and security protection to banknotes embedded with RFID labels. In a traditional setting, the entity conducting the re-encryption will not be aware of the plaintext. However, in the re-encryption scheme discussed in [43], the plaintext is known to the entity performing the re-encryption. The scheme is elaborate and the details are complicated, thus, only an overview of the scheme is given below.

The security of the mechanism is based on the ciphertext created by encrypting the digital signature stored on the RFID chip by a central bank authority, the serial number of the bank note and a random number. The authenticity of the banknote can be verified by comparing the ciphertext stored on the banknote to the ciphertext obtained by encrypting the digital signature, the serial number, and the random number using a public key stored on an RFID reader. A match indicates an authentic banknote. Figure 6 provides an overview of the data placed on each banknote. In addition, an access control mechanism prevents the data on an RFID label from being read without making optical contact first. This prevents remote alteration and interrogation of an RFID label's memory contents.

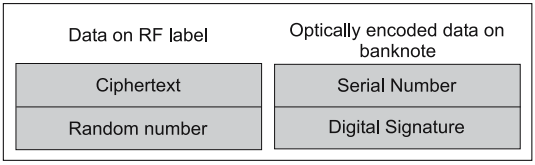


Fig. 6 Data on a banknote.

The significant privacy and security achievements outlined include consumer privacy (tracing of individuals or banknotes is only possible with the use of a key), forgery resistance, fraud detection and tamper resistance. In contrast with the aforementioned mechanisms outlined in Section 9.1, the primary significance of the re-encryption mechanism is that a banknote is not in possession of any secret keys and the RFID label is not required to perform any resource intensive operations. The encryption engines and secret keys have been shifted away from the RFID label to more secure locations, such as the readers and the central bank authority.

Despite the inventiveness of the re-encryption scheme, the significant drawback is the adequacy of the information obtainable from a banknote to create fraudulent banknotes. The digital signature is not verified during a transaction; hence, the fake banknotes can be created with ciphertext obtained from a collection of believeable serial numbers. Other shortcomings that might be exploited by a resourceful adversary are provided in [43].

9.9 Lightweight Cryptography

Lightweight cryptography is a branch of cryptography that aims to develop fast and efficient cryptographic mechanisms for resource constrained environments. Hence this branch of cryptography has been the most promising avenue to generate secure cryptographic solutions to low cost RFID systems. Several lightweight cryptographic models relevant to RFID are summarized below.

Building cost effective cryptographic hardware for RFID is still not a reality. Although certain advances have been made towards the development of hardware optimized encryption engines in [66, 67, 68], they still present a performance hindrance and an expensive solution to current RFID systems.

Lightweight Hardware

The process of developing or optimizing the existing hardware of security mechanisms has been an active area of research with respect to smart card processors. However, there has recently been some focus on more resource-constrained environments such as low cost RFID ICs. There have been a number of advances towards developing such low cost hardware [66, 67, 69, 70]. Elliptic Curve Cryptography (ECC) has presented itself as a public key cryptosystems for RFID [68] due to the smaller key sizes required to provide an adequate level of computational security. A relatively low cost implementation of an ECC processor suitable for RFID can be found in [71].

Lightweight Protocols

Building on prior work, Hopper and Blum have suggested two shared-key authentication protocols, HB and HB + protocols [72]. HB protocol is proven secure against a passive (eavesdropping) adversary. The HB + protocol is proven secure against active attacks. Security of these protocols are based on the conjectured hardness of the "learning parity with noise" (LPN) problem. Their extremely low computational overhead makes them very suitable for low power, bandwidth and low cost devices such as RFID. In [73] it has been proven that the security of these protocols only holds for sequential executions and the question of whether the security also holds in the case of parallel or concurrent executions is explicitly left open.

Katz and Shin [74] suggest that, in addition to guaranteeing security against a stronger class of adversaries, a confirmation of the security in parallel and concurrent operations would allow the HB + protocol to be parallelized. This would also reduce substantially its round complexity. Katz and Shin prove the security above. They also suggest simpler security proofs for these protocols which are more complete. In effect they also explicitly address the dependence of the soundness error and the number of iterations.

An improved version of the HB + developed by [75] is analyzed in [76] where a number of improvements have been made against various vulnerabilities outlined in [75]. HB + + [77] is a modified version of the HB + protocol designed to thwart a wider adversarial attack in [78].

9.10 Minimalist Cryptography

The formulation of mechanisms to achieve security and/or privacy objectives under the constraints presented by low cost RFID systems to real-world tags using a weak, but perhaps a realistic, security model form the basis for minimalist cryptography.

9.10.1 Pseudonyms

An early version of minimalist cryptography was proposed in [79] where a list of randomly generated tag identifiers was used on a tag. On querying a tag, a reader is able to hash the response and access tag related data on a secure hash table. The idea of using completely random EPCs and an outline of such a scheme was given in [79]. A similar version was also published in [80] with a minimalist security model and accompanying protocols for low-cost tags. The proposed method has every tag containing a collection of pseudonyms; it releases these pseudonyms on each interrogator query. Both schemes have left open the possibility for a valid reader to renew the list of pseudonyms on a tag.

The use of pseudonyms in [80] is based on the assumption that the intruder only comes into the scanning range of a tag on a periodic basis, as a complete analysis of the limited number of pseudonyms will allow the identification of the tag. The security model is also based on the underlying assumption that the tags release their data at a limited rate [80]. The minimalist model sets an upper limit on the number of times an intruder or an adversary can scan a given tag or try to spoof a valid reader.

9.10.2 One Time Pads and Random Numbers

In addition to the schemes presented above, there are many security schemes in the patent literature. Information security is a secretive realm, with many holding firmly onto their intellectual property with a whole array of patents. It is the nature of the beast.

Most methods outlined in patent literature are too complicated for low cost RFID. However, [81] demonstrates a very simplistic approach. The patented scheme relies on a simple one-time pad concept, where the intended application is that of bank notes. The scheme involves the recording of a random number, a time, and a date stamp on an RFID label of a bank note on release of the note for circulation. The bank note keeps a track of the number of times it has been scanned and this number is used as part of its authentication process. When a bank note is read by a bank teller, the random number, date, time stamp and the number of scans are sent to a central bank computer to verify the authenticity of the note based on comparing the same information securely stored on the computer.

This scheme is subject to imitation, simply because the label can not be trusted as a secure place of storage for valuable information because bank notes provide adequate incentives for a physical attack on the RFID label.

A different application of one-time pads can be found in another patent [82]. In the novel scheme, labels are equipped with a small rewritable memory. Prior to

the release of a label, a set of random numbers (authentication keys) generated by a completely random physical process is stored into the label along with a label ID. A back end database stores a copy of the random codes and the associated label IDs.

The label ID may be read from the label during an interrogation. The ID provides knowledge of the label being authenticated by the reader by consulting the relevant records in the secure back end database. In consultation with the database, the interrogator may transmit one or more of the random numbers stored in the database. One of the numbers should match a series of random numbers stored on the label. If a match occurs, the label responds with a return authentication code known exclusively to the database and increments a counter to select a set of new random numbers for the next authentication procedure. An identical counter that determines which of several authentication numbers is next in force is incremented at the database to synchronize the database entries with that of the label.

Hence, the above mechanism prevents an eavesdropper from obtaining any information regarding the next correct authentication key, or the next label authentication response. The only available information to an eavesdropper is an apparent burst of random numbers.

In the event of an unauthorized reader, the label will not respond unless the reader knows the next random number expected by the label. In case a counterfeit label is interrogated, the label may respond with a random number but the interrogator will fail to find a match, and thus detect the counterfeit.

Nevertheless, this scheme still leaves the possibility of a physical attack where the contents of the label may be discovered. However, in the worst case, this information cannot be used to counterfeit labels in massive quantities as the set of authentication keys and authentication responses are all different and completely random on each individual label.

9.11 Exploiting Noise

RFID systems are based on limited computing power and are not suitable for public key cryptography. Hence a protocol with low computation burden is outlined in [83].The protocol in [83] takes advantage of signal noise on a communication channel to secretly exchange a key in the presence of an eavesdropper.

An alternative proposal was presented in [84]. Similar to the blocker tags [85], the special tag in this proposal is named as the noisy tag. Noisy tags are owned by a reader's manager and set out within a reader's field. They are regular RFID tags that generate noise on the communication channel between the reader and the queried tag. This is done in such a manner that the intruder or eavesdropper cannot differentiate the messages sent by the queried tag and those sent by the noisy tags. Hence the intruder is unable to identify the secret bits that are sent to the reader. Afterwards, the secret shared by the reader and the tag can be used to launch a secure channel in order to protect communications against eavesdroppers. In addition, the tag's identifier can be refreshed by exclusively-or'ing the new identifier with the exchanged secret [84].

9.12 Radio Fingerprinting

If tags have distinct "radio fingerprints" that are difficult to reproduce, then these fingerprints, on their own, could help strengthen device authentication [149]. This technique, while sound in theory, is not a practicable avenue because obtaining such a radio fingerprint is expensive and difficult.

9.13 Distance Implied Distrust

This scheme is based on the assumption that an unauthorized reader attempting to read a tag will generally be more physically distant from the tags than a legitimate reader. The latter assumption is based on the realization that a closer and more visible reader will draw greater investigation by tag owners or tag bearers. Thus, the measurement of distance of a reader to a tag is proposed as a measure of trust [87].

9.14 Authentication Protocols

The YA-TRAP protocol proposed in [88], based on [89], provides location privacy and allows the authentication of the tag by using monotonically increasing timestamps stored on the tag which are in synchronicity with timestamps on a secure backend database. This protocol requires the implementation of an iterated keyed hash function on the tag. However, the proposal is vulnerable to a DoS attack initiated by desynchronizing the timestamp between the tag and the backend databases. Nevertheless by using hash tables that are pre-computed, the search time required for correlating a tag response is reduced to an $O(1)$ operation which requires less workload than the randomized hash lock scheme outlined in Section 9.1.

Another version of the YA-TRAP protocol in [90] also addresses some of the weaknesses in [88] albeit at the cost of increasing the workload of backend systems.

10 Conclusion

Despite the vast array of RFID systems, those that are within the low cost spectrum pose the greatest threat due to the possibility of wide scale deployment and inherent constraints that place limitations on the number of possible solutions. This paper has introduced in detail systems characterized as low cost RFID systems and identified numerous vulnerabilities of low cost RFID systems operating under the UHF EPCglobal air interface protocol. These vulnerabilities lead to both security and privacy related issues. Addressing these issues requires implementing various security and privacy objectives. Providing services to achieve those objectives traditionally requires the implementation of various cryptographic primitives.

However, low cost RFID, due to its resource intensive environment, presents a number of difficult challenges to more robust cryptographic mechanisms with a high level of security. Most of these robust cryptographic mechanisms are too area or power hungry to fit well within the limitations of RFID systems, and much of the encryption hardware available is for smart card technology. Even though the solutions can be applied directly to RFID, the main obstacle is that smart card processors are much more powerful than a typical RFID label consisting of only 200 – 4000 gates. Thus, the solutions are not portable to an RFID platform if we expect the cost of the secure labels to remain below the 5 cents mark.

It is also clear from the discussion on the level of security that all security system designers aim to develop a system where the only possible means of attack is by way of an exhaustive search. Such a system will then have a computational security determined by the size of its key space. Thus any security system considered must surely have a large enough key space to ensure security, especially considering the fact that CPUs used in personal computers in this century have benchmark performances rated in terms of teraflops (floating point operations per second).

Considering Moore's Law, cryptographic solutions that seem too expensive for low cost RFID may become the solutions in the future. However, it is not possible to wait for a future time frame while the deployment of RFID systems is taking place around the world. Since advances in cryptography are slow to arise, due to the time taken to scrutinize new mechanisms and find faults, the best option might be to fall back on simple and proven techniques, such as those presented in minimalist encryption and lightweight cryptography.

The survey of various research efforts to address the security and privacy issues provides an insight into current developments in the area of security for resource intensive, low computation capable devices.

It is important to recognize that the resource limitation of low cost labels suggests that the simplicity of small one-time pads, which involve one or more small shared secrets between a label and an interrogator, and relatively simple chip implementations, should also be considered and must not be discounted. Some of the concerns arising from privacy and security may also be removed by occasional use of shielded electromagnetic communications between the label and the reader system.

There are unique opportunities within the label class hierarchy to develop various schemes for meeting the security and privacy levels expected by labels belonging to their respective classes. This opens the gate to a vast number of research avenues that could be pursued with regard to providing both security and privacy to low cost RFID systems.

It must be realized that security will come in many flavors and strengths, but 'low cost' implies that we find mechanisms that are 'good enough' and are deterrents, rather than mechanisms that are impossible to crack.

Perfect secrecy is a fine mathematical concept; in reality, there will always be a human element that is difficult to quantify into any mathematical formulation. Thus, it is practically impossible to have a perfectly secure system. Once this is understood, it is possible to move onto addressing realistic security and privacy issues overshadowing RFID.

References

1 Verichip corporation home page. Available from:http://www.4verichip.com/ (06.2006)
2 Sarma, S.: Towards The 5c Tag. In: Technical Report MIT-AUTOID-WH-006 (2001). Available from: http://www.autoidcenter.org/research/ MIT-AUTOID-WH-006.pdf
3 EM Micro Readies New RFID Chip. In: RFID Journal news article, March (2003). Available from: http://www.rfidjournal.com/article/articleview/350/1/1 (06.2006)
4 Takaragi, T., Usami, M., Imura, R., Itsuki, R., Satoh, T.: An Ultra small individual recognition security chip. In: IEEE Micro, November–December (2001)
5 EPCglobal Inc.: Specification for RFID air interface (2007). Available from: http://www.epcglobalinc.org/standards_technology/EPCglobal2UHFRFIDProtocolV 109122005.pdf.
6 ITU, International Telecommunication Union. Available from: http://www.itu.int/ home/index.html (06/2006)
7 FCC Regulations, Title 47, Telecommunications, Paper 1, Part 15, Radio frequency devices, http://www.fcc.gov (2005)
8 ETSI, European Telecommunications Standards Institute, ETSI EN 302 208-1 V1.1.1 (2004–09), http://www.etsi.org/ (2006)
9 Cole, P.H., Ranasinghe, D. C., Jamali, B.: Coupling relations in RFID systems. In: Auto-ID Center white paper, June (2003)
10 Cole, P.H.: A study of factors affecting the design of EPC antennas and readers for supermarket shelves. In: Auto-ID Center workshop, October (2003)
11 Finkenzeller, K.: RFID Handbook: Radio Frequency Identification Fundamentals and Applications. John Wiley & Sons, New York (1999)
1. RFID Privacy and corporate data. In: RFID Journal, 2 June (2003). Available from: http://www.rfidjournal.com (08.2005)
13 Ranasinghe, D.C.: New directions in advanced RFID systems. In: PhD Thesis submitted to the University of Adelaide, School of Electrical and Electronic Engineering (2007)
14 A. Juels, “RFID Security and Privacy: A research Survey”, RSA Laboratories, September 2005
15 Bono, S., Green, M., Stubblefield, A., Juels, A., Rubin A., Szydlo, M.: Security analysis of a cryptographically-enabled RFID Device. In: Proceedings of 14th USENIX Security Symposium (2005) 1–16
16 Westhues, J.: Hacking the prox card. In: RFID: Applications, Security and Privacy, Addison-Wesley (2005) 291–300
17 Albrecht, K.: Chipping workers poses huge security risks, February (2006). Available from: http://www.freemarketnews.com/Analysis/139/3812/ 2006-02-15.asp?wid = 139&nid = 3812 (06.2006)
18 Ker, Z., Wool, A.: Picking virtual pockets using relay attacks on contactless smartcard systems. In: Proceedings IEEE/CreateNet SecureComm (2005) 47–58
19 Rieback, M. R., Crispo, R., Tanenbaum, A. S.: Is your cat infected with a computer virus? In: Fourth IEEE International Conference on Pervasive Computing and Communications (percom) (2006) 169–179
20 Oren, Z., Shamir, A.: Power analysis of RFID tags (2006). Available from: http://www.wisdom.weizmann.ac.il/~yossio/rfid/ (03.2006)
21 Avoine, G, Oeschlin, P.: RFID traceability: a multilayer problem. In: Financial Cryptography (2005)

22 Weigart, S.H.: Physical security devices for computer subsystems: a survey of attacks and defences. In: Workshop on Cryptographic Hardware and Embedded Systems, LNCS, Vol. 1965. Springer-Verlag, Berlin Heidelberg New York (2000) 302−317
23 Anderson, R., Kuhn, M.: Low cost attacks on tamper resistant devices. In: International Workshop on Security Protocols, LNCS. Springer-Verlag, Berlin Heidelberg New York (1997)
24 Bovenlander, E.: Invited talk on smartcard security. In: Eurocrypt 97 (1997)
25 Boycott Benetton web site. Available from: http:/www.boycottbenetton.com (12.2005)
26 Benetton, M.: Benetton explains RFID privacy flap. In: RFID Journal, 23 June (2004). Available from: http://www.rfidjournal.com/article/articleview/471/1/1/ (06.2006)
27 Roberti, M.: Analysis: RFID and Wal-Mart. News article, September (2003). Available from: http://www.cioinsight.com/article2/0,1540,1455103,00.asp(06.2005)
28 Jha, A.: Tesco tests spy chip technology. In: The guardian, 19 July (2003). Available from: http://www.guardian.co.uk/uk_news/story/0,3604,1001211,00.html (06.2005)
29 Spychips web site. Available from: http://www.spychips.com (12.2005)
30 J Collins. Marks & Spencer expands RFID retail trial. In: RFID Journal, 10 February (2004)
31 Molnar, D., Wagner, D.: Privacy and security in library RFID: Issues, practice, and architectures. In: Pfitzmann, B., McDaniel, P. (eds.): ACM Conference on Communications Security, ACM Press (2004) 210−219
32 RFID Upgrades Gets Goods to Iraq. In: RFID Journal, 23 July (2004)
33 Menezes, A., Van Oorchot, P., Vanstone, S.: Handbook of Applied Cryptography. CRC Press, New York (1996)
34 Stajano, F., Anderson, R.: The resurrecting duckling: security issues for ad-hoc wireless networks. In: International Workshop on Security Protocols, LNCS, Vol. 1796. Springer-Verlag, Berlin Heidelberg New York (1999) 172−194
35 Clarke, R.: Introduction to data surveillance and information privacy and definition of terms, August (1997). Available from: http://www.anu.edu.au/people/ Roger.Clarke/DV/ Intro.html#Id (01.2006)
36 Subirana, B., Bain, M.: Towards Legal Programming of Software Agents. Research Monograph, Kluwer (2004)
37 Commonwealth Freedom of information Act 1982, Australia (1982)
38 Commonwealth Privacy Act 1988, Australia (1998)
39 Electronic Privacy Information Centre, EPIC web site. Available from: http://www.epic.org (03.2004)
40 Beresford, A., Stajano, F.: Location privacy in pervasive computing. In: Pervasive computing, January−March (2003)
41 Schnier, B.: Applied Cryptography Protocols: Algorithms, and Source Code in C. John Wiley & Sons, Inc., New York (1994)
42 Stinson, D.R.: Cryptography Theory and Practice. CRC Press, New York (1995)
43 Juels, A., Pappu, R.: Squealing euros: privacy protection in RFID Enabled banknotes. Financial Cryptography (2002)
44 Weis, S.A., Sarma, S.E., Rivest, R.L., Engels, D.W.: Security and privacy aspects of low-cost radio frequency identification systems. In: Security in PervasiveComputing (2003)
45 Rabaey, J.M., Chandrakasan A., Nikolic, B.: Digital integrated circuits − A design perspective. 2nd edn., Prentice Hall, New Jersey (2003)
46 Rabaey, J., Pedram, M.: Low-Power Design Methodologies. Kulwer Academic Publishers, (1996)

47 Juels, A., Weis, S.A.: Defining strong privacy for RFID. In: RSA Laboratories (2006)
48 Avoine, G.: Adversary model for radio frequency identification. In: Technical Report, Security and Cryptography Laboratory, Swiss Federal Institute of Technology; Lausanne; (2005)
49 Nohara, Y., Inoue, S., Baba, K., Yasuura, H.: Quantitative evaluation of unlinkable ID matching schemes. In: Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society. ACM Press (2005) 55–60
50 Aigner, M.: Crypto implementations for RFID tags, presentation, Graz University of Technology (2006)
51 Wolfram, S.: A New Kind of Science. 2nd edn., Wolfram Media (2002)
52 Wolfram, S.: Cryptography with cellular automata. In: Advances in Cryptology: Crypto '85 Proceedings, LNCS, Vol. 218. Springer-Verlag, Berlin Heidelberg New York (1986) 429–432
53 Daemen, J., Govaerts, R., Vandewalle, J.: Hash functions based on block ciphers: a synthetic approach. In: Advances in Cryptology, LNCS. Springer-Verlag, Berlin Heidelberg New York (1991)
54 Sen, A., Shaw, C., Chowdhuri, D.R., Ganguly, N., Chaudhuri, P.P.: Cellular automata based cryptosystem (CAC). LNCS, Vol. 2513. Springer-Verlag, Berlin Heidelberg New York (2003) 303–314
55 Blackburn, S.R., Murphy, S., Paterson, K.G.: Comments on "theory and applications of cellular automata in cryptography". In: IEEE Transactions on Computers, Vol. 46 (5) (1997) 637–638
56 Bardell, P.H.: Analysis of cellular automata used as pseudorandom pattern generators. In: Proceedings of 1990 International Text Conference, (1990) 762−768
57 Meyer, C.H., Tuchman, W.L.: Pseudo-random codes can be cracked. In: Electronic Design, Vol. 23 (1972)
58 Meyer, C.H., Tuchman, W.L.: Design considerations of cryptography. In: Proceedings of the NCC, Vol. 42. Montvale, N.J. AFIPS Press (1972) 594–597
59 Hoffstein, J., Pipher, J., Silverman, J.H.: NTRU: A ring-based public key cryptosystem. In: Proceedings of ANTS III, Portland, June (1998)
60 Micciancio, D.: The hardness of the closest vector problem with pre-processing. In: IEEE Transactions on Information Theory, Vol. 47(3) March (2001) 1212−1215
61 Goldreich, O., Goldwasser, S., Halvei, S.: Public-key cryptosystems from lattice reductions problems. In: MIT LCS (1996)
62 McEliece, R.J.: A public key cryptosystem based on algebraic coding theory. In: JPL Pasadena, (1978)
63 NTRU web site. Available from: http:www.ntru.com/products/genuid.html (08.2003)
64 Wheeler, D., Needham, R.: TEA, a Tiny Encryption Algorithm. Computer Laboratory, Cambridge University, England (1994). Available from: http://www.ftp.cl.cam.uk/ftp/papers/djw-rmn/djw-rmn-tea.html (07.1995)
65 Stnadaert, F., Piret, G., Gershenfeld, N., Quisquater, J.: SEA: A scalable encryption algorithm for small embedded applications. In: CARDIS 2006. LNCS, Vol. 3928, Springer-Verlag, Berlin Heidelberg New York (2006) 222–236
66 Aigner, M., Feldhofer, M.: Secure symmetric authentication for RFID tags. In: Telecommunications and Mobile Computing TCMC2005, March 8th–9th (2005)
67 Feldhofer, M., Dominikus, S., Wolkerstorfer, J.: Strong authentication for RFID systems using the AES algorithm. In: Proceedings of the Workshop on Cryptographic Hardware and Embedded Systems. LNCS, Vol. 3156. Springer-Verlag, Berlin Heidelberg New York (2004) 357–370

68 Wolkerstorfer, J.: Is elliptic-curve cryptography suitable to secure RFID tags? In: Workshop on RFID and Light-Weight Cryptography, Graz, Austria (2005)
69 Martin, F., Manfred, A., Sandra, D.: An application of RFID tags using secure symmetric authentication. In: Proceedings of 1st International Workshop on Privacy and Trust in Pervasive and Ubiquitous Computing, Santorini Island, Greece, July 14 (2005) 43–49
70 Tillich, S., Großschädl, J.: Accelerating AES using instruction set extensions for elliptic curve cryptography. In: Proceedings of Computational Science and Its Applications. LNCS, Vol. 3481. Springer-Verlag, Berlin Heidelberg New York (2005) 665–675
71 Batina, L., Guajardo, J., Kerins, T., Mentens, N., Tuyls, P., Verbauwhede, I.: An Elliptic Curve Processor Suitable For RFID-Tags. In: Cryptology ePrint Archive, Report 2006/227, (2006). Available from: http://eprint.iacr.org/ (09.2006)
72 Hopper, N.J., Blum, M.: Secure human identification protocols. In: LNCS, Vol 2248. Springer-Verlag, Berlin Heidelberg New York (2001) 52
73 Juels, A., Weis., S.: Authenticating pervasive devices with human protocols. In: Advances in Cryptology, Crypto 2005. LNCS, Vol. 3621. Springer-Verlag, Berlin Heidelberg New York (2005) 293–308
74 Katz, J., Shin, J.S.: Parallel and Concurrent Security of the HB and HB + Protocols. In: Eurocrypt 2006 (2006)
75 Dimitriou, T.: A Lightweight RFID Protocol to Protect Against Traceability and Cloning Attacks. In: Proceedings of IEEE Conference on Security and Privacy for Emerging Areas in Communication Networks – SECURECOMM (2005)
76 Pramuthu, S.: HB and related lightweight authentication protocols for secure RFID tag/reader authentication. In: CollECTeR Europe Conference, Basel, Switzerland
77 Bringer, J., Chabanne, H., Dottax, E.: HB + + : A lightweight authentication protocol secure against some attacks. In: Security, Privacy and Trust in Pervasive and Ubiquitous Computing, June (2006)
78 Gilbert, H. Rodshaw, M., Sibert, H.: An active attack against HB + – a provabaly secure lightweight authentication protocol. In: IEE Electronic Letters, Vol 41 (21) (2005) 1169–1170
79 Ranasinghe, D.C., Engels, D.W., Cole, P.H.: Security and privacy solutions for low cost RFID systems. In: Proc. of the 2004 Intelligent Sensors, Sensor Networks & Information Processing Conference, Melbourne, Australia (2004) 337–342
80 Juels, A.: Minimalist cryptography for low cost RFID tags. LNCS, Vol. 3352. Springer-Verlag, Berlin Heidelberg New York (2001) 149–164
81 Szewczykowski: United States Patent, Patent number 5818021,Date of patent Oct. 6 (1998)
82 Cole, P. H.: Secure Data Tagging Systems. In: International Patent Application, Applicant TagSys Australia Pty. Ltd, Patent number PCT/AU02/01671, 10 Feb. (2003)
83 Chabanne, H., Avoine, G.: Noisy cryptographic protocols for low RFID tags. In: Workshop on RFID lightweight Crypto (2005)
84 Castelluccia, C., Avoine, G.: Noisy tags: a pretty good key exchange protocol for RFID tags. In: International Conference on Smart Card Research and Advanced Applications (CARDIS'06), Spain, April (2006)
85 Juels, A., Rivest, R.L., Mszydlo: The blocker tag: selective blocking of RFID tags for consumer privacy. In: Atluri, V. (ed.): 8th ACM conference on Computer and Communications Security. ACM Press (2003) 103–111

86 Jules, R., Daniels, T., Mina, M., Russell, S.: A small fingerprinting paradigm for physical layer security in conventional and sensor networks. In: IEEE/CreateNet Secure Comm, (2005)
87 Hancke, G.P., Kuhn, M.G.: An RFID distance bounding protocol. In: IEEE SecureComm, Athens (2005)
88 Tsudik, G.: YA-TRAP: Yet another trivial RFID authentication protocol. In: International Conference on Pervasive Computing and Communications – PerCom, Pisa, Italy, March (2006)
89 Herzberg, A., Krawczyk, H., Tsudik, G.: On traveling incognito. In: IEEE Workshop on Mobile Systems and Applications, December (1994)
90 Chatmon, C., Le, T.V., Burmester, M.: Secure anonymous RFID authentication protocols. Technical Report TR-060112, Florida State University, Department of Computer Science, Tallahassee, Florida, USA (2006)

Chapter 7
RFID Tag Vulnerabilities in RFID Systems

Behnam Jamali, Peter H. Cole, and Daniel Engels

bjamali@eleceng.adelaide.edu.au. cole@eleceng.adelaide.edu.au. dragon@csail.mit.edu

Abstract. More than half a century after its inception, radiofrequency identification (RFID) technologies are finally living up to their long promised capabilities. They are being rewarded with pervasive deployments in closed loop applications and the initial deployments in the even more pervasive open loop supply chain management applications. By providing accurate, real-time, human out-of-the loop asset and product monitoring throughout the world's supply chains, RFID technologies are beginning to improve the efficiency and security of these chains. The use of RFID technologies in these open loop supply chains is still in its infancy with all of the learning and growing pains that the introduction of a new technology entails. Security is of paramount importance in the deployment of RFID systems, particularly when they are being deployed, in part, to enhance the security of the supply chains. It is therefore appropriate that we examine now the potential security vulnerabilities inherent in the RFID systems currently being deployed in the supply chains of the world. Instead of covering the expansive RFID security landscape in this paper, we focus on the security vulnerabilities in the use of the data retrieved from an RFID tag. We conclude that the data stored on an RFID tag provides no more a security vulnerability to a system than any other manner of importing data into that system. Furthermore, the limited and highly structured nature of the data stored on the license plate RFID tags being used for supply chain management eliminates the potential for any security vulnerability due to the use of the tag data in a competent system.

1 Introduction

Radio frequency identification (RFID) technologies are rapidly becoming ubiquitous for product identification as businesses seek to improve supply chain operations and respond to mandates from key customers. As with the introduction of any new technology into an application, the use of RFID technologies to improve supply chain management, among the many applications currently adopting RFID technologies, creates new security and privacy issues that must be addressed to ensure the successful integration of the technologies.

RFID is a form of enabling technology that enables machines to collect information through sensors. Warehouses will sense whether they become low on stock

or, perhaps, overstocked. Luggage will be routed automatically from airport to airport. Healthcare, libraries and many other institutions will be all influenced if not changed by RFID. The benefits are potentially immense, but so are the potential security vulnerabilities, if the RFID systems are not designed properly.

One of the hotly debated issues currently is the ability of RFID tags to transmit computer viruses. No viruses targeting RFID technology has been released live as yet, but RFID systems have several characteristics that could be engineered to exploit vulnerabilities in the system and its back-end databases.

A hypothetical situation could be that a virus attack might start with a single RFID tag carrying evil data, read by a reader with a particular form of vulnerability. When that reader scans that tag, the readers bug would be triggered, causing the reader to execute a command specified by that tag. The command would reconfigure the reader to make it write copies of the evil data onto tags that it saw in the future. This would spread the evil data onto more tags. When any of those tags come in contact with the vulnerable reader, that reader would be infected, turning it into a factory for making more infected tags. The infection would spread from readers to new tags, and from tags to new readers. Before long many tags and readers would be infected.

The real potential vulnerability with the use of RFID tag data is how the data is used and what data "cleaning" is performed on the data before it is used. Poorly designed database systems are vulnerable to attack (both intentional and unintentional) regardless of where the data comes from. Data is used not just in the reader, which will have very limited use of the data collected, but in the information systems supporting the data. That information system is the target of any virus. The security vulnerabilities due to poor software architecture or implementation are distinctively different from inherent security vulnerabilities in the use of data. For example the use of a simple identifier as a pointer or a key into a database has no inherent security vulnerabilities in its use, but a poorly formed SQL substitution query is an implementation security vulnerability.

In this paper we focus on vulnerabilities due to the use of data retrieved from an RFID tag. An RFID tag is an unsecured and untrusted database. All data retrieved from it should be treated as such until sufficient confidence is gained that the data is accurate. In the software design, care must always be taken to ensure that data does not unintentionally cause problems. Failure to protect against unintentional malware creates security vulnerabilities that malware may take advantage of.

The paper is divided into four sections. The Code Injection scenario and approaches to manage its effects are discussed in the first section. In the second section we talk about Buffer Overflow in the context of RFID systems and reader design. Next section details the Denial of Service (DoS) attack on RFID systems. Finally the last section explains and reviews some final thoughts on how to mitigate the effect of these vulnerabilities.

2 Code Injection

2.1 Introduction

Databases form a critical part of RFID systems and are considered the backbone of RFID automation. An RFID reader interacts with the database through middleware software. Databases are susceptible to code injection attacks. To demonstrate the plausibility of this scenario, the authors of [1] adopted simple reader architecture and wrote their own middleware, giving it a common type of bug called an SQL Injection Vulnerability. They then constructed the precise diabolical data needed to exploit that vulnerability, and demonstrated that it would spread automatically as described. In light of this demonstration, it is clear that RFID viruses can exist, if RFID readers have certain types of bugs.

Structured Query Language (SQL) is a textual language used to interact with relational databases. SQL Injection occurs when an attacker is able to insert a series of SQL statements into a 'query' command by manipulating data input into an application. This type of attack can be carried out using any number of computer programming languages [2], [3].

2.2 Implementing an SQL attack

As an example one can corrupt a query operation by providing, while a query is being assembled, a response string containing special characters such as < > " ; & , etc, so that the effects of the query is modified. For example an original statement may be ignored and a new malicious statement introduced.

SQL injection occurs when an attacker is able to insert a series of SQL statements into a 'query' by manipulating data inputs into an application. Following is a typical SQL query statement.

select username, password from users where username = 'joe';

This statement will retrieve the 'username' and the 'password' columns from the 'users' table, displaying the rows in the table where username is equal to joe. An important point to note here is that the string literal 'joe' is delimited using single quotes. Assuming that the username field is being supplied by user input, an attacker might be able to inject some SQL statement into this query by inputting values into the application like:

username: 'joe';drop table—

The query string now becomes:

select username, password from users where username = 'joe';drop table--;

When database attempts to run this query, it will return an

error:

ERROR 1064: You have an error in your

SQL syntax near 'joe'

The reason for this is that the insertion of a single quote character breaks out the single quote delimited data. The database then tries to execute the next statement 'drop table', which results in the 'users' table being deleted. This makes the RFID systems OR database vulnerable to code injection viruses.

2.3 Summary

Using SQL code injection, tags could be employed to instigate a number of malicious attacks on the databases and middleware used in an RFID network, including buffer overflow and SQL injection, and even open a back door to the RFID application server. The authors in [1] wrote a poly-morpheus virus and used it to demonstrate the virus propagation within an RFID system. Their virus is based on SQL code injection. They used off-the-shelf commercially available equipment and software, but have developed their own middleware. The attacks can come in the form of an SQL injection, even though the tags themselves may only store a small bit of information, the paper said.

SQL injection attacks are possible because the SQL language contains a number of features that make it quite powerful and flexible, namely: the ability to embed comments in a SQL statement using a pair of hyphens; the ability to string multiple SQL statements together and to execute them in a batch, and the ability to use SQL to query metadata from a standard set of system tables.

SQL code injection attacks are well known and have been studied for many years [4], [5]. If one designs an RFID system that is vulnerable to SQL injection attack, the designer simply may have overlooked the fundamentals of RFID design features necessary in automatic data collection systems and good relational database design [4]. In other words, a poorly designed system has many security vulnerabilities when it comes to entering data, whether it is done capturing RFID tag information, bar code information or even using a keyboard.

Code injections that target database engines are easily prevented using the standard well-known protection techniques as described in this paper. Therefore the chance of this kind of virus spreading through the RFID network is very slim if the system is well designed.

3 Buffer Overflow

3.1 Introduction

Over the past few years buffer overflow attacks have become a major security threat to computers[6]. They have accounted for more than 50% of the advisories issued by CERT [7], demonstrating how serious the issue is. Most of the buffer overflows aim at forcing the execution of malicious code, mainly in order to provide a 'root shell'1 to the user.

The buffer overflow operating principle is quite simple: malicious instructions are stored in a buffer, which is overflowed to allow an unexpected use of the process, by altering various memory sections. In order to make use of these phenomena an attacker must have an intimate knowledge of the system that he/she is attacking [8].

Buffer overflow attacks aiming at the heap or data segment also pose a threat, though they are typically harder for an attacker to exploit. In this case an attacker must find and change some value in the memory that is security critical such a function pointer, so when the pointer is dereferenced the code of the attacker will be executed [9].

Technically, buffer overflow is a problem with a program's internal implementation [10]. A buffer overflow occurs when you write a set of variables, usually in form of a string of characters, into a fixed length buffer with a size smaller than the length of the string. It can occur when reading input from a user into a buffer.

Buffer overflow vulnerability comes in many flavours including stack smashing and heap smashing attacks. In this section we discuss how buffer overflow can be used to corrupt the stack. The stack is used to store information regarding function arguments, local variables, function return address and some information allowing the stack to be returned to the state before the function call. The stack is based on a LIFO (Last In, First Out) principle, and grows toward the low memory addresses.

A computer program is a collection of functions, when a function call is made the processor stores vital information on the stack and then it starts executing the code in the function by setting the Instruction Pointer register to the memory address of the function. After reaching the end of the function, the processor retrieves the return address from the stack, which is the address of the next instruction to be executed. Now if the stack has been corrupted during the function call (through buffer overflow), the return address (the address of the instruction to be executed next) could be any point in the memory. In case of an attack, we place our executable code, referred to as shell code, in a buffer and then we put the function return address to point to the buffer we have just modified with our own code. Therefore the processor will execute our code rather than what it was intended to do. This is not an easy task, as the function's return address on the stack is not known. An attacker has to find this address by trial and error. It is done by shifting the content of the buffer (a byte at a time) and padding it with nop instructions until eventually the return address is found. Substantial resources are available on how to mitigate the effect of buffer overflow [14]. Many contemporary compilers come with built in functionality to check for these types of flaw. This is an active research with many challenges [15], [16].

3.2 RFID reader architecture

A contemporary RFID reader design contains two main processors [11], [12], [13], a Digital Signal Processor (DSP) and a General Processor (GP). The DSP chip is responsible for modulation, demodulation, waveform shaping, coding and decoding of RF signal, while the GP is used for general data manipulation, storage, query

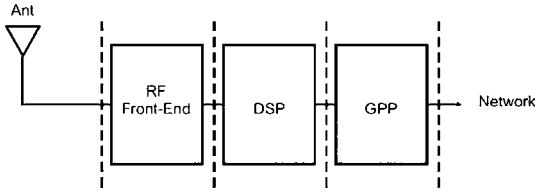


Fig. 1 Architecture of a typical RFID reader.

and networking. The development of this mechanism is one of the right moves made in the direction of security.

The DSP chip is the main interface between an RFID tag and the rest of the RFID system. As such any buffer overflow attempts would affect the DSP chip first. Most of the workstations and home PC run standard pieces of software, such as MS Windows or UNIX. An attacker can obtain intimate knowledge about the software run on those platforms because of their popularity and standardization. Contrary to the PC world, the software run on DSPs (referred to as firmware), are chip dependent and are different from one chip to another and even from one revision to the next.

A GP receives its input from the DSP chip. It does that usually by reading data (a fixed length) from the DSP memory. In case of a TI DSP this can be done through the HPI (Host Port Interface) interface [17]. If the DSP security has been compromised by a rogue tag and its software has crashed, this still would not cause buffer overflow in the GP processor, because the GP only reads a fixed amount of data from DSP's memory. In the worst case scenario all it gets is garbage. The obtained data might not be valid but it cannot cause buffer overflow in GP either. Therefore buffer overflow attack is only limited to the DSP chip. Considering that a DSP chip has no control over the main processor and has no access to the network, it is unlikely to gain control of an RFID system by means of buffer overflow.

Another difficulty in applying a buffer overflow exploit to an RFID system is that an attacker needs to have an interactive connection to an RFID reader, i.e., receive feedback from the system, while an RFID reader does not except in exceptional case provide feedback to a tag.

A typical EPC tag has 96 bits of memory with an ID number. For such a threat to be credible, there would have to be more memory, a read-writable tag and an RFID reader capable of reading and writing to such tags.

3.3 Summary

A buffer overflow occurs when a program or process tries to store more data in a buffer than it was intended to hold. Buffer overflow can be easily prevented if the application that stores the data in a buffer does a proper boundary checking to make sure that the data offered is not larger than the buffer allocated to it.

In case of an RFID reader, the reader knows how many bits of information it is pulling out of the noise so it always returns data of proper length.

Also RFID interrogators to comply with the EPC global standards need to be able to detect rogue tags or rogue software on tags as part of the verification process of reading them.

In an RFID system, most applications, including the EPC Gen 2, look for specific kinds of data. Poor reader design might allow the reading of an unrecognized tag, but a good system will verify that data against pre-defined parameters. If the incoming information does not satisfy the conditions then it is typically ignored. This is a common practice among all the applications developed today.

Given the constrains mentioned in this section, and, based on multiple processor architecture of RFID readers as discussed above, code injection through buffer overflow using a rogue or customized tag seems very far fetched in an RFID context. Therefore even if a buffer overflow takes control of an RFID reader, the damage would be very limited. The dual DSP/CPU configuration of RFID readers provides a buffer zone (or a firewall) where buffer overflow on DSP cannot propagate to the GP, as such it cannot propagate through to RFID network.

4 Denial of Service

A Denial of Service (DoS) attack is designed to causes a loss of service to users, typically the loss of network connectivity and services by consuming the bandwidth of the victim network or overloading the computational resources of the victim system.

In case of an RFID system, it is possible to perform DoS attack on tags using cheap store-bought radio transmitters that transmits signal at the operating frequency of the tags. It also is easily possible to cause Tag Confusion. When a tag receives what it considers an intelligent signal from a reader, it attempts to decode but then it may find the signal to be an un-correctable error. The tag then generally resets itself to an error state or reset its circuitry to the initial power-up state.

In order to strengthen security of the EPC Gen 2 protocol[18], the designers have implemented two fundamental key features. First, the ability to lock a tag so that only an authorized interrogator can write any data to it, and secondly the use of encryption, which adds a random number to a tag's ID and requires the tag and reader to exchange what it likens to a handshake before they can exchange any data.

These features make it much harder to introduce a virus into the system using the methods shown in the aforementioned paper, but they still are vulnerable to the air interface from an RF jammer. From all the different attacks that can be made against an RFID system, Denial of Services (DoS) is the easiest one to achieve. As such tag confusion and reader confusion are two possible methods of attacking RFID systems. Researchers at university of Adelaide are working actively to address these issues.

A faulty/noisy tag or any device such as a jammer can cause DoS in a reader. In such situation the RFID reader creates an immediate error condition that should be flagged by the middleware to create a high priority ticket for maintainers to raise an alarm and human intervention can solve the problem.

5 Conclusion

Protection against viruses is a common process today in computer world. Global standards and commercial products are created to ensure that there is adequate data security, much the same way that running current virus software prevents virus attacks to computers.

The methods showed in this paper are the most common attackers on computers and databases. They are implementation specific as opposed to fundamental to RFID system. RFID middle-ware designers must be and normally are aware that they must install, and have already installed, methods to authenticate any data into their software.

Critical advancements in information processing and logistics made possible by RFID technology cannot be underestimated. Corporations, government agencies and consumers will enjoy greater confidence when they select standardized RFID technology because the associated security issues have been addressed and resolved by the world's leading experts.

RFID systems are a collection of tags, middleware, software and database systems, as well as custom software to bind these elements together. In contrast to PC desktop every RFID system is unique. Therefore an attacker would need intimate knowledge of the system's vulnerabilities (if there exist any) to be able to make use of them. And even then it can only affect one particular brand of RFID reader.

At the end of the day if some one wants to manipulate the pricing of an item it can easily be detected by the cashier. The same as when one can peel off a printed bar code from one item and stick on another item, RFID tags can be misplaced as well. In case of an RFID tag, there is a possibility that one can change the content of a tag to represent a different identification code rather than physically switching the tag. All this happens only in a place where someone tries to cheat the system and in such situation there always exist a cashier who can check the validity and correlate the tag id number with the item it is attached on.

So far the only virus propagation/infection method demonstrated relies on the Buffer Overflows and SQL Code Insertion techniques. As shown in this paper even the most primitive RFID readers are immune to these kinds of attacks. Therefore in conclusion we must say that the idea of virus being spread through RFID tags is very slim, but the denial of service attacks are easily achievable.

Acknowledgement

The authors would like to thank Alfio Grasso, Raja Ghosal and Damith Ranasinghe for their valuable inputs.

References

1 Melanie R. Rieback Bruno Crispo Andrew S. Tanenbaum, “Is Your Cat Infected with a Computer Virus?”, Vrije Universiteit Amsterdam, Computer Systems Group, IEEE PerCom 2006

2 Chris Anley, “Advanced SQL Injection In SQL Server Applications”, 2002 Next Generation Security Software Ltd, http://www.nextgenss.com/papers/advanced sql injection.pdf

3 Chris Anley, “(more) Advanced SQL Injection”, 2002 Next Generation Security Software Ltd, http://www.nextgenss.com/papers/more advanced sql injection.pdf

4 SQL Server Security Checklist, http://www.sqlsecurity.com/checklist.asp

5 “Stop SQL Injection Attacks Before They Stop You”, http://msdn.microsoft.com/msdnmag/issues/04/09/SQLInjection/default.aspx

6 C. Cowan, P. Wagle, C. Pu, S. Beattie, and J. Walpole “Buffer overflows: Attacks and defenses for the vulnerability of the decade.” In Proceedings of the DARPA Information Survivability Conference and Expo, 1999

7 Centre of Internet security expertise, http://www.cert.org

8 Pierre-Alain FAYOLLE, Vincent GLAUME, “A Buffer Overflow Study, Attacks & Defenses”, ENSEIRB Networks and Distributed Systems 2002

9 Cert coordination center, vulnerability note vu#363715. http://www.kb.cert.org/vuls/id/363715.

10 David A. Wheeler, “Secure Programming for Linux and Unix HOWTO”, http://www.dwheeler.com/secure-programs/Secure-Programs-HOWTO.pdf

11 “ALR-9800 ENTERPRISE RFID READER” http://alientechnology.com/docs/AT DS 9800 v3 WEB.pdf

12 “Mercury 4 EPCglobal Gen2 certified in all modes, including Dense Reader Mode Intelligent, Network Ready, Reads Any Tag” http://www.thingmagic.com/html/pdf/m4brochure.pdf

13 “XR400 RFID Reader” http://www.thingmagic.com/html/pdf/m4brochure.pdf

14 Eric Haugh, Matt Bishop “Testing C Programs for Buffer Overflow Vulnerabilities”, University of California at Davis http://www.isoc.org/isoc/conferences/ndss/03/proceedings/papers/8.pdf

15 “Windows Server 2003 in a Managed Environment”,http://www.microsoft.com/technet/prodtechnol/windowsserver2003/technologies/security/ws03mngd/13 s3iis.mspx

16 David Litchfield, “Defeating the Stack Based Buffer Overflow Prevention Mechanism of Microsoft Windows 2003 Server.”http://www.ngssoftware.com/papers/defeating-w2k3-stack-protection.pdf

17 “Host Port Interface Reference Guide”, http://focus.ti.com/lit/ug/spru588b/spru588b.pdf

18 “EPCglobal Class 1 Gen 2 RFID Specification”,http://www.alientechnology.com/docs/AT wp EPCGlobal WEB.pdf

Chapter 8
An Evaluation Framework

Damith C. Ranasinghe, and Peter H. Cole

Auto-ID Lab, The School of Electrical and Electronic Engineering,
The University of Adelaide, Adelaide SA 5005, Australia.
{damith,cole}@eleceng.adelaide.edu.au

Abstract. There are various solutions expounded upon to address security vulnerabilities and privacy violations of low cost RFID systems. This paper will formulate a framework for defining the problem space around low cost RFID systems to enable the engineering of solutions and for evaluating those solutions for their effectiveness in the contest of networked low cost RFID systems.

1 Introduction

Achieving security and privacy objectives using cryptographic solutions in low cost RFID is a challenging but a necessary proposition. Design, analysis and evaluation of security mechanisms are other challenging aspects due to the lack of an established evaluation criteria or a clear security model. The following sections develop simple evaluation criteria for security mechanisms and a simple, yet sufficient model of a low cost RFID system for analysing security mechanisms.

It is important to define certain boundaries and assumptions taking into account the challenging aspects of implementing cryptographic primitives on low cost RFID. Implementing mechanisms to address security or privacy otherwise is inconceivable. Defining and modelling the problem space will also aid future research in low cost RFID security. Table 1 summarises a set of important characteristics that needs to be understood and reasonable assumptions that need to be made prior to implementing any cryptosystems to address the vulnerabilities of low cost RFID systems outlined in [1, 2].

Table 1 An outline of low cost RFID system characteristics (Modified from Damith C. Ranasinghe, Daniel W. Engels, Peter H. Cole "Low cost RFID systems: confronting security and privacy", 2005 Auto-ID Labs White Paper Journal Volume 1, © 2005 Auto-ID Labs).

Class of labels	Class I and Class II type of labels, (as they are low cost RFID labels)
Unique identifier	EPC of length 96 – 256 bits. As defined in by EPCglobal in their tag data specification standard.
Read range	3 m – 5 m for UHF and 200 – 500 mm for HF operation under FCC regulations.
Label reads per second	200 – 1500 as demanded by end user performance requirements.
Hardware cost	Current fabrications of Class I labels consist of around 7,000 to 10,000 logic gates while Class II labels may have several thousand more gates.
Power consumption	10s of microwatts, and should not exceed that required for E^2PROM operation, so the tag read range requirements can be maintained.

2 Evaluating Security Measures

While designing cryptographic solutions is challenging it is important to be able to evaluate security measures devised to ensure that various goals outlined in Table 2, suggested in [1] for providing security and privacy, are satisfied while meeting expected performance levels and costs. Table 2 outlines a security evaluation matrix to appraise the suitability of various mechanisms for providing security and privacy to low cost RFID and various applications constructed around low cost RFID technology.

Table 2 Criteria for evaluating security mechanisms.

Achieved Security Objectives	Confidentiality
	Message content security
	Tag Authentication
	Reader Authentication
	Product Authentication
	Access control
	Availability
	Integrity
Achieved Privacy Objectives	Anonymity
	Untraceability
Cost and Performance Estimates	Tag implementation cost estimate (gate count estimation)
	Back end resource requirements (online or offline)
	Overhead costs (initialization costs, requirements or time)
	Time estimate (time to complete a process or hardware throughput or clock cycles)
	Estimation of power consumption (maximum bound)

2.1 Achieved Security and Privacy Objectives

The definition and description of the security and privacy objectives in Table 2 have been expounded upon in [1]. Therefore, given a solution, it is then possible to evaluate that solution to analyse the set of security and privacy goals accomplished using the matrix in Table 2.

2.2 Evaluating Cost and Performance Objectives

Estimating the cost of security mechanisms, their power consumption and their performance is vital to assess the suitability of the security mechanism to a networked low cost RFID application. The following sections consider the cost and performance objectives in Table 2.

2.2.1 Tag Implementation Cost

Evaluating the tag cost of a security measure generally refers to the cost implication for its implementation on a tag IC measured in terms of the silicon area required for fabrication. It is generally both expensive and complex to carry out such a physical implementation, and an estimate may be found by implementing the hardware required on a FPGA (Field Programmable Gate Array). Alternatively, it may be possible to evaluate the cost of an IC in terms of the number of gates required for its implementation.

One NAND gate is considered to have a unit area in CMOS standard cell based hardware and it is common to express the area evaluation in terms of the number of gates (NAND) required. Implementing a NAND gate in hardware requires at least four FETs. Typical cost estimations in terms of the gate count are given in Table 3.

When referring to the cost of a tag, for security purposes the associated cost implied is the cost of the digital components required to implement the security measure

Table 3 Cost estimation guide for cryptographic hardware based on static CMOS designs.

Functional Block	Cost (gate count equivalent)
2 input NAND gate	1
2 input XOR gate	2.5
INV (Inverter)	0.5
2 input AND gate	2.5
FF (Flip Flop)	12
D latch	2
2-1 MUX	5
n-bit LFSR	$n \times 12$
n-byte RAM	$n \times 12$
HA (Half Adder)	1 XOR + 1 AND
FA (Full Adder)	3 AND + 2 OR + 2 XOR

on chip. In respect to the cost constraints placed on these labels and taking the current cost of fabricating a transistor to be 1/1000th of a US cent, the low cost labels can be expected to have 2000 – 5000 gates available for security purposes, although the number of gates available is expected to increase over the years as manufacturing techniques and processes improve or as RFID IC manufacturing begins to fill the excess capacity of obsolete processes that are still in operation. Overhead costs will be reflected in the performance of the system.

2.2.2 Backend Resources and Overhead Costs

It is generally difficult to implement a security mechanism without the aid of proxy systems or secure backend system for storing secret information such as keys. Security mechanisms of this kind require online and real time access to secure resources. The monetary and time cost of implementing such a mechanism must be accounted for in the evaluation process. Timing constraints placed on RFID security mechanisms may require expensive database system implementations and expensive networking infrastructure. However, it may be possible to design a security mechanism that performs its operations off-line (that without requiring online access to secure resources). Hence backend resource costs can be generally expressed as those requiring online access or those that can be performed off-line.

Overhead costs may result from the need for initializing tags with secure information, or the need for performing some operations prior to their use or periodically during their use. For instance a security mechanism may require the replenishment of secret keys on a tag.

2.2.3 Power Consumption

Any security mechanism design will eventually involve an IC implementation. Currently, static CMOS is the choice of most digital circuit designs built for low power consumption and robustness. Hence, it is appropriate to consider power consumption analysis based on an implementation using static CMOS technology. However a drawback of this technology is the extra silicon area required for implementing logic compared to dynamic CMOS.

An important aspect of the design process and the establishment of its suitability is to ensure that the power dissipation of the integrated circuits do not exceed that outlined in Table 1. There are several techniques for measuring power consumption [3, 4] and most methods rely on simulation techniques based on various models using simulation tools such as HSPICE [5].

While the use of direct methods to measure power dissipation may be possible a simple method for estimating the dynamic power dissipation is based on formulating the power loss during the charging and discharging of capacitances. Equation (1) models the power dissipation of a node that may consist of a number of logic gates and it is probably most practicable when the logic circuit complexity is a minimum.

$$P = p_{0\to 1} C_L V_{dd}^{\ 2} f_{clk} \text{ W} \tag{1}$$

In (1), C_L is the output load capacitance along the critical path and $p_{0\rightarrow1}$ is the fraction of time the node makes a power consumption transition (that is a logic 0→1 and 1→0) in a single clock cycle. The combination of $p_{0\rightarrow1}$ C_L can also be stated as the average capacitance switched during each clock cycle. In (1), f represents the clock frequency and V_{dd} is the maximum signal swing, more generally taken as the operating supply voltage. It is difficult to apply this formula to ICs of large sizes. However, it is adequate for estimating the power consumption in small hardware, especially if circuit design tools can be used to evaluate the capacitance estimates.

2.2.4 Performance

Furthermore, it can be assumed that the time available for a label operation (and thus, any implementation of a security mechanism that needs to be performed in real-time) is in the range of 5 – 10 milliseconds considering the performance criteria of an RFID system that demands a minimum label reading speed of at least 200 labels per second. In accordance with C1G2 protocol, a maximum tag to reader data transmission rate bound of 640 kbps and a reader to tag data transmission rate bound of 126 kbps, based on equi-probable binary ones and zeros in the transmission, can be calculated. These data transmission rates can be used to estimate the execution time of a protocol and hence system performance.

3 Security Model

Given the list of vulnerabilities posed by low cost RFID systems and the unique nature of the technology, it is natural to consider the nature of adversaries in an RFID systems context with its various limitations, and to provide a clear security model. It is important to develop the nature of various adversaries in an RFID systems context in order to be able to analyse the level of protection provided by various security mechanisms and the various circumstances under which the security mechanism may be deployed to achieve the security and privacy objectives outlined in Table 2. A number of ideas modelling possible adversaries and system models have been published in [6], [7], [8] and [9]. However, analysing the security of an RFID security mechanism is still challenging, partly because of a lack of a universal model or the narrow scope of the models developed.

The following sections develop a simple model of a low cost RFID system and a number of adversaries with various capabilities and goals.

3.1 Authorised and Legitimate

A discussion regarding security and privacy requires a clear understanding of the trusted parties. The term “authorised” will be used in relation to readers or interrogators who are registered in a given RFID system’s database and are equipped with the necessary security mechanisms to access secure resources of that system.

The term "legitimate" will be used in relation to tags that are registered in a given RFID system's database as verified by an authorised reader. Given the above concepts of authority and legitimacy, a reader that is not authorised will be referred to as an "unauthorised" reader while a tag that is not legitimate will be referred to as a "fraudulent" tag. In the event a legitimate tag is copied to produce a copy that may for all purposes has the ability to be verified as a legitimate tag, it will be called a "cloned" tag.

3.2 Tamper Proofing

The long-term security of label contents cannot be guaranteed since these contents are vulnerable to physical attacks. Tamper proofing to prevent physical attacks is an expensive option. Hence, it is assumed that labels cannot be trusted to store long-term secrets such as secret keys that apply to a range of RFID labels, but secrets pertinent to an individual label that is unrelated to another label can be considered to be acceptable, as long as the information obtained is unhelpful in defeating the security mechanism of other tags.

3.3 System Model

While RFID systems consists of various components a system model need only be concerned with interrogators, tags, and the communication channels between them. The problems involved with securing reader communications with a secure backend database will not be considered; instead, it is assumed that legitimate readers have secure connections to backend systems. The general notions of eavesdropping, as described in [1], can be assumed. In such a context it is clear that that the forward channel (interrogator to reader) is exposed to undetectable eavesdropping from several hundred meters away, while the backward channel (label to interrogator) can only be monitored from close range (several meters away, for UHF systems and a few centimetres, in HF systems).

Some implementations of readers may store security related information onboard (providing an offline security mechanism), while other readers may simply obtain security related information from a backend system (as required in an online security mechanism). In either scenario a reader has adequate resources to secure the stored information or use a secure communication channel for obtaining the information.

Since other components of the system are not confronted with any constraints with regards to implementing necessary cryptosystems to secure information it is possible to consider an interrogator and the rest of the back end systems as a single entity, which can be referred to as an authorised interrogator. Now it is possible to outline the communication participants; authorised interrogators, legitimate tags, fraudulent tags, clone tags and unauthorised interrogators (refer Section 3.1). Figure 2 shows the possible ways in which the various communication participants can interact.

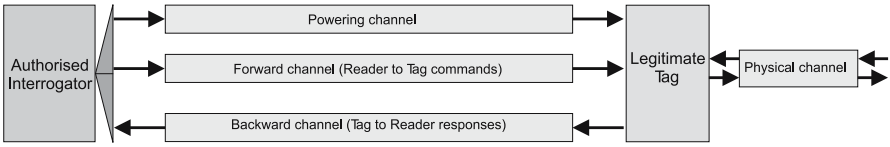


Fig. 1 System model describing the information channels of an RFID system (Damith C. Ranasinghe and Peter H. Cole, "Confronting Security and Privacy Threats in Modern RFID Systems", 40th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, California, 29 Oct. -1 Nov. 2006. © 2006 IEEE).

A communication model describing the various information channels is shown in Figure 1.The sources of information available to an adversary in a low cost RFID system are restricted to those which can be obtained over the insecure communication channels, the contents of the tag memory by way of the memory channel from tags which are constantly in an untrusting environment and from the interrogation of legitimate tags as shown in Figure 2. A discussion on the eavesdropping distance of the forward channel, and the backward channel can be found in [1]. While the forward channel is generally longer than the backward channel, in a UHF RFID system, the backward channel is in the range of several metres and can be easily observed by a hand held reader, or an RF receiver with a higher sensitivity from a greater distance. Hence in a UHF RFID context it is not appropriate to consider the two channels separately while it is worthwhile considering the two channels separately in the context of an HF RFID system, where a tag antenna must be inductively coupled at very close range for a reader to be able to decode the backward channel. Hence protocols based on considering the forward and backward channels separately in their security model cannot be considered to be secure for systems operating in the UHF region while the model may hold true for HF systems.

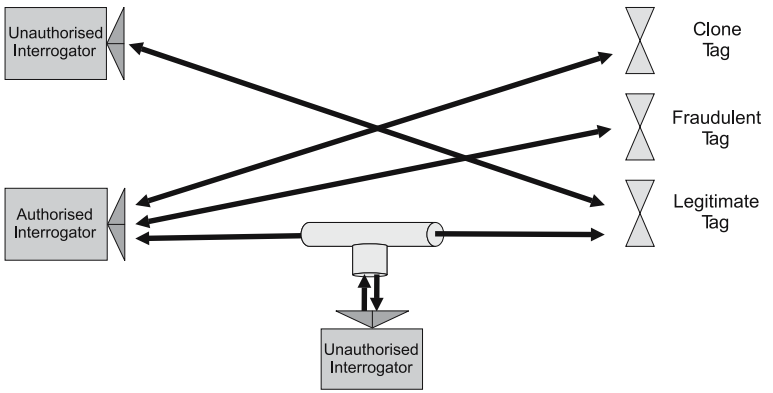


Fig. 2 Possible interactions between communication participants (Damith C. Ranasinghe and Peter H. Cole, "Confronting Security and Privacy Threats in Modern RFID Systems", 40th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, California, 29 Oct. – 1 Nov. 2006. © 2006 IEEE).

Although it is theoretically possible for an adversary to read or write to the forward channel, backward channel or the memory channel, in practice such duality is not possible. It is possible for an adversary to provide a powering channel and observe the powering channel as in the case of a power analysis attack. The adversary may read or write to the forward and or the backward channel, depending on the frequency of operation and the nature of the adversary being modelled, as in the case of a man-in-the-middle attack.

The complete set of data stored on a legitimate tag will be unique, and not dependent on other tags. This assumes that a tag's complete set of memory contents is unique (such as secret keys, unique tag identifiers and passwords) and it is not identical to any other tag legitimate tag.

The physical channel is considered to be read only once, as it may be used to read the memory contents of the tag, and such an operation is destructive to the legitimate tag. However, it may then be possible for the adversary to create a clone of the tag using the information obtained, but creating many clones of the same tag is considered to be not possible in supply chain applications due to the availability of a track and trace facility and the fact that a given tag has a complete set of data that is unique. However, for a general application such as an access control system, such an assumption is not valid, hence the validity of the read only once assumption must be utilised with care.

3.4 Adversary Model

Threats exist to RFID systems because of the value of the object to which an RFID tag is attached or the asset that the RFID tag is attempting to protect or the value derived from correlating object identification information with human identification information. Who are the adversaries and what goals do these adversaries have? An RFID system deployed in a consumer goods supply chain application will surely have different adversaries from those in a supply chain application for the Department of Defence. An adversary looking at a system will only see the weaknesses of the system, outlined in [1], and will aim to exploit them to achieve various objectives: theft, defeating an access control system or mass production of counterfeit goods to illustrate a few. Such objectives are only limited by imagination and they depend on the application. However it is possible to have a simple classification of adversaries according to their objectives, level of interference, presence, and available resources.

3.4.1 Objectives of an Adversary

It is clear that an adversary model should be determined with respect to applications. However, achieving an adversary's objective will ultimately involve using various methods, such as cloning, or eavesdropping to defeat one of the security or privacy objectives outlined in Table 2.

For instance an adversary may violate confidentiality to obtain a tag's access password by eavesdropping on a conversation between a tag and a reader, or an adversary may collect tag data to associate object data with human data or use

a physical attack to obtain secure information stored on a tag to create counterfeit products. If there is a gain to be made by disabling a system, the adversary may initiate a DoS attack (defeating the security objective of availability).

3.4.2 Level of Interference

It is possible to simply take the more traditional notion of a passive and an active adversary in an RFID context. A passive adversary is capable of eavesdropping on both the forward and the backward channel or just the forward channel (in case of HF systems) without being detected, then record and analyse the data in real time or at a later time. An active adversary may play a number of different roles discussed below.

A malicious adversary has no respect for electromagnetic compatibility regulations and has the ability to communicate with tags without being discovered by avoiding raising any suspicions but with a bound on the number of possible tag queries. The limitation on the number of tag queries comes from the assumption that tags will implement a throttling mechanism to prevent repeated tag queries from being executed in a brute force style attack.

A disruptive adversary can attempt to hinder the process of an interrogation by an authorised reader or the responses from a legitimate tag. Such an attempt will amount to a DoS attack. The adversary may also interrupt protocols, alter authorised reader commands or legitimate tag responses.

3.4.3 Presence

The presence of an adversary is an important and a defining aspect of an adversary's character. A stationary adversary may eavesdrop on RFID communications in a confined space such as a store, a warehouse or a room.

A mobile adversary seeks to clandestinely track the items. The adversary can maintain logs of all observed tag identifiers and locations while being able to interact with the tag using its standard air interface protocol. Such an attacker has the ability to use the recorded information to interact with legitimate readers to obtain tag related information. However a mobile or a local attacker may be a proximity adversary or a distant adversary.

A proximity adversary is able to record both the forward channel, the backward channel and utilize the powering channel while a distant adversary is only able to record the forward channel and utilize the powering channel.

3.4.4 Available Resources

It is necessary to assume various levels of capability for an adversary. Generally, these capabilities can be expressed as the level of resources available in the following categories.

- Financial
- Equipment
- Knowledge
- Time

The aim of solutions to secure and protect privacy concerns should be to achieve a practical and adequate level of security and protection by designing security measures to address various security and privacy goals to defend against realistic attackers. Thus attackers with unlimited resources such as highly funded government or non-governmental organisations need not be considered for most supply chain applications. Equipment availability is another important consideration, especially when dealing with physical security of devices. Generally computing power is considered in the same category. Deciding on adequate key lengths based on current computing power is not difficult but it is difficult to estimate future computing power to forecast the length of time a key size or a cryptographic algorithm may be used. However, the general rule of thumb is that the efficiency of computing equipment divided by price, increases by a factor of ten every five years [10]. This is only true of PCs (Personal Computers) [11].

A difference may be made withrespect to the type of knowledge available to an adversary: outsider, insider, or a permanent insider. An outsider is generally not in possession of any special knowledge beyond that available, for instance by eavesdropping on messages between tags and readers. An insider however, is an adversary with access to additional information such as private keys or other securely stored information. An insider may be a permanent insider if the adversary has continual access to secure information.

4 Conclusion

Various solutions proposed for RFID systems have been discussed in [1]. This paper has formulated a simple, yet adequate, security model for evaluating security and privacy services suitable for low cost RFID systems. We have also provided a framework for developing an adversary model suitable for analysing the level of security provided by services devised to address the vulnerabilities of low cost RFID systems. Using the characteristics of the adversaries outlined above it is possible to consider likely adversaries to a given RFID application.

It is now possible to use the framework outlined to analyse the effectiveness of security services using the evaluation framework developed to establish their merits for low cost RFID and the level of security provided using an appropriate adversarial model.

References

1 Ranasinghe, D.C., Cole, P.H.: Adressing Insecurities and Violations of Privacy. In: Ranasinghe, D.C., Cole, P.H. (eds.): Networked RFID Systems and Lightweight Cryptography, Springer-Verlag, Berlin Heidelberg New York (2007)

2 Ranasinghe, D.C., Engels, D.W., Cole, P.H.: Low cost RFID systems: confronting security and privacy. In: Auto-ID Labs White Paper Journal. Vol 1 (2005)

3 Kang, S.M.: Accurate simulation of power dissipation in VLSI circuits. In: IEEE Journal of Solid-State Circuits. Vol. 21(2) (1986)

4 Macci, E., Poncino, M.: Estimating worst-case power consumption of CMOS circuits modelled as symbolic neural networks. In Rabaey, J. M., Chndrakasan, A., Nikolic, B. (eds.): Digital Integrated Circuits, Second Edition, Prentice Hall (2002)
5 Synopsis website. Available from:http://www.synopsys.com/products/mixedsignal/hspice/ hspice.html (2007)
6 Juels, A., Weis, S. A.: Defining Strong Privacy for RFID. In: RSA Laboratories publication (2006)
7 Juels, A.: Minimalist cryptography for low cost RFID tags. In: LNCS. Vol. 3352, Springer-Verlag (2001) 149-164
8 Avoine, G.: Adversary model for radio frequency identification. In: Technical Report, Security and Cryptography Laboratory, Swiss Federal Institute of Technology, Lausanne (2005)
9 Avoine, G.: Radio frequency identification: adversary model and attacks on existing protocols. In: Technical Report LASEC-REPORT-2005-001, September (2005)
10 Beth, T., Frisch, M., Simmons, G. J. (eds.): Public-Key Cryptography: State of the Art and Future Directions. Lecture Notes in Computer Science, Springer-Verlag, (1992)
11 Schnier, B.:Applied Cryptography Protocols: Algorithms, and Source Code in C, John Wiley & Sons, Inc, New York (1994)

Chapter 9

From Identification to Authentication – A Review of RFID Product Authentication Techniques

Mikko Lehtonen[1], Thorsten Staake[2], Florian Michahelles[1], and Elgar Fleisch[1,2]

[1]Information Management, ETH Zurich, 8092 Zurich, Switzerland,
{mlehtonen,fmichahelles}@ethz.ch
[2]Institute of Technology Management, University of St.Gallen,
9000 St.Gallen, Switzerland, {thorsten.staake,elgar.fleisch}@unisg.ch

Abstract. Authentication has an important role in many RFID applications for providing security and privacy. In this paper we focus on investigating how RFID can be used in product authentication in supply chain applications and a review of existing approaches is provided. The different categories of RFID product authentication approaches are analyzed within the context of anti-counterfeiting and fields where future research is needed are identified.

Keywords: Product authentication, RFID, security

1 Introduction

Since the *identification, friend or foe* (IFF) systems used in the Second World War, radio frequency devices have been applied to identify physical objects [1–3]. Nowadays radio frequency identification (RFID) enables automated data gathering in various applications like pallet identification [4], cattle tracing [5] and access control [6]. However, identification itself does not guarantee that the acquired identity corresponds to the genuine identity and thus also verification or validation of the claimed identity – authentication – is needed.[1]

Product authentication in supply chain provides great opportunities to fight illicit trade by detecting counterfeit products. Counterfeiting is a rapidly growing

[1] One should note that this anti-counterfeiting terminology, adopted from biometrics, can conflict with that of cryptology where identification per se can stand for securely establishing the identity of a communicating entity.

world wide problem that affects a great number of industries and harms societies in many ways [7]. Counterfeit players work to get a return on investment for their illegal actions. The overriding requirement of any anti-counterfeiting system is to change the risk-return profile for the counterfeiters – raising the risk and thereby minimizing the return [8]. Product authentication techniques form an important tool in turning the expected return less favorable for the illicit actors thus supporting the legal trade.

In this paper we concentrate on the use of RFID technology in product authentication. Our contribution is to present the requirements of product authentication in supply chain applications and to show how RFID can be used as an enabling technology for product authentication. Our focus is on security and therefore the attack scenarios of counterfeit players and their implications to RFID are presented. Then, categorization and review of existing RFID product authentication approaches are provided. In discussion the presented approaches are analyzed. We finish by identifying fields where future research is needed.

2 Product Authentication in Supply Chain Applications

The role of product authentication is to answer whether a given product is genuine or counterfeit (e.g. product that infringes a trademark). An explicit way to authenticate products is needed in supply chain applications because counterfeits can be very similar or even identical to authentic products. The starting point of automated non-destructive product authentication is to insert a special label or security feature into products, like a hologram or a water mark, and to authenticate this label.

Product authentication can take place in single item-level or in aggregated levels. Generally, multiple similar units are authenticated simultaneously, for example when a shipment arrives to a retail store. The desired level of security, which can be defined as the effort an illicit actor has to undertake to break or bypass the security mechanism, has a major impact on the cost of a product authentication system. While minimizing the cost, the level of security should be high enough to protect the item over its entire life-span. Because different products have very varying security requirements, different levels of security and thus different solutions are needed.

The level of security of product authentication system is defined by the level of security of a single security feature and by the granularity of the security features. By granularity we mean here how many products use an identical security feature; for example, applying weak but unique security features to all products can be more secure than using strong but identical feature on the same products. One conceptual problem of automated product authentication is that it is only the security feature that is authenticated and not the product itself – therefore difference between label and product authentication should be made. The general requirements of product authentication system in supply chain application are listed below:

- The system needs to be used by multiple parties from multiple locations
- Authentication of products that are unknown to the system should be supported
- The cost and effort to perform a check need to be low
- The optimal solution should allow also the customers to authenticate products
- The product authentication system needs to have an appropriate level of security

Among the requirements listed above, the level of security demands most attention in the system design. The level of security can be considered as the resistance against attacks that are conducted against the authentication system. In supply chain applications, product authentication is typically performed under the supervision of authorized personnel. This restricts the possible attacks of counterfeit players. The general attack scenarios of illicit actors against product authentication system can be divided into following four categories:

Omission of security features which are applied on the genuine objects refers to the counterfeiters not taking any explicit actions to fool the authentication. These products form a considerable part of the counterfeit trade for example due to consumer demand of counterfeits.

The use of misleading security features means that the fake products are equipped with security features whose role is to make the products avoid closer inspection. Interviews with brand owners and customs reveal that this scenario together with the aforementioned one is dominant especially for all goods which are mass produced or where the consumers do not regularly check for the object's authenticity.

The removal and reapplying of authentic security features remains a threat in all automated product authentication systems if not explicitly addressed by binding the product and the label. However, because acquiring and reapplying authentic labels is costly, this attack does not threaten authentication systems in large scales.

The cloning and imitation of security features is the most obvious attack that a product authentication system has to resist. As the underlying problem of counterfeits is that the products themselves can be cloned, the first line of defense is to integrate such security features into products that are hard to be replicated.

3 RFID Product Authentication Techniques

RFID has considerable potential in product authentication. The benefits of RFID compared to old authentication technologies include non line-of-sight reading, item-level identification, non-static nature of security features, and cryptographic resistance against cloning. RFID systems in general comprise transponders, readers or interrogators, and online database, sometimes referred to as the back-end server. The potential of RFID in anti-counterfeiting is discussed further in [9] and [10].

There are many applications where RFID transponders are already used for authentication, for example access control. While RFID product authentication is very close to RFID access control when it comes to the authentication protocols used. Product authentication needs specific solutions because of the specific application

requirements discussed in the previous section. RFID product authentication can be based on transponder authentication or identification and additional reasoning using online product data. Furthermore, RFID supports secure ways to bind the label and the product.

To resist cloning and forgery are the most important security properties of authentication tags. The simplest cloning attack against an RFID tag only requires reading the tag serial number and programming the same number into an empty tag. There are two essential obstacles against this kind of replication. First, even the low-cost transponders (e.g. EPC Class-1 Generation-2 [11]) have a unique factory programmed chip serial number (or transponder ID, TID) that is similar to the unique MAC address of PC network cards. To clone a transponder's TID would therefore also require access to hardware manufacturing or a fully programmable tag.

The second obstacle against cloning is to use read-protected secrets residing on tags and to check if the tag knows these secrets, for example by cryptographic challenge-response protocols. Even though this can provide significant improvements to a tag's cloning resistance, there remain many ways to conduct a cloning attack against a single tag. These attacks include *side channel attack* [12], *reverse-engineering and cryptanalysis* [13], *brute-force attack* [14], *physical attacks* [15] and different *active attacks* against the tag [16]. In addition, shared secrets based product authentication approaches are always vulnerable to *data theft,* where the secret PIN codes or encryption schemes of valid products are stolen or sold out by insiders, which would enable criminals to create phony tags. This scenario is especially interesting for adversaries because it would allow them to clone a large number of tags.

Other RFID security issues that have to be considered in product authentication comprise resistance against a denial of service (DoS) attack. In general, DoS causes loss of service to users. Even though it cannot be used to fool the product authentication, it can pose a threat for the overall process. In RFID DoS attack can be conducted, for example, by jamming the readers with hidden blocker tags [17] or by desynchronizing tag and a database entry [18].

We assume that product authentication is normally performed under the surveillance of authorized personnel or by the customer, which narrows down the possible attack scenarios. Therefore active attacks, where the adversary would need to participate in the authentication session and use special devices in the proximity of the reader (e.g. replay, relay and man-in-the-middle attack), are not considered as realistic threats against RFID product authentication.

4 Review of RFID Product Authentication Approaches

In this section we provide a review of existing and proposed RFID product authentication approaches. The approaches are categorized into four categories depending on what the authentication is based on. The approaches presented in

subsections 4.1 and 4.2 authenticate the products without tag authentication, while in approaches presented in subsections 4.3 and 4.4 the tag or the data the tag stores is authenticated.

4.1 Unique Serial Numbering

By definition, one of the fundamental assumptions in identification, and thus also in authentication, is that identified entities possess an identity. In supply chain applications, issuing unique identities can be efficiently accomplished with RFID. We recognize unique serial numbering and confirmation of validity of identities as the simplest RFID product authentication technique. The potential of unique numbering of objects without tag authentication is discussed by Juels in [19]. There the author provides an example from the art world, where a Victorian painter Alma-Tadema evaded the problem of counterfeiting by writing unique serial numbers on his paintings and cataloging the numbers. Product authentication without tag authentication has been proposed also by **Takaragi et al.** [20].

Koh et al. [21] proposed ways of RFID product authentication in the pharmaceutical supply chain. One of the proposals was to keep a list of valid product ID numbers in a secure online server so that the absence of a product's ID from that list would serve as an indication of counterfeit. The security of this approach relies on keeping the list secret from counterfeiters while providing needed access for it to licit parties.

Counterfeiters can always try to guess the valid serial numbers, especially when the numbers are issued in a systematic way. Therefore unique serial numbering can be made more secure by assigning the serial numbers in a random way from a large name space. This is possible with RFID, due to the supported long identifiers. The clear un-addressed weakness of unique serial numbering approaches is tag cloning. However, duplicated tags can be detected and are an important indicator of counterfeit. Furthermore, these approaches can be implemented in RFID enabled supply chain systems with little additional cost, as RFID tags are already being used for pallet and case level identification in large scales [4].

4.2 Track and Trace based Plausibility Check

Track and trace [10, 21, 22] refers to generating and storing inherently dynamic profiles of individual goods as products move through the supply chain. The product specific records allow for heuristic plausibility checks, for example a product with a serial number registered for sale in Switzerland is suspicious if offered in an American store at the same time. The plausibility check is suited for being performed by customers who can reason themselves whether the product is original or not, though it can also be automated by suitable artificial intelligence.

Track and trace is a natural expansion of unique serial numbering approaches. Furthermore, track and trace will be used in supply chains also for other purposes, such as for deriving a product's history and for organizing product recalls. In

addition, some industries like the pharmaceutical industry have legislation that demands companies to document product pedigrees [23]. Therefore track and trace based product authentication can be cost-efficient for companies. However, generating and gathering track and trace profiles of products in multi-party supply chains can be hard and requires cooperation between the partners.

4.3 Secure Object Authentication

Secure object authentication techniques make use of cryptography to allow for reliable authentication while keeping the critical information secret in order to increase resistance against cloning. Because authentication is needed in many RFID applications, the reviewed protocols come from different fields of RFID security and privacy.

One of the first cryptographic privacy enhancing technologies for RFID is the hash-lock of **Weis et al.** [24]. The design principles behind the proposed scheme include the assumption that tags cannot be trusted to store long-term secrets when left in isolation. The authors proposed a way to lock the tag without storing the access key, but only a hash of the key on the tag. The key is stored in a back-end server and can be found using the tag's meta-ID. This approach can be applied in authentication, namely unlocking a tag would correspond authentication. However, the cloning resistance of the scheme is based only on the locked state of the tags and so it is more suitable for protecting privacy. **Henrici et al.** [25] have later extended the randomized version of the original hash-lock scheme [24] for increased privacy and scalability.

Avoine et al. [26] proposed another hash-based RFID protocol that provides modified identifiers for improved privacy and that can be applied for authentication. In the proposed protocol the authors solve scalability issues of the privacy-enhancing scheme from [27] by introducing a specific time-memory trade-off. In addition, hash-based RFID protocols for mutual authentication have been proposed in [28–30]. All these protocols rely on synchronized secrets residing on the tag and back-end server and they require a one-way hash function from the tag. These approaches show how guaranteeing the un-traceability by updating tag identifier increases the workload of back-end servers.

Texas Instruments has developed RFID based authentication techniques for pharmaceutical industry. The model presented in [22] bases on authenticating the products through digital signatures that are written on tags. By using TID and a public-key, the transponder can be linked to the signer of the data in a provable way. To improve the traceability of products, tag memory is also used to store chain-of-custody events.

Juels et al. [31] presented an approach to increase tracing and forgery resistance of RFID-enabled banknotes by using digital signatures for RFID authentication. The approach uses re-encryption to avoid static identifiers and optical data on the banknote to bind the RFID tag and the paper. Authentication is performed by verifying that the data on the tag is signed using a valid public-key. In order to increase cloning resistance, the authors suggest including some distinctive

characteristics of the physical media into the signature (i.e. physical fingerprint of the banknote) and verifying the validity of these characteristics as a part of the authentication process. **Zhang et al.** [32] have later enhanced the protocol by addressing some integrity issues.

Tsudik [33] proposed an authentication protocol called YA-TRAP which provides tracking-resistant tag authentication through monotonically increasing timestamps on the tag. YA-TRAP requires a pseudo-random number generator (PRNG) from the tag and its basic version is vulnerable to DoS attack through timestamp de-synchronization between the tag and the server. The approach does not require on demand computation for the back-end as a result of a pre-computed hash-table for later tag verification, which means less load for the server than for example in [34]. **Chatmon et al.** [35] proposed anonymous RFID authentication protocols based on YA-TRAP that provide anonymity for authenticated transponders and address some vulnerabilities of the original design, while increasing the server workload.

Juels [36] discussed minimalist cryptography based authentication and proposed a tracking-resistant pseudonym-throttling scheme. This mutual authentication protocol bases on a list of pseudonyms and keys residing on tag and on back-end server. The protocol needs additional memory on tag and uses a way to update the tag's pseudonym list using one-time pads to resist cloning and eavesdropping. However, the communication cost is relatively high because of the tag data updates.

Juels proposed another low-cost authentication in [37], where the read-protected 32-bit kill passwords of EPC Class-1 Generation-2 tags are used to implement *ad-hoc* tag authentication protocol. The protocol bases on the fact that even though the EPC of a transponder can be skimmed, the kill-password remains secret. Cloned tags can be found by testing, without killing the tag, if the kill password matches the original one stored in a database. Furthermore, the protocol supports for mutual authentication.

Vajda et al. [38] discussed lightweight authentication protocols for low-cost tags. The proposed set of challenge-response protocols includes simply XOR encryption with secret keys (although also complex encryption like RSA was proposed, it's not considered here because it's infeasible in low-cost tags [39]). The cryptographic problem with keys being static in XOR encryption is addressed by re-keying schemes that make use of keys from multiple previous protocol runs.

Juels et al. [39] introduced an approach for low-cost authentication based on the work of Hopper and Blum (HB) [40]. The proposed HB^{+} protocol makes use of the hardness assumption of statistical "Learning Parity with Noise" (LPN) problem and can be implemented on low-cost tags, as it only requires bitwise AND and XOR operations and one random "noise bit". The security of HB+ against active adversaries has gained publicity in the scientific community and is discussed in details in [41]. The first version of the original protocol [39] was found to be vulnerable against a realistic active attack [16]. Proposals to address the security issues have emerged, including the modified HB^{++} by **Piramuthu** [42].

Dimitriou [43] proposed a protocol that addresses privacy issues and aims at efficient identification of multiple tags. The enhanced version of the protocol is considered here, since the basic one does not protect the tags against cloning. In

this approach the tags need a PRNG and a pseudo random function (PRF) for symmetric-key encryption. The proposed protocol is efficient in terms of tag-to-reader transaction and protects the privacy by avoiding transmission of static IDs. However, since the tags share secret keys, compromise of one tag may reveal information about others. In another work [44] the author proposed a lightweight RFID protocol against traceability and cloning attacks. This approach bases on a refreshing a shared secret between tag and back-end database and requires hash calculations and PRNG from the tag.

Duc [45] proposed communication protocol for RFID devises that supports for tag-to-reader authentication based on synchronization between tag and back-end server. The proposed scheme is tailored for EPC Class-1 Generation-2 tags so that it requires only a PRNG on the tag and pre-shared keys. The approach also takes advantage of the CRC function that is supported by Generation-2 tags. The underlying idea is to use the same PRNG with the same seed on both RFID tag and on back-end side and to use it for efficient key sharing. The encryption and decryption can then be done by XORring the messages.

Ranasinghe et al. [46] presented ways to implement challenge-response authentication protocol on RFID tags without using costly cryptographic primitives. These proposals are based on a Physical Unclonable Function (PUF) residing on the tag, which allows for calculation of unique responses using only some hundreds of logical gates. A possible candidate for the PUF can be found from [47], where the manufacturing variations of each integrated circuit are used to implement a secret-key on a tag. The back-end server needs to store a list of challenge-response pairs for each PUF (i.e. for each tag) because, without encryption, a PUF challenge-response pair that is once used, can not be used again since it may have been observed by an adversary. The PUF based security is still an area of active research. Also **Tuyls et al.** [48] proposed the use of PUFs to increase RFID transponders resistance against both physical and communication based cloning attacks and defined an offline authentication protocol. The authors estimated that their anti-clone tag can be built with on the order of 5,000 gates.

Engberg et al. [49] proposed so called zero-knowledge device authentication as an answer to consumer privacy issues. In their proposal the tag must authenticate the reader before it returns any traceable identifier. The scheme is based on shared secrets and requires hash function from the tag. Also **Rhee et al.** [50] proposed a challenge-response protocol for user's privacy. The proposed protocol doesn't update the tag ID and therefore can be applied in an environment with distributed databases. The protocol relies on hash calculations by the back-end database, so that the tag ID is the only necessary shared secret between the devices taking part in the authentication.

Molnar et al. [51] proposed private authentication protocols for library RFID, where the tag and the reader can do mutual authentication without revealing their identities to adversaries. The protocols made use of PRNG residing on the tag. **Molnar et al.** presented in [34] another privacy enhancing scheme where an RFID pseudonym protocol takes care of emitting always a different pseudonym using PRF. In order to relate pseudonyms and real tag IDs, the authors presented an entity called Trusted Center (TC) that is able to decode the tag responses and obtain the tag's identity. In the same work the authors introduced term *ownership*

transfer that refers to TC giving permissions to only readers of a certain entity to read an RFID tag.

Gao et al. [52] proposed protocols for improved security and privacy of supply chain RFID. In their proposals the tags store a list of licit readers to protect the tags against skimming and therefore need rewritable memory. Other tag requirements include PRNG and hash function. Though the protocol burdens the back-end server with some computational load, the approach is designed to be suitable for a large number of tags. **Yang et al.** [53] proposed a mutual authentication protocol that provides protection against replay attack and MITM attack even when the reader is not trusted and the communication channel is insecure. This mutual authentication protocol provides privacy protection and cloning resistance with the expense of tag's hash calculations and storing two secrets in the tag and in the back-end server.

Dominikus et al. [54] discussed symmetric RFID authentication protocols in practice and presented five standard challenge-response protocols for reader, tag and mutual authentication. The design focuses on strong authentication for advanced, about 50 ¢ tags with available silicon area of 10,000 gates. The presented protocols use AES encryption (and decryption) on tags in such a way that energy constraints of Class-2 RFID systems are met.

Feldhofer [55] presented an implementation of standard symmetric two-way challenge-response protocol as extinction to the standard ISO/IEC 18000 RFID protocol. The use of standard authentication protocols with standard communication protocols is important for ensuring the security and interoperability of an approach. Hardware implementation of the same protocol can be found from [56], where **Feldhofer et al.** presented a novel minimalist approach of a 128-bit Ad_vanced Encryption Standard (AES) implementation. The approach provides a promising choice for strong authentication in RFID systems and the proposed low-cost AES hardware implementation is used in various other proposals as an enabler of cost-efficient RFID cryptography.

Also **Bailey et al.** [57] concentrate on integrating common cryptographic standards into RFID by proposing techniques to create RFID tags that are compliant with the EPC Class-1 Generation-2 tags, but offer cryptographic functionality of standards like ISO 7816-4. The proposed challenge-response protocols make use of AES on the tag and can be used for mutual authentication. In particular, the authors define a 32 or 64 bit "one-time password" that could be included in transmitted EPC data fields.

4.4 Product Specific Features

To explicitly address transponder removal and reapplying (but also tag cloning) attacks with low-cost tags, **Nochta et al.** [58] proposed a cryptographic way to bind the RFID transponder and the product that it authenticates. Because of the uniqueness of the approach, we consider it as a separate category of RFID product authentication. In this approach the authentication is based on writing on the tag memory a digital signature that combines the TID number and product specific

features of the item that is to be authenticated. These features can be physical or chemical properties that identify the product and that can be verified, such as very precise weight. The chosen feature is measured as a part of the authentication and if the feature used in the tag's signature does not match the measured feature, the transponder-product pair is not original.

The proposed authentication needs a public-key stored on an online database. Also an offline authentication is proposed by storing the public-key on the tag, though this decreases the level of security. The disadvantage of this approach is that each unit has to be physically verified as a part of authentication.

5 Tag Requirements for Authentication

In order to evaluate RFID product authentication in practice, the cost of authentication needs to be considered. One of the most important cost drivers of RFID product authentication is transponder cost that is, for its part, mostly defined by the complexity of the chip (or integrated circuit, IC). The complexity of the chip can be described by several informal metrics [59] like the number of transistors or the *gate equivalent (GE),* or *gate count,* that is about a fourth of the number of transistors. The gate count of current low-cost transponders is 5,000 – 10,000 [53, 60], limiting their computational power to only a fraction of that of computers. In addition, the number of gates available for security features is even smaller and estimated to be below 2,000 [61] or below 5,000 [53]. The rule of thumb of gate cost says that every extra 1,000 gates increase the chip price by 1 ¢ [61].

In order to be able to evaluate the transponder cost more precisely, we quantify the transponder's technical requirements. The requirements we consider include first of all additional non-volatile memory (NVM) which is typically EEPROM. Different types of NVM exist: factory-programmable memory (or read only), field programmable memory (or write-once-read-many, WORM) and read-write (RW) memory. Other requirements relate to the transponder's ability to perform logical operations. Logical functionalities can be implemented on chips basically by increasing the gate count and they include first of all the ability to perform primitive bitwise operations (e.g. AND, XOR) that can be implemented with a small number of gates. Other requirements include hash function that is a common cryptographic primitive but so far out of the scope for low-cost RFID transponders – standard cryptographic hash functions like SHA-1 need roughly 20,000 gates [61]. Weis discusses non-linear feedback shift registers as one possible low-cost hash function [61] as it has no complex hardware requirements (besides the register). Interestingly, Yüksel [62] presented implementations of low-cost hash functions, taking only 1,700 gates for block size of 64 bits.

Another tag requirement is pseudo-random number generator (PRNG) that can be implemented for example by keying a hash function. However, it is still unclear how and when adequate PRNG can be deployed on inexpensive RFID tags [24, 63]. The final tag requirement considered is symmetric key encryption, or in general pseudo random function (PRF). Public-key encryption is not considered because it is too expensive for RFID transponders [54]. A common example of

Table 1 Summary of example transponders' resources

	NVM	RW	Bitwise Operations	PRNG	Hash function	Symmetric key encryption
Label Tag	64 bits					
Smart Label	96 bits	Yes	Yes	16 bit		
Crypto Tag	256 bits	Yes	Yes	64 bit		Yes

symmetric key encryption is the Advanced Encryption Standard (AES) block-cipher which can be used to encrypt data using a secret-key. Hardware implementations of AES take on the order of 20,000 – 30,000 gates [61], which seemed to constrain it out of the scope of low-cost transponders till a few years ago. However, Feldhofer et al. [56] presented an implementation of 128 bit AES encryption which requires only 3,600 gates (and 256 bits RAM) which is considerably fewer than the smallest AES circuit published so far, bringing cost-efficient strong authentication closer to reality for RFID tags.

To illustrate available tag resources for product authentication, the properties of three example tags are summarized in Table 1. The simplest example tag, denoted *label tag*, provides only a factory programmed label, like the EPC Class-0 [64]. This tag can be used in approaches where only tag identification is required (subsections 0 and 0). The more advanced *smart label* presents an EPC Class-0 Generation-2 tag [11] with RW memory (even though Class-1 tags were originally designed for WORM memory [65], also tags with RF memory are available, e.g. [66]). The cheapest EPC Class-1 tags cost on the order of 15 ¢ in high volumes [67]. The *crypto tag* presents an advanced (e.g. Class-2) transponder. Such tags cost about 50 ¢ and have silicon area of about 10,000 gates [54].

6 Discussion

In this paper, we have provided a review of existing RFID product authentication techniques. Four categories of approaches are distinguished based on what is the reasoning behind the check. In general, either the transponder is authenticated or the reasoning is based on identification and additional information in online databases.

The focus of the review is on cryptographic secure object authentication approaches which are by far the most discussed category of RFID authentication techniques within the scientific community. This is partly explained by the fact that the considered secure protocols originate in the field of RFID security and privacy in general and thus they can be applied in transponder and product authentication also. The main motivation to use cryptographic tags for product authentication and anti-counterfeiting is the increased cloning resistance. Even though secure object authentication approaches remain vulnerable to many attacks that can enable tag

cloning, they can provide a significantly improved level of security for original products.

However, also other, potentially more cost-efficient solutions exist – for example, also a reliable way to find the duplicated tags could be used to make cloning non-profitable for counterfeiters. The presented categories of low-cost product authentication approaches are unique serial numbering and track and trace based plausibility checking. Even though these approaches do not prevent tag cloning, also they can be used to significantly increase the barrier of counterfeit players to distribute fake products. The better cost-efficiency of these approaches compared to cryptographic techniques is supported by two facts. First, they need only low-cost tags and they support relatively simple authenticity checks. Second, unique serial numbering and track and trace are used also in other supply chain applications and so authentication is not the only application responsible for the hardware costs, whilst the increased transponder costs of secure object authentication approaches must be justified entirely by the increased cloning resistance.

All the approaches presented in the review provide a careful trade-off between complexity and security. In order to evaluate the optimal product authentication system for anti-counterfeiting, the costs and benefits of different techniques have to be evaluated. As stated in the introduction, the overriding requirement of any anti-counterfeiting system is to change the risk-return profile for the counterfeiters. The counterfeiter will carry out some form of direct or in-direct cost-benefit analysis before embarking on criminal enterprises [8]. Product authentication increases the illicit players' risk of getting caught and decreases the number of counterfeit products in the market. The affected companies will benefit from this for example through additional sales. Though the precise mechanism of how companies benefit from product authentication is very hard to be quantified, security of authentication plays an important role as an enabler of those benefits. Therefore the appropriate way to compare different product authentication approaches for anti-counterfeiting is to consider their security and cost.

Security of RFID product authentication can be evaluated by considering cloning resistance, ability to detect cloned tags and resistance against tag removal and reapplying. Active attacks against readers are not considered as realistic threats against product authentication system. The cost of an approach can be evaluated by considering the general complexity of the check and the cost of transponder. Table 2 summarizes these abovementioned properties of the four general product authentication categories. For more detailed comparison, a comprehensive summary of technical requirements of the presented approaches is presented in Table A-1 (Appendix A).

Based on the discussion so far, unique serial numbering and track and trace based approaches are most probable to provide convenient authentication techniques for consumer goods and other low-cost items; secure object authentication techniques can be applied for more expensive products when tag cloning needs to be addressed. However, there are also promising low-cost methods to increase the cloning resistance of all RFID tags, such as the use of unique transponder ID number, which could make cryptographic tags unnecessary for most product categories.

Table 2 Comparison of different product authentication categories

Approach	Complexity of check	Cost of tag	Cloning resistance	Clone detection	Tag reapplying resistance
Serial Numbering	Low	Low	No	No	No
Track and Trace	Medium	Low	No	Yes	Yes
Secure Authentication	Medium-High	Low-High	Yes	No	No
Product Specific Features	High	Low	Yes	No	Yes

The cost of cryptographic product authentication transponders (e.g. the crypto tag, Table 1) will be determined by the development of minimalist hardware implementations of two most important cryptographic primitives, hash functions and pseudo random functions. The importance of these functions as enablers of secure object authentication approaches can be clearly seen in Table A-1. However, the development of secure protocols that can be implemented using only simple bitwise operations on tags can create a family of truly low-cost tags (e.g. the smart label, Table 1) for secure authentication.

Finally, the review reveals that offline authentication remains unsolved as practically all existing techniques need online servers. Current development of RFID protocols is driven mostly by privacy concerns and the goal is often an efficient use of back-end server to protect customers against tracing. Attempts to be independent from a network are rare and they might need further development from the field of physically unclonable functions. Also many network issues remain unsolved. The open questions include key distribution, scalability, generation of track and trace profile in multi-partner environment, ownership transfer and the need for trusted third parties. Furthermore, as in all RFID applications, the role of standards is of primary importance in product authentication and should be taken into account in solution design.

7 Conclusions

This work shows that there is no silver-bullet approach for moving from radio-frequency identification to authentication and therefore accurate and well justified ways to compare the different techniques are needed. The focus of recent development in RFID authentication has been on consumer privacy, but product authentication needs also specific solutions to address the application requirements. Further research is still needed in the field of offline authentication and many network issues, before RFID product authentication will meet all its promises in practice.

References

1 Lampe, M., Strassner, M.: The potential of RFID for moveable asset management. In: Workshop on Ubiquitous Commerce at Ubicomp (2003)

2 NJE Consulting: RFID in waste management (2006). Available from: http://www.nje.ca/Index_RFIDWasteManagement.htm (22.6.2006)

3 RFID in Japan: Shoe RFID expands. News Article, July 10, 2005 (via Nikkei Ryutsu Shimbun MJ, July 6, 2005). Available from: http://ubiks.net/local/blog/jmt/archives3/004067.html (22.5.2006)

4 RFID Journal: Wal-Mart draws line in the sand. News Article, June 11 (2003). Available from: http://www.rfidjournal.com/article/articleview/462/1/1/ (11.5.2006)

5 RFID Journal: Can RFID save the cattle industry? Vertical Focus, December 23 (2003). Available from: http://www.rfidjournal.com/article/articleview/1032 (22.5.2006)

6 RFID Journal: Long-range RFID for access control. News Article, July 8 (2003). Available from: http://www.rfidjournal.com/article/articleview/493/1/1/ (22.6.2006)

7 International Chamber of Commerce: IP Roadmap 2005: Current and emerging intellectual property issues for business. ICC, Paris (2005) 52. Available from: http://www.iccwbo.org/iproadmap/ (19.5.2006)

8 Organization for Economic Co-operation and Development (OECD): The Economic Impact of Counterfeiting (1998). Available from: http://www.oecd.org/dataoecd/11/11/2090589.pdf (3.5.2006)

9 U.S. Food and Drug Administration: Combating Counterfeit Drugs – A Report of the Food and Drug Administration, February (2004). Available from: http://www.fda.gov/oc/initiatives/counterfeit/report02_04.html (2.5.2006)

10 Staake, T., Thiesse, F., Fleisch, E.: Extending the EPC network – the potential of RFID in anti-counterfeiting. In: Proceedings of the 2005 ACM Symposium on Applied Computing (2005) 1607 – 1612

11 EPCglobal: Class-1 Generation-2 UHF RFID Conformance Requirements Specification v. 1.0.2. EPCglobal public document, February (2005)

12 RFID Journal: EPC tags subject to phone attacks. News Article, February 24 (2006). Available from: http://www.rfidjournal.com/article/articleview/2167/1/1/ (4.5.2006)

13 Bono, S., Green, M., Stubblefield, A., Juels, A., Rubin, A., Szydlo, M.: Security analysis of a cryptographically enabled RFID device. Pre-print (2006). Available from: www.rfidanalysis.org (4.5.2006)

14 RFID Journal: RFID, Privacy and Corporate Data. Feature Article, June 2 (2003). Available from: http://www.rfidjournal.com on subscription basis

15 Weingart, S.: Physical security devices for computer subsystems: A survey of attacks and defense. In: Koc, C.K., Paar, C., (ed.): Lecture Notes in Computer Science, Vol. 1965. Springer-Verlag, Berlin Heidelberg New York (2000) 302–317

16 Gilbert, H., Robshaw, M., Sibert, H.: An active attack against HB^+ – a provably secure lightweight authentication protocol. Manuscript, July (2005)

17 Juels, A., Brainard, J.: Soft blocking: Flexible blocker tags on the cheap. In: Vimercati S.de.C., Syverson, P., (ed.): Workshop on Privacy in the Electronic Society – WPES, Washington. ACM, ACM Press October (2004) 1–7

18 Kang, J., Nyang, D.: RFID authentication protocol with strong resistance against traceability and denial of service attacks. In: Molva, R., Tsudik, G., Westhoff, D. (ed.): European Workshop on Security and Privacy in Ad hoc and Sensor Networks – ESAS'05, Lecture Notes in Computer Science, Vol. 3813. Springer-Verlag, Berlin Heidelberg New York (2005) 164–175

19 Juels, A.: RFID security and privacy: A research survey. Condensed version to appear in 2006 in the IEEE Journal on Selected Areas in Communication (2005)
20 Takaragi, K., Usami, M., Imura, R., Itsuki, R., Satoh, T.: An ultra small individual recognition security chip. In: IEEE Micro, November–December (2001)
21 Koh, R., Schuster, E., Chackrabarti, I., Bellman, A.: Securing the pharmaceutical supply chain. In: White Paper, Auto-ID Labs, Massachusetts Institute of Technology (2003)
22 Pearson, J.: Securing the pharmaceutical supply chain with RFID and public-key infrastructure (PKI) technologies. Texas Instruments White Paper, June (2005). Available from: http://www.ti.com/rfid/docs/docntr.shtml (28.4.2006)
23 RFID Journal: Congress weighs drug anti-counterfeiting bill. News Article, March (2005). Available from: http://www.rfidjournal.com/article/articleview/2180/1/1/ (19.5.2006)
24 Weis, S., Sarma, S., Rivest, R., Engels, D.: Security and privacy aspects of low-cost radio frequency identification systems. In: Hutter, D., Müller, G., Stephan, W., Ullmann, M. (ed.): International Conference on Security in Pervasive Computing - SPC 2003, Lecture Notes in Computer Science, Vol. 2802. Springer-Verlag, Berlin Heidelberg New York (2003) 454–469
25 Henrici, D., Müller, P.: Hash-based enhancement of location privacy for radio-frequency identification devices using varying identifiers. In: Sandhu, R., Thomas, R. (ed.): International Workshop on Pervasive Computing and Communication Security – PerSec (2004)
26 Avoine, G., Oechslin, P.: A scalable and provably secure hash based RFID protocol. In International Workshop on Pervasive Computing and Communication Security – PerSec, Kauai Island, Hawaii (2005)
27 Ohkubo, M., Suzuki, K., Kinoshita, S.: Cryptographic approach to “privacy-friendly” tags. In RFID Privacy Workshop, MIT, MA, USA, November (2003)
28 Lee, S.M., Hwang, Y.J., Lee, D. H., Lim, J.I.: Efficient authentication for low-cost RFID systems. In: Gervasi, O., Gavrilova, M., Kumar, V., Lagana`a, A., Lee, H.P., Mun, Y., Taniar, D., Tan, C.J.K., (ed.): International Conference on Computational Science and its Applications - ICCSA 2005. Lecture Notes in Computer Science, Part I, Volume 3480, Springer-Verlag, Singapore (2005) 619—627
29 Choi, E.Y., Lee, S.M., Lee, D. H.: Efficient RFID authentication protocol for ubiquitous computing environment. In International Workshop on Security in Ubiquitous Computing Systems – Secubiq. Lecture Notes in Computer Science, Springer-Verlag, Nagasaki, Japan (2005)
30 Lee, S., Asano, T., Kim, K.: RFID Mutual Authentication Scheme based on Synchronized Secret Information. In Symposium on Cryptography and Information Security, Hiroshima, Japan, January (2006)
31 Juels, A. Pappu., R.: Squealing Euros: Privacy Protection in RFID-Enabled Banknotes. In: Wright R. (ed.): Financial Cryptography – FC’03. Lecture Notes in Computer Science, Volume 2742, Springer-Verlag, Le Gosier, Guadeloupe, French West Indies, IFCA (2003) 103—121
32 Zhang, X., King, B.: Integrity Improvements to an RFID Privacy Protection Protocol for Anti-counterfeiting. In: Zhou, J., Lopez, J., Deng, R., Bao, F. (ed.): Information Security Conference – ISC 2005. Lecture Notes in Computer Science, Vol. 3650, Springer-Verlag, Singapore (2005) 74–481
33 Tsudik, G.: YA-TRAP: Yet another trivial RFID authentication protocol. In International Conference on Pervasive Computing and Communications – PerCom 2006, Pisa, Italy, March (2006)

34 Molnar, D., Soppera, A., Wagner, D.: A scalable, delegatable, pseudonym protocol enabling ownership transfer of RFID tags. Handout of the Ecrypt Workshop on RFID and Lightweight Crypto, July (2005)
35 Chatmon, C., Le, T.V., Burmester, M.: Secure anonymous RFID authentication protocols. Technical Report TR-060112, Florida State University, Department of Computer Science, Tallahassee, Florida, USA, (2006)
36 Juels, A.: Minimalist cryptography for low-cost RFID tag. In Conference on Security in Communication Networks – SCN'04, LNCS, Amalfi, Italia, September (2004) Springer-Verlag
37 Juels, A.: Strengthening EPC Tags Against Cloning. In: Jakobsson, M., Poovendran, R. (ed.): ACM Workshop on Wireless Security (2005) 67—76
38 Vajda, I., Buttyán, L.: Lightweight authentication protocols for low-cost RFID tags. Workshop on Security in Ubiquitous Computing (2003)
39 Juels, A., Weis, S.: Authenticating pervasive devices with human protocols. In: Shoup, V. (ed.): Advances in Cryptology – CRYPTO'05. Lecture Notes in Computer Science, Vol. 3126, IACR, Springer-Verlag, Santa Barbara, California (2005) 293–308
40 Hopper, N., Blum, M.: A Secure Human-Computer Authentication Scheme. Tech. Rep. CMU-CS-00-139, Carnegie Mellon University (2000)
41 Katz, J., Shin, J.S.: Parallel and concurrent security of the HB and HB+ protocols. In Serge Vaudenay, editor, Advances in Cryptology – EUROCRYPT'06, Lecture Notes in Computer Science, IACR, Springer-Verlag, Saint Petersburg, Russia (2006)
42 Piramuthu, S.: HB and related lightweight authentication protocols for secure RFID tag/reader authentication. In Collaborative Electronic Commerce Technology and Research – CollECTeR 2006, Basel, Switzerland (2006)
43 Dimitriou, T.: A Secure and Efficient RFID Protocol that could make Big Brother (partially) Obsolete. In International Conference on Pervasive Computing and Communications – PerCom 2006, Pisa, Italy, March (2006)
44 Dimitriou, T.: A Lightweight RFID Protocol to protect against Traceability and Cloning attacks. In Conference on Security and Privacy for Emerging Areas in Communication Networks – SecureComm, Athens, Greece, September (2005)
45 Duc, D.N., Park, J., Lee, H., Kim, K.: Enhancing Security of EPCglobal Gen-2 RFID Tag against Traceability and Cloning (2006)
46 Ranasinghe, D., Engels, D., Cole, P.: Security and privacy: Modest proposals for low-cost RFID systems. In Auto-ID Labs Research Workshop, Zurich, Switzerland (2004)
47 Lee, J., Lim, D., Gassend, B., Suh, G.E., Dijk, M., Devadas, S.: A Technique to Build a Secret Key in Integrated Circuits for Identification and Authentication Applications. Symposium on VLSI circuits (2004)
48 Tuyls, P., Batina, L.: RFID-tags for Anti-Counterfeiting. In: Pointcheval, D. (ed.): Topics in Cryptology – CT-RSA – The Cryptographers' Track at the RSA Conference. Lecture Notes in Computer Science, No. 115–131, Springer Verlag, San Jose (2006) 3860
49 Engberg, S., Harning, M., Damsgaard-Jensen, C.: Zero-knowledge device authentication: Privacy & security enhanced RFID preserving business value and consumer convenience. In Conference on Privacy, Security and Trust – PST, New Brunswick, Canada, October (2004)
50 Rhee, K., Kwak, J., Kim, S., Won, D.: Challenge-response based RFID authentication protocol for distributed database environment. In: Hutter, D., Ullmann, M. (ed.): International Conference on Security in Pervasive Computing – SPC 2005. Lecture Notes in Computer Science, Vol. 3450, Springer-Verlag, Boppard, Germany (2005) 70–84,

51 Molnar, D., Wagner, D.: Privacy and Security in Library RFID: Issues, Practices, and Architectures. In: Pfitzmann, B., Liu, P. (ed.): Conference on Computer and Communications Security – ACM CCS, pages 210–219, Washington, DC, USA, October (2004)
52 Gao, X., Xiang, Z., Wang, H., Shen, J., Huang, J., Song, S.: An approach to security and privacy of RFID system for supply chain. IEEE International Conference on E-Commerce Technology for Dynamic E-Business (2004)
53 Yang, J., Park, J., Lee, H., Ren, K., Kim, K.: Mutual authentication protocol for low-cost RFID. Handout of the Ecrypt Workshop on RFID and Lightweight Crypto, July (2005)
54 Dominikus, S., Oswald, E., and Feldhofer, M.: Symmetric authentication for RFID systems in practice. ECRYPT Workshop on RFID and Lightweight Crypto, Graz, Austria, July (2005)
55 Feldhofer, M.: A Proposal for Authentication Protocol in a Security Layer for RFID Smart Tags. Stiftung Secure Information and Communication Technologies SIC (2003)
56 Feldhofer, M., Dominikus, S., Wolkerstorfer, J.: Strong authentication for RFID systems using the AES algorithm. Workshop on Cryptographic Hardware and Embedded Systems – CHES 2004. Lecture Notes in Computer Science, Vol. 3156, Springer-Verlag (2004) 357−370
57 Bailey, D., Juels, A.: Shoehorning security into the EPC standard. Manuscript in submission, January (2006)
58 Nochta, Z., Staake, T., Fleisch, E.: Product Specific Security Features Based on RFID Technology. Saint-w, pp. 72−75, International Symposium on Applications and the Internet Workshops (SAINTW'06) (2006)
59 Sarma, S.: Towards the 5¢ Tag. White Paper, Auto-ID Center, MIT (2001). Available from: http://www.autoidlabs.org/whitepapers/mit-autoid-wh-006.pdf (5.5.2006)
60 Sarma, S., Weis, S., Engels, D.: Radio-Frequency Identification: Security Risks and Challenges. In RSA Laboratories Cryptobytes, Vol. 6, No. 1 (2003)
61 Weis, S.: Security and Privacy in Radio-Frequency Identification Devices. Master's Thesis, MIT, May (2003)
62 Yüksel, K.: Universal Hashing for Ultra-Low-Power Cryptographic Hardware Applications. Master's Thesis, Dept. of Electronical Engineering, WPI, (2004)
63 Juels, A., Syverson, P., Bailey, D.: High-Power Proxies for Enhancing RFID Privacy and Utility. In Workshop on Privacy Enhancing Technologies (2005)
64 EPCglobal: 900 MHz Class 0 Radio Frequency (RF) Identification Tag Specification. EPCglobal public document, February (2003)
65 EPCglobal: Class-1 Generation-2 UHF air interface protocol standard version 1.0.9. EPCglobal public document, January (2005). Available from: http://www.epcglobalinc.org/standards_technology/EPCglobal2UHFRFIDProtocolV109122005.pdf (8.5.2006)
66 Alien Technology: EPC Class 1 RFID Tags Datasheet (2005). Available from: http://www.alientechnology.com/products/documents/alien_915mhz_128_bit.pdf (19.5.2006)
67 Supply Chain Digest: RFID News: Tag Prices Drop, but is it Real? Wal-Mart, Target Push for Sunsetting Class 0 and 1 Tags. News and Views, October 13 (2005). Available from: http://www.scdigest.com/assets/newsviews/05-10-13-2.cfm (19.5.2006)

A Summary of Technical Requirements of Different Approaches

We assume that tags always carry an ID number, such as EPC. Considered tag memory requirements include additional non-volatile memory (NVM) and read-write (RW) capability. Functional requirements include tag's ability to perform basic *bitwise operations*, pseudo-random number generator (PRNG), *hash function*, and *symmetric key encryption*. For the sake of simplicity we assume that all approaches that require any of the last three functionalities implicitly require also bitwise operations.

The network requirements include the needed level of secrecy for the online data. This data can be *public* (e.g. a public-key), *secret* (e.g. a secret-key), or *semi-public* when it is not or it cannot be kept completely secret due to its nature, such as tag serial number. The level of secrecy of back-end data affects how easily an approach can be implemented – if the authentication cannot be performed without access to secret data, for example, mode complex system is required than when only public data is used. The final network requirement considered is the need to update data or to perform computations on the server relating the authentication process. This requirement is referred to as *complex database*. Reader requirements include *complex reader* which refers to need to perform computations (e.g. encryption) on the reader side. *Physical verification* stands for the need to verify a physical property of the product as a part of the authentication.

Table A-1. Comparison of technical requirements of different product authentication approaches (dashed lines separate categories, subsections 4.1 – 4.4)

Approach	*Tag Memory Requirements*		*Tag Functional Requirements*				*Network Requirements*				*Reader Requirements*	
	NVM	*RW*	*Bitwise operations*	*PRNG*	*Hash function*	*Symmetric key encryption*	*Public*	*Semi public*	*Secret*	*Complex database*	*Complex reader*	*Physical verification*
Unique Serial Numbering [20.21]								X				
Track and Trace [10, 21, 22]								X		X		
Juels [37]									X			
Ranasinghe [46] [1] *(3.A)*									X			
Pearson [22]	X	X					X				X	
Juels et al. [31]	X	X					X				X	(X) [2]
Zhang et al. [32]	X	X					X				X	
Juels [36]	X	X	X						X	X		
Vajda et al. [38] (4.1-4.3)	X	X	X						X		X	
Tuyls et al. [48] [1]	X		X								X	
Juels et al. [39]	X		X	(X) [3]					X		X	
Piramuthu [42]	X		X	(X) [3]					X		X	
Tsudik [33]	X	X	X	X					X	X		
Chatmon et al.[35]	X	X	X	X					X	X		
Duc [45]	X	X	X	X					X	X		
Molnar et al. [51]	X		X	X					X		X	
Engberg et al. [49] (III. A)	X		X		X				X		X	
Avoine et al. [26]	X	X	X		X				X	X		
Gao et al. [52]	X	X	X	X	X				X	X	X	
Rhee et al. [50]			X	X	X				X	X		
Dimitriou [44]	X	X	X		X				X	X		
Yang et al. [53]	X	X	X		X				X	X	X	
Weis et al. [24] (5.1.)	X		X		X				X			
Henrici et al. [25]	X	X	X		X				X	X		
Lee et al. [30]	X	X	X	X	X				X	X		
Choi et al. [29]	X	X	X		X				X	X		
Lee et al. [28]		X	X		X				X	X		
Molnar et al. [34]	X	X	X	X		X			X	X		
Dimitriou [43]	X		X	X		X			X	X		
Feldhofer [55],[56]	X		X			X			X		X	
Bailey et al. [57]	X		X	X		X			X		X	
Dominikus et al. [54]	X		X	(X) [4]		X			X		X	
Nochta et al. [58]	X						(X) [4]				X	X

[1] *the transponder needs a physical unclonable function (PUF);* [2] *optional;* [3] *only one random bit is required,* [4] *not necessary for all proposed approaches*

Part III
Network Based Solutions

Chapter 10

EPC System for a Safe & Secure Supply Chain and How it is Applied

Tatsuya Inaba[1]

[1] 5322 Endo, Fujisawa, Kanagawa 252-8520 Japan. tinaba@autoid.sfc.keio.ac.jp

Abstract. Threats in supply chains, such as counterfeiting, product piracy and product recall, are ubiquitous, and Japan is no exception to this trend. In addition, these threats are not limited to industrial products; supply chains of agricultural products are also under threat. In order to eliminate these threats, various efforts have been made, some of which are the applications enabled by the EPC System, a technology that connects the physical world with the information world. In this paper, we first analyze safe and secure supply chain issues in Japan and identify the fundamental issues through abstracting these issues. Then, we analyze how the EPC System works effectively to deal with these issues and propose potential research topics that can enhance the security level of supply chains. Although this study starts from issues in Japan, since those issues can be generalized, the analyses and proposals are applicable to issues in other countries/regions.

Keywords: RFID, Anti-counterfeit, SCM, Security, EPC System

1 Introduction

Safety and security of the supply chain is one of the concerns in Japanese society, and efforts to improve safety and security have been made by both public and private sectors for years [1], [2], [3]. The importance of safety and security of the supply chain is often argued in the context of food safety and security. This is because the society learned the importance from the past bitter experiences, such as the Bovine Spongiform Encephalopathy (BSE or Mad Cow Disease) epidemic and the fake labeling of agricultural products issues [4], [5]. In the arguments of these issues, two points are often highlighted. The first is whether producers/manufacturers make products appropriately or not, and the second is whether the products are distributed securely. In addition, since we learned that the fact that authentic products are traded securely in the legitimate supply chain is not sufficient from the drug contamination case, the argument of the traceability after shipment from

producers/manufacturers is included in the second point. We make these arguments as a background and categorize safe and secure supply chain issues into three: 1) issues about the authenticity of the products (counterfeit), 2) issues about the legality of the product trade (illegal trade), and 3) issues about the status of the product after shipment (wrong status).

Regarding the counterfeit issues, fake products made outside of Japan and brought to Japan becomes increasingly common these days. Since counterfeiting is sophisticated, the actual damage is not known. According the World Trade Organization (WTO), total damage of counterfeiting in Japan would be about 2.3 trillion Japanese yen [6]. In addition to this fake product issue, we include issues regarding labels on the product, such as re-labeling, substitution, fake labeling, as parts of counterfeit issues.

Illegal trade consists of gray market and black market, both of which are also problems in Japan. A gray market is created when products are sold in the different markets through different channels from the market and channels through which the original manufacturers and the authorized distributors intend to distribute. With the progress of international parcel services, e-business and Internet auction, the way products are shipped become dynamic and it is difficult for the original manufacturers and authorized distributors to keep track of their products, which could be a foundation of the gray market. Not only that, this complexity of the distribution channels also can cause emergence of the black market.

The ultimate goal of realizing safe and secure supply chains is to deliver safe and secure products to the end consumers. Because of this, the ability to guarantee the quality of the products and eliminate wrong status products even after shipment from the manufacturers or producers is crucial. Wrong status can be both a short term issue and a long term issue. For short term, wrong status could be expired products or mishandling of the inappropriate products, such that low quality products somehow enter the legitimate supply chain. On the other hand, long term issues include the issue that products have some kind of defects that are not known now but will be known in the future. Contaminated drugs are a good example of this long term issue.

These are the issues about the "products," which are categorized into "physical objects." However, these issues are not limited to "physical objects"; both physical objects and information about the objects need to be considered to resolve these issues. As a technology to connect physical things with information, the EPC System is expected to solve these issues effectively and achieve the safe and secure supply chain.

In this paper, we will introduce emerging supply chain issues in Japan first, analyze the nature of these issues, and then explain how the EPC system improves safety and security of the supply chain effectively. In addition, we will propose types of measures to improve safety and security using the EPC system based on the characteristics of the products and supply chains, and briefly introduce potential research topics in the EPC system.

2 Issues in Japan

2.1 Counterfeit

Fake labeling of agricultural products

Fake labeling of agricultural products became famous when the Bovine Spongiform Encephalopathy (BSE or Mad Cow Disease) epidemic hit Japan. At that time, sales of Japanese beef plunged suddenly because people were afraid of eating beef. To help the industry the Japanese government decided to subsidize companies that deal with Japanese beef. The subsidy was paid corresponding to the amount of beef that the company had and the company had to dispose of the beef if it wanted to receive the subsidy. However one company that wholesaled both Japanese beef and imported beef illicitly put fake labels on the imported beef, claiming it to be domestic beef, in order to fraudulently exploit the subsidy. The case was found by inside information, and the company was condemned for the dishonesty. Moreover, people boycotted its products, and eventually the company had to leave the market [4]. This company was not the only case for this kind of exploitation of government subsidy; there are several similar cases in which companies put fake labels to get government money illicitly [5].

There is another fake labeling case as well. In Japan, locations and species of the agricultural products become the "brand," in just the same way as the SONY brand for televisions and Toyota for automotives. Each brand has its reputations, most of which are deliciousness of the brand product. Therefore, even though they may look exactly the same, the consumer is willing to pay more money to Kobe Beef than the beef from an unknown region. There are several famous brands in most of the agricultural products, and, because of the higher margin that they can get by dealing with these famous brands, producers are working hard to keep the quality of the products and the image of the brands. However, since it is difficult to tell the difference, say, between two pieces of meat, malicious producers, wholesalers, repackagers, and/or retailers put fake labels on the packages and mislead consumers [5].

Putting aside the issue that this fake labeling is done by producers, because it is more of an ethical issue than a safe and secure supply chain issue, there are still many issues in this fake labeling: Companies in the supply chain can easily remove the labels from what they buy from their suppliers and put fake labels onto them. If there is a mechanism to stop this fake labeling, not only industrious producers of the brand products can secure their brand and get fair profits but also consumers can buy products after confirming that what they buy is what they see on the labels.

Fake Products

Fake products cause tremendous damage to both consumers and manufacturers of the products. According to JETRO (Japan External Trade Organization), a broad range of products, such as electronic home appliances, machine tools, auto-parts, office equipment, motorbikes, toys, cosmetics, and food, are being counterfeited [7]. There are two main characteristics in the Japanese fake product cases: 1) fake

products of the Japanese manufacturers' brand are made in the foreign countries and sold in both foreign and domestic market, and 2) fake products of the foreign manufacturers, such as luxury goods manufacturer, are made in the foreign countries and sold in the domestic market. Since fake products are made in foreign countries in both cases, anti-counterfeit measures taken by the Japanese government is to work on the countries in which fake products are made to settle intellectual property legislations and enforce strict regulations, and to work together with the international community to develop international laws and guidelines for these fake product issues. At the same time, the government as well as industry organizations produce public campaigns and educate consumers not to buy fake products [8].

At the same time, the private sector is also taking anti-counterfeit measures. A common approach is to use physical anti-counterfeit measures, such as holograms and invisible ultra-violet ink [9], [10]. Although they are effective to some extent, these physical anti-counterfeit measures are also copied by the counterfeiters. Since Auto-ID technology is said to be effective for anti-counterfeit, expectations of the technology are high in many industries. Recently the electronic appliance industry started a new consortium called the Home Appliance Electronic Tag Consortium. The goal of this consortium is to develop industry rules and guidelines about how to use RFID, and it is said that one of the purposes is to study how RFID can be used for anti-counterfeit [11].

2.2 Illegal trade

Gray market

If the market is different, the price of the same product, even though they are in the same condition, may not be the same. This difference mainly comes from the difference in tariffs and exchange rates. This price differentiation was not a big issue when the distribution of the products was controlled by the manufacturers and the authorized distributors. However, with the development of the Internet and the progress of the transportation network, the movement of both humans and products is becoming dynamic, and, as a result, products are sold in many markets through the channels that the original manufacturers and the authorized distribution channel companies do not intend. This kind of practice is called parallel import; and this unintended distribution has brought many consequences; such as deteriorating distributor relations, brand image, profits, sales force morale, and customer service efforts; and is becoming a major issue [12], [13], [14].

In addition to these gray market issues caused by parallel import, there is another gray market issue that is created by the discount stores that buy excess inventory from the authorized distribution channel companies with cheaper wholesaler price and sell them to consumers with discount price. Whether this practice becomes an issue or not is dependent on each industry. For example, in the home electronic appliance industry, these kinds of discount stores became one of the major distribution channels for the manufacturers; manufacturers and discount stores became interdependent. Now those discount stores realize low cost procurement by directly purchasing products from manufacturers instead of buying excess inventory from

the authorized distribution channels. It is not gray anymore. In other industries, on the other hand, manufacturers are still making efforts not to lower brand image caused by the proliferation of the unintended cheap products in the market and trying to maintain the authorized distribution channel [15].

Although it is not necessarily illegal as indicated in its name, gray market can become a barrier for the manufacturers that want to deliver safe and secure products to the end customers. In this sense, we will argue that this gray market phenomenon is a kind of safe and secure supply chain issue in this paper.

Black Market

Although national television networks sometimes broadcast news about criminal organizations that steal luxury cars and smuggle them to foreign market or sell them as used cars [16], [17], the black market, in which stolen and illegal products are traded, is not an immediate threat to ordinary people in Japan. However, it is not something that we can ignore completely because it is said that shoplifted products are sometimes sold in the secondary market, which can be categorized as a black market issue. All kinds of goods sold in stores can become a target for shoplifting, but, among them, books are a major target for shoplifters. According to the statistics from the Ministry of Economy, Trade and Industry, the average damage of shoplifting is about 2.1 million Japanese yen per book store. Considering the fact that the number of the book stores in Japan is about 80,000, the total damage of shoplifting adds up to 150 billion Japanese yen [18]. Since selling the stolen books is not a major reason for the shoplifting, it is not likely that this huge amount of books floods into the legitimate market. However, since the secondary market for books has been well developed in Japan, the possibility of the stolen books' re-entering the market is high.

2.3 Wrong status

Infection caused by contaminated blood or blood-origin drugs

Transmission of HIV and other blood-borne viruses through blood transfusions or the use of human origin drugs became a huge social problem in Japan, too [19]. To respond to the situation, the industry and the government have been tightening the quality check of the raw materials of the drugs, which are human and animal blood or organs, in order to stop manufacturing contaminated products. Also they have developed after infection action rules and regulations. Those include identifying the root cause of the infection and affected patients, and treating them as soon as possible.

What makes this issue more complicated is the existence of the infectious prion diseases, such as Variant Creutzfeldt-Jakob Disease (vCJD). vCJD is said to be a disease when a human is infected with BSE. Although researchers are working hard to investigate the infectious mechanism and the treatment, since so far the number of patients is not so many, the entire process of studying the disease is not as smooth as is expected. One of the characteristics known so far is the long latency period that the disease has. The exact period is not clear, but it is said to be

as long as 20 years [20], [21]. This piece of information gives a tremendous impact to the community because the effects of long latency period may grow exponentially. When one patient who is given human or animal origin drugs is found to be infected with vCJD, the possibility of having other patients is high; those who took the same drug could be infected. However, since the latency period is so long, it is difficult to envision the entire impact by only using conventional record retention system. The long latency period also becomes a cause of secondary and tertiary infections, assuming that the patients may donate their blood and organs.

To respond to the current and future problem, the Ministry of Health, Labor and Welfare issued a new regulation enacted on July 31, 2003. In the new regulation, manufacturers of human origin drugs have to retain information of their customers, which are pharmacies and hospitals, for thirty years, and pharmacies and hospitals that used the drugs have to retain information about their patients for ten years. Manufacturers also have to retain customer information for ten years if the drug is of animal origin [22], [22].

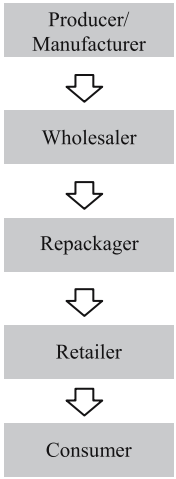
The existing concept of drug recall is to collect defective drugs from the drug market. This process is not perfect but effective considering the nature of the current drug problem. However, this recall process may not work well for the contaminated drugs that can cause diseases like vCJD, whose infectious mechanism is not totally known and latency period is very long. Other measures such as new government regulations as well as new technologies such as Auto-ID technology will be required to effectively minimize the impact of this kind of diseases.

3 Nature of the Issues

3.1 Analysis of threats of safe and secure supply chain

This subsection explores the nature of the safe and secure supply chain issues and how the EPC System can work effectively on the nature of the issues. Figure 1 depicts a supply chain from producer/manufacturer to consumer. This figure is showing the entities in the supply chain and the general flow of products but is not limiting the relations of each entity. For example, a producer/manufacturer may ship directly to its consumers in some cases, and a wholesaler may sell products to another wholesaler in other cases. In addition, there might be a case in which a person may buy products from a retailer and sell them through an internet auction. At the time the person and the retailer work as a retailer and a wholesaler, respectively.

Each entity in the supply chain has its own vulnerabilities in terms of safety and security of the supply chain (Table 1). Although it is not necessarily illegal, we include gray market issues as parts of safe and secure supply chain issues because 1) these issues cause both manufacturers and consumers problems and 2) we believe that supply chain members should be notified of the risk of the products before they actually sell them to their customers.



Producer/Manufacturer sources raw materials and makes products. They also label their products (physically, logically).

Wholesaler sources products from producer/manufacturer and sells them to repackager and retailer. Sometimes, they pack the products (e.g., from large quantity package to small quantity package).

Repackager sources products from wholesaler (or producer/manufacturer), repacks without processing the products, and sells them to retailer (or wholesaler).

Retailer sources products from producer/manufacturer, wholesaler, or repackager, and sells them to consumer. They sometimes repack the products (e.g., small quantity packages)

Consumer buys products from retailers.

Fig. 1 Supply chain

3.2 Analysis of securing supply chain applications

3.2.1 Basic applications

In order to realize safe and secure supply chains, companies have to do two types of verifications, physical verification and informational verification, with three applications. Firstly, each entity of the supply chain must verify the authenticity and

Table 1 Threats of supply chain security and entry points of them

		Producer/ Manufacturer	Wholesaler	Repackager	Retailer	Consumer
Counterfeit	Fake label	X		X		
	Adulteration	X		X		
	Re-label		X		X	
	Substitute		X		X	
	Fake product		X	X	X	
Illegal trade	Stolen		X	X	X	
	Gray market		X	X	X	
Wrong status	Scrapped		X		X	
	Recall/Contamination	X		X		

Table 2 Basic applications for securing supply chain

Application	Verification type	Description
On-site status verification	Physical	Verify authenticity by checking the ingredient of the product (component elements, DNA etc.), verify packages for broken packaging, broken seals etc.
	Informational	Verify status by checking information about the products, such as serial number, certificate, trade history, etc
Track	Informational	Verify status by checking prepositioning information
Trace	Informational	Identify suspicious products, eliminate the products, and stop proliferation of the products

appropriateness of the products (hereafter "status verification") on hand in order not to catch problematic products (On-site status verification). This practice includes both physical verification, such as checking if the packaging is broken, and information verification, such as the serial number verification. It is also important for consignees to verify the products they will receive before they physically receive the products by getting prepositioning information about the products from shippers, such as serial number and certificates (Track). Secondly, each entity of the supply chain as well as regulatory bodies must identify the entry point of the problematic products, assess the impact of the case, eliminate the products, and expose the parties that introduced the problematic products. This can be implemented by accumulating the track information (Trace).

3.2.2 Analysis of intentional mislabeling by Manufacturer/Producer

In the case that the producer/manufacturer labels their products (either physically or logically) and that the products are shipped directly to their customers, if the producer/manufacturer ships low quality products or substandard products, saying that they are appropriate, it is impossible to detect the problem by solely using informational status verification. For example, if a producer of Kobe beef falsely sells normal beef as Kobe beef, downstream members of the supply chain can not recognize the issue, and other physical status verification measures, such as DNA testing, may be required. Considering the difficulty of using physical status verification measures at the individual consumer level, government and/or industry bodies need to implement effective measures. For example, government agencies or organizations from the public sector sample the product and check the status, or issue the licenses to the authorized producers/manufacturers if it meets a certain qualification; or trusted third parties certify the authenticity of both producers/manufacturers and their products so that consumers can have information to judge the quality of the products. As analyzed different nature of the fake labeling done by producer/manufacturer, we put these issues out of scope of this study.

Table 3 Threats of the supply chain security and entry points of them (Revised)

		Producer/ Manufacturer	Wholesaler	Repackager	Retailer	Consumer
Counterfeit	Fake label	(out of scope)		X		
	Adulteration	(out of scope)		X		
	Re-label		X		X	
	Substitute		X		X	
	Fake product		X	X	X	
Illegal trade	Stolen		X	X	X	
	Gray market		X	X	X	
Wrong status	Scrapped		X	X	X	
	Recall/ Contamination	X		X		

3.3 Measures to enable suggested applications

3.3.1 Mass serialization

In order to verify the legitimacy of the product status, the ability to uniquely specify the individual products with unique identifiers or serial numbers is essential. Moreover, to have a unique identifier itself can show the authenticity of the product to some extent. The problem of the unique identifier will be discussed in the next subsection.

3.3.2 Verification by using information

Importance and constraints of physical status verification
In the previous subsection, we explained that to specify the individual product is crucial for product status verification and that mass serialization is necessary for this purpose. One verification method is a physical status verification, with which those who check the product status confirm the physical characteristics of the products, such as breaks of the shrink wrap and clarity of the hologram. However, it is not enough because these physical characteristics can also be copied by parties such as criminal organizations. Moreover, to check these physical characteristics may delay the speed of the supply chain, which can undermine the benefit of using RFID. Therefore, informational verification, especially using automated data capturing, is required.

Status verification by using information

Introduction of information to verify the legitimacy of the product status can improve the verification speed and accuracy of verification, but to introduce a new concept, “information” in this case, will add a new vulnerability because this “information” could be wrong. Therefore, it is necessary to verify the status of the “information” as well.

There are several pieces of information that can be used for product verification. One fundamental one is the unique identifier. However, if the unique identifier is scanned by anyone, it is easily duplicated. In the case that the numbering scheme is too simple, identifiers may be forged by malicious parties. In order to improve security levels regarding this unique identifier, it is necessary to employ some kind of access control and encryption technologies when identifiers are read. In addition, to manage the lifecycle of unique identifiers and to verify the legitimacy of the identifiers including detecting duplication will be necessary to improve the security level. Related information of the products is also used for status verification. It includes information such as certificates and trade history of the product. Just the same as unique identifiers, this information also needs to have its status verified its status before it is used for product status verification.

In addition, since all or part of the information used to verify the product status is supposed to be managed by using network resources, the availability of the network at each check point in the supply chain is also important. Types of network availability are 1) always available (on-line), 2) sometimes available (on/off-line), and 3) not available (off-line). Security levels will become lower according to this order. This is because supply chain members need to use real time information when they verify the information legitimacy but they can not do so without the network. In addition, it is necessary to guarantee the authenticity of the information that links unique identifiers to network resources and information.

Track and trace enabled by the related information

We proposed three applications for product status verification in the previous subsection: on-site status verification, track, and trace. These three applications, especially track and trace, require information to uniquely identify the product (unique identifiers), and to describe the attributes of the product (related information). Regarding tracking, originally it is used to streamline the shipping process, and it also improves the accuracy of the shipment. In this case, unique identifiers will be sent from a shipper to a consignee by using communication systems such as EDI. The consignee uses this information to confirm whether the product is what it ordered, prepare for receipt of the shipment, and to trigger permission of the new orders from its customers. This is a short description of tracking applications for shipment, but tracking for product verification can be implemented easily by adding a few more processes, such as verification of the unique identifiers and certificate/trade history documents, to this tracking for shipment processes.

Trace, on the other hand, is an application using the product trail archive. Therefore, once you use unique identifiers and product related information for tracking, you can implement trace by just archiving the same information. There are several ways to implement trace, and they are cetegorized based on the location of the

archive information. This difference also affects the speed of the retrieval and difficulty of the implementation. These are three different types of trace:

- store infiormation about track in one place and retrieve trace information (Centralized),
- store information about track in each of the supply chain member and retrieve trace information by visiting each member with the unique identifier as a key (De-centralized), and
- circulate trace information with the product and retrieve trace information by checking locally stored information (Pedigree).

Each measure has its own advantages and disadvantages and is chosen based on the requirements from the product and the industry.

As explained here, track and race applications to verify the status of the products, can be implemented by using unique identifiers and related information that describes attributes of the products.

3.3.3 Tampering and re-labeling alert

In the previous subsection, we explained that it is essential to check the appearance of the products, such as tampering and re-labeling, in terms of product status verification and that this check has a down side for slowing down the speed of the supply chain if it is done through human intervention. In order to resolve this disadvantage, it will be useful to detect package tampering automatically, alert and convert the change into information signals. This automatic detection can be done with RFID tags. One example of this kind of tags is the type used for international

Table 4 Measures to secure supply chain

		Measures covered by the EPC System			Measures not covered
		Mass serialization	Related information	Tamper & Relabel alert	Physical verification
Counterfeit	Fake label	X	X		
	Adulteration	X	X		
	Re-label			X	X
	Substitute			X	X
	Fake product	X	X	X	X
Illegal trade	Stolen	X	X		
	Gray market	X	X		
Wrong status	Scrapped	X	X	X	X
	Recall/ Contamination	X	X		

shipments, such as e-seals. E-seals can emit signals when they detect tampering of the container [23][24].

3.4 Mapping measures to supply chain issues

Based on the argument in this section, we map security measures (3.3) onto potential threats in the supply chain (2). Table 4 shows the relation.

3.5 Functions of the EPC System necessary for safe and secure supply chain

In the previous section, we explained that there are three effective measures in securing supply chain. They are:

1) to introduce unique identifiers for confirmation of individual products (Mass serialization),
2) to confirm the legitimacy of the product status by using information including both unique identifiers and related information such as certificate and trade history (Related information), and
3) to detect breach of the products and/or product packaging by automatically alerting and converting physical changes into information (Tampering and relabeling alert).

Then what kind of functions are required for the EPC System to realize these measures? Table 5 shows the relation between proposed measures and components that constitute the EPC System.

Thiss map is also used to make decisions when a company has to prioritize security measures for its products. For example, if a product has characteristics that

Table 5 Mapping of security measures to EPC System components

Security measures		EPC	Tag	Reader	Middle ware	EPC-IS	ONS
Mass serialization	Encryption	X	X	X	X		X
	Access control		X	X	X		
	ID management					X	X
Related information	Electronic document verification					X	
	Product status management					X	X
Tamper and Relabel alert	Tamper proof tag	X	X	X			

the number of shipments is many but the price is not high, then naturally the manufacturer will take low functionality/low price tags with high functionality network system because the more it ships its products, the higher the total cost of the system becomes. This is just one example for the product characteristics, but characteristics of the supply chain, including network availability, also affect the choice of the security measures. In the next chapter, we will qualitatively analyze what kind of measures companies should take in order to make their supply chain safe and secure.

4 Choice of security measures

There are several measures to secure supply chain, and these measures are different in their effect, cost, and difficulty of implementation. Each measure has its own advantages and disadvantages, and companies choose measures considering the characteristics of the product and its supply chain to make the effect feasible and sufficient. In addition, the availability of the network is also important when companies select security measures as explained in the previous chapter. In this chapter, we will analyze the relation between these characteristics and security measures.

4.1 Relation between characteristics of product and security measures

Products have many characteristics that will affect the choice of security measures. They are:

- products sold with outer package (e.g., cosmetics, drugs) and without outer package (e.g., auto parts, second hand distribution of luxury goods),
- products with different monetary value, and
- products that have privacy issues with non line of sight read (e.g., drugs).

Packaging of the product is one important factor when companies select security measures. When products are usually shipped without outer packages, the company has to consider the possibility of the tag detachment. Rather than sticking a tag on the product, it may be effective to embed a tag into the product or directly engrave/inscribe the necessary information onto the product so that the binding between the product and the unique identifier and related information can be guaranteed. The impact here is that if the company chooses a solution that uses engraved two-dimensional barcodes as a carrier of unique identifiers, it can not use high functionality that can be provided by RFID tags and has to rely on the network system for high functionalities such as tag data encryption.

Next is the monetary value of th product. When comparing high and low functionality, it is more expensive to use the high functionality tags than the low functionality ones. Although it depends on the number of products to be shipped and the level of the required security, relatively expensive products may justify the use of high

functionality (expensive) tags, whereas low functionality (cheap) tags will be used for relatively cheap products. If you need to meet a security level for the supply chain system, more sophisticated functionality is required for the network system of the low functionality tag solution than that of the high functionality tag solution.

Privacy and security issueswill become important if the information stored in the tag can be easily read without any access control. Suppose a manufacturer makes controlled drugs like morphine and it adopts to use low functionality RFID tags that do not have access control functionality, anyone can get the tag information by just scanning the tags, and the possibility of the theft by criminal organizations will become higher. In such a case, the company should pay more attention to the security and should choose high functionality tags that have access control functionality or at least remove product information from the unique identifier so that companies that do not have business contracts with the manufacturer cannot identify the product type.

4.2 Relation between characteristics of supply chain and security measures

Supply chains also have many characteristics that will affect the choice of security measures. They are:

- simple distribution (e.g., fresh produce)
- complex distribution (e.g., medical equipment, parallel import)
- secondary distribution (e.g., internet auction)

In the simple distributions, since originally the number of participants is small and the distribution route is rather static, the possibility of having adverse events will be small. Therefore, relatively low security solutions will be sufficient for this kind of distribution. Whereas, in the complex distributions, since not only the total length and time of the product movement will be long but also the route becomes dynamic, the chances of counterfeited product entry and parallel import product entry will become higher. In the distribution that has these characteristics, the assumption that the network connectivity is always available may not be true all the time. Therefore, solutions with high tag functionality and low network dependency will be feasible.

In the secondary distribution, such as Interne auction, it is also expected that products are routed dynamically. But if ordinary consumers are assumed to work as either wholesalers or retailers, high functionality tag solution may not be appropriate. This is because high functionality tag solution is effective in securing the information stored in the tag with technologies like access control, but the solution requires complex network system including expensive high functionality readers. As a result, consumers can not read the information stored in the tag, update the status of the unique identifiers and related information, or verify the status of the product. In such a case, a solution with which a unique identifier can be read easily and the product is verified easily is required even if the security level might be sacrificed to some extent.

4.3. Relation between network dependency and security measures

We touched a little on the importance of the network availability in analyzing characteristics of the supply chain. In this subsection, we will show the relation between network availability and possible security measures thoroughly.

Off-line means that measures do not require network access. For example, to verify whether a tag can be read with a legitimate identifier that follow the identifier's schematic rule and to detect whether the product package is broken or not can be a basis of entire security measures. On/off-line means that the system has network access but not always available. In this case, measures that use network access can be used, but the security level is not as high as that of on-line. The reason why the security level is considered to be low is because the network is used to exchange access control key and verify against a list of illegitimate (or legitimate) identifiers, but real time information is not always available to check the legitimacy of identifiers, including identifier duplications. Lastly, on-line means that network, and consequently necessary information, is always available and higher security can be achieved.

As shown in the table, the choice of the security measures depends on availability of the network. For example, if a high functionality solution with tag encryption is adapted in a supply chain without network availability, the solution will work so long as the encryption is not broken, but once it is broken, it can not change security keys through network and therefore, the security level becomes extremely low. A similar problem is the security level of the network, since all the measures that use the network access assume the security of the network itself, the security level of the measure becomes lower if the network is not secure enough because secret information could be exposed to unknown third parties. Therefore, if companies in the supply chain can not implement sufficient network security

Table 6 Network availability and security measures

		Network availability		
		Off line	On & Off	On line
Mass serialization	Schematic	X	X	X
	Encryption		X	X
	Access control		X	X
	ID management (cached)		X	X
	ID management (real time)			X
Related information	Electronic document verification		X	X
	Product status management (cached)		X	X
	Product status management (real time)			X
Tamper and Relabel alert	Tamper proof tag	X	X	X

because of some constraints, the entire security level of the supply chain will be lower.

5 Potential research topics

The effectiveness of the EPC System to realize the safe and secure supply chain was explained in section 3, but the EPC components that have been standardized or are being standardized may not be sufficient to realize all the measures proposed in the chapter. That is, studies regarding the EPC System must be done to realize the proposal. Moreover, these studies are not enough; business rules and guidelines, such as document formats to exchange related information, will be necessary in order to implement the security measures. In this chapter, we will introduce potential research topics in both EPC System components and outside of EPC System components.

5.1 Mass serialization

ID Encryption
When companies verify the product status by using information, to keep the legitimacy of the unique identifiers is crucial. If it is read by anyone and duplicated easily, the entire security measures proposed in chapter 3 are undermined. Even if a malicious third party gets the identifier from the tag, but if it is encrypted and the party can not decrypt and get the real unique identifier that is used to connect the identifier with the object, companies can reduce the risk of counterfeiting. Encryption alone does not solve identifier duplication, but at least obscuring identifier schema will prevent malicious parties from generating schematically legitimate identifiers. Therefore, with the help of identifier duplication check, which is a function of the identifier management, companies can improve security level of supply chains.

The mechanism is also effective for some of the privacy issues. One of the privacy concerns is that anyone who has an RFID reader may be able to get the information of the product you have through casual contact. But if the identifier is encrypted and does not have any meaning by itself, that will solve this casual contact privacy issue.

Access control
The same problem described in the ID encryption part can be solved by implementing access control to the tag. By access control, tag access is only allowed to the supply chain members and/or the authorized parties. Through this access control, not everyone can get the information stored in the tag, and the possibility of duplication and forging by malicious third parties will become lower.

Moreover, just the same as ID encryption, this access control function is also effective in dealing with privacy concerns. One of the strong solutions to protect

privacy in the current RFID implementation is the function to kill the tag (i.e. to render it permanently inoperable), e.g. at the checkout, when the products are bought by consumers. But if the access of the tag is properly controlled, consumers do not need to kill the tag and they may be able to utilize tag information in maintaining the products, collecting the quality information, and recycling the product.

Regarding methods of access control, access control technologies developed for the Internet security and smart card security can be applied to this RFID security.

Identifier management
ID encryption and access control are the measures to prevent unique identifiers from being exposed unnecessarily, but this identifier management is a measure to deal with after exposure or in the case of no access limitation to tags. Even if identifiers are not protected with encryption or access control, furthermore if there is a mechanism to detect duplicates and wrong status (e.g., identifiers of stolen products are found or identifiers of used and scrapped products are found), it will improve safety and security of the supply chain because the possibilities of suspicious activities or system errors are high in those cases. The mechanism will require network services to manage the lifecycle of the identifier as well as requiring new functionality from the EPC System components.

5.2 Related information to secure supply chain

Electronic document validation
It is useful to use product related information, such as certificate and trade history to secure the supply chain. However, this information could also be counterfeited or duplicated; therefore, measures to prevent these malicious activities need to be identified and eliminated. Unfortunately current EPC System does not have functionalities to exchange this kind of related information; therefore, it is necessary either to standardize this information within the EPC System or to develop industry standards by analyzing business processes to secure the supply chain. One example of this related information is the electronic pedigree document discussed in the Healthcare and Life Science Business Action Group (HLS BAG) at EPCglobal [27].

Product status management
Just the same as the argument in 5.1, even after shipment, products in the wrong status should be eliminated from the supply chain. Wrong status includes product recall and expiry. In order to realize this function, information about the products need to be exchanged. This exchange of the related information can be done by using EPCIS, a component in the EPC System, but business processes that are out side of the EPC System standards scope also need to be agreed among the companies in the supply chain.

In addition, trade history information is not always queried and updated through the network considering the availability of the network. One of the current

EPC System's design policies is to limit functions of tags in order to make them cheap and to control related information using the network, such as the internet [28]. But even without network connectivity, companies may have to verify the status of the products, and, in such a case, some amount of information might be stored in the tags for redundancy purpose. With the new air protocol, user memory areas became available, but guidelines of how to use the user memory, including memory allocation and fine-grained access controls, should be necessary to use the storage capacity in these tags [29].

ONS Security
As analyzed in 3.3, to guarantee the authenticity of the information that links the unique identifiers with network resources is necessary. In the EPC System, the only resolver function currently provided is the Object Naming Service (ONS), which does not have the mechanism to authenticate both client and network resource information. If the supply chains are static and members of the supply chain have settled business deals, this resolver function might not be required; however, if the transaction is dynamic, securing the resolver mechanism becomes mandatory. It is planned that 'EPC Discovery Services' will provide serial-level lookup services for tagged objects throughout their lifecycle. Massive scalability, secure access controls, authentication and the ability to prevent unauthorized data mining of the serial-level tracking data are some of the business requirements for Discovery Services.

5.3 Tampering and re-labeling alert

Tamper-evident tag
In case a unique identifier of the product is attached not directly to the product but the package of the product, managing the information about the product is not enough to guarantee the safety and security of the product. This is because the package may be opened and the product inside is switched for other products. Of course this package tampering may be detected through manual checks, but, since it will delay the supply chain process, some measures to detect package tampering and send it to the system managing the supply chain is required. With this function, package tampering will be detected without spoiling the merits of the RFID system.

6 Conclusion

In this study, we explored issues that threaten supply chains in Japan, analyzing the nature of the issues, and showed the potentials of using the EPC System as a solution to the issues. In arguing how the EPC System is used to secure the chain, we proposed three potential applications to make supply chain safe and secure and explained how each application is effective in addressing supply chain vulnerabil-

ities. Moreover, we showed the relation between characteristics of the supply chain factors, such as products, distribution, and network availability and the recommended measures. We also introduced potential research topics to realize the proposed applications.

Although we analyzed both supply chain issues and the EPC System as a solution, more careful and quantitative analysis is required to apply the EPC System to an actual issue because each case must have different supply chain members, damage structure, and urgency; and these factors are interdependent. One of the important factors is the security level that the system needs to realize. Of course, it is better if members of the chain can achieve the level, but, even if they can not, they may lower the possibility of having adverse events to some extent.

This section started from the introduction of the supply chain issues in Japan, but, since these issues can be applied to other parts of the globe, we believe all the arguments made in this paper will be useful when analyzing the safe and secure supply chain issues throughout the world.

References

1 Nikkei Shimbun: e-retail special – IT as enabler of safe and secure life, Food traceability is available at store front, Nikkei Ryutsu Shimbun MJ, September 21, 2005 (in Japanese)

2 Nikkei Shimbun: e-retail special – RFID is coming! Attendees are counted by using RFID embedded ticket at EXPO, Nikkei Ryutsu Shimbun MJ, September 21, 2005 (in Japanese)

3 Nikkei Shimbun: SANRIO called for help to combat counterfeit products to customs offices, Nikkei Sangyo Shimbun, August 29, 2005 (in Japanese)

4 Nikkei Shimbun: Snow Brand Food decided liquidation, Nikkei Shimbun, April 27, 2002 (in Japanese)

5 Nikkei Shimbun: Reasons of stumble (1) Trust toward food manufacturers, Nikkei Shimbun, August 13, 2002 (in Japanese)

6 Sankei Shimbun: Japanese government asks China for strict law enforcement to IPR abuse, Sankei Shimbun, December 4, 2001 (in Japanese)

7 Intellectual Property Rights Department JETRO Beijing. Available from: http://www.jetro-pkip.org/ (in Japanese)

8 Japan Ministry of Economy, Trade and Industry: Issues and measures for counterfeit products, October 16, 2002. Available from: http://www.kantei.go.jp/jp/singi/titeki/dai7/7siryou4.pdf (in Japanese)

9 SECUTAG. Available from: http://www.secutag.com/

10 Hewlett-Packard: UV/IR Invisible Ink System. Available from: http://www.hp.com/oeminkjet/products/C6121A/overview.html

11 Nikkei Shimbun: Home appliance industry set up a consortium to study RFID – Four companies including SONY will develop industry guidelines, Nikkei Shimbun, November 30, 2005 (in Japanese)

12 Chaudhry, P.E., Walsh, M.G.: Gray marketing of pharmaceuticals. Journal of Health Care Marketing, Vol. 15, No.3 (1995) 18–22

13 Myers, M.B.: Incidents of gray market activity among U.S. exporters: occurrence, characteristics, and consequences. J. of Int. Business Studies, Vol. 30, No. 1 (1999) 105–128
14 Cavusgil, S.T., Sikora, E.: How multinationals can counter gray market imports. Columbia J. of World Business Winter (1988) 75–85
15 Tsuchida, K.: A study of discrimination of sales in E-Business and limitation of re-selling. Waseda Hougakkai Waseda Hogaku, Vol 76, No. 3 (2001) 209–238 (in Japanese)
16 Nikkei Shimbun: Arrest organized crime of automotive. Damage is JPN 1 billion and 46 are arrested, Nikkei Shimbun, October 10, 2005 (in Japanese)
17 Nikkei Shimbun: Worst record of automotive theft. 64,000 cases a year, Nikkei Shimbun, January 20, 2004 (in Japanese)
18 Japan Ministry of Economy, Trade and Industry: Survey for shoplifting at bookstore, October 25, 2002 (in Japanese)
19 Defense council for Osaka HIV law suit: International Conference for AIDS caused by drug contamination, Sairyu-sha, Tokyo (1998) (in Japanese)
20 BBC: vCJD may take 30 years to show, BBC News, March 22, 2001. Available from: http://news.bbc.co.uk/1/hi/health/1235241.stm
21 Japan Ministry of Health, Welfare and Labor: (Draft) Regulation for human and animal origin drugs record retention, January 10, 2003 (in Japanese)
22 Japan Ministry of Health, Labor, and Welfare: Pharmaceutical Affairs Law, §68.9 (2002) (in Japanese)
23 Hadow, R.: E-seals and RFID. The Journal of Commerce (2004) 58
24 Tirschwell, P.: Container seals: here they come. The Journal of Commerce (2005) 52
25 Yamada, Y.: Marking technology trend and point of utilizing marking technology, Tool Engineer, September (2005) 84–85 (in Japanese)
26 Brandner, S.: Privacy as an after thought. In: NetworkWorld, March 1 (2004) 24
27 EPCglobal, Healthcare Life Science Business Action Group (HLS BAG). Available from: http://www.epcglobalinc.org/action_groups/hls_bag.html
28 Sarma, S., Brock, D., Ashton, K.: The Networked Physical World, Auto-ID Center. In: Auto-ID Labs White Paper. Available from: http://www.autoidlabs.org/whitepapers/mit-autoid-wh-001.pdf
29 EPCglobal: EPC Radio-Frequency Identity Protocols Class-1 Generation-2 UHF RFID Protocol for Communications at 860 MHz–960 MHz Version 1.0.9 (2005). Available from: http://www.epcglobalinc.org/standards_technology/EPCglobal2UHFRFIDProtocolV109122005.pdf (01.2005)

Chapter 11

The Potential of RFID and NFC in Anti-Counterfeiting

Mikko Lehtonen[1], Thorsten Staake[2], Florian Michahelles[1], and Elgar Fleisch[1,2]

[1]Information Management, ETH Zurich, 8092 Zurich, Switzerland, {mlehtonen,fmichahelles}@ethz.ch
[2]Institute of Technology Management, University of St.Gallen, 9000 St.Gallen, Switzerland, {thorsten.staake,elgar.fleisch}@unisg.ch

Abstract. In this paper, we investigate how RFID and NFC could improve current customs processes to fight illicit trade. During the import process, customs officers have to evaluate which consignments are inspected and, when an inspection takes place, whether intellectual property rights have been infringed. We propose and evaluate new micro processes that leverage the dual-existence of products and logistic units in order to enable easier, faster and more reliable inspection of goods. The impact of the improved processes is analyzed with an example case.

Keywords: Anti-counterfeiting, customs, NFC, product authentication, RFID

1 Introduction

Counterfeit and pirate goods cause increasing economic losses to companies, industries and countries and threaten the consumer health and safety while fostering other illegal activities [9]. Though the actual value of counterfeit trade is unknown, it is estimated up to EUR 500 billion annually and is escalating rapidly [20]. Taking into consideration the high growth rate in the world container port traffic – having had average annual increase of 11.8% between 2000 and 2004 [26, 25], compared to 2% average annual increase in the world GDP during the corresponding period [27] – we can conclude that an increasing number of containers with an increasing number of counterfeit and pirate goods flows to affected markets every year.

The majority of counterfeit products in the Western countries are imports and the primary sources of counterfeits are in Asia, China alone representing the source of more than 50% of fakes stopped at the European borders [4]. The most

important means of transport of counterfeit products is sea, being responsible of 70% of stopped fake products in Europe [4]. Modern technology and high level of industrialization have made it possible for illicit manufacturers to produce fake products in high volumes and to imitate the original grade and quality [9].

Fortunately, development in technology has also enabled novel countermeasures. Most notably, the emerging of radio-frequency identification (RFID) technology has opened many opportunities to fight illicit trade [22],[8]. RFID allows for automatic identification of tagged objects and establishes a link between the physical world and a virtual world. This so called dual existence of objects enables a number of new ways to manage the physical world, and thus RFID is being used for example to automate many supply chain processes [12].

Combined with Near Field Communication (NFC), technology for short-range wireless connectivity between hand-held devices and tags [23], RFID could be effectively used to authenticate tagged products. NFC uses the 13.56 MHz Industrial, Scientific and Medical (ISM) radio band that is globally available. The devices apply touch to read principle which makes communication easy and intuitive, and the typical reading ranges vary from 0 to 20 cm [30]. Besides reading NFC tags, the protocol allows for secure two-way communication between the reader devices. This differentiates NFC from RFID technology used in supply chain applications, where the goal is mostly to read multiple tags at once without line of sight. NFC devices are expected to become widespread when integrated in mobile phones, and Nokia unveiled first such product in 2004 [19]. It should be noted that since NFC is based on inductive coupling, also it can be considered as RFID technology. Therefore in the remainder of this paper we include NFC in the term RFID if not explicitly stated otherwise.

Even though RFID is used more and more to label logistic units as they flow through the supply chain, it is not yet realistic to assume that majority of individual items would be tagged. Therefore RFID-based product authentication should take place inside the supply chain, while the products travel in tagged units such as pallets and cases. As a consequence, RFID has substantial potential to improve customs countermeasures against illicit trade. Customs is very important authority in protecting societies against counterfeit products and in many cases the only gatekeeper between the manufacturer of counterfeit goods and the end customers. Due to the rapidly increasing workload and obligations concerning other interests than finding fake products, customs resources in anti-counterfeiting are very limited. Therefore increasing the efficiency of customs processes for finding counterfeit goods is important for the success in the fight against illicit trade.

In this paper, we propose ways how RFID and NFC could be used to improve existing customs processes to fight illicit trade. Illicit trade is a roof term for trade with goods that infringe Intellectual Property Rights (IPRs) and for a number of grey- and black market activities. Infringing goods can be counterfeit or pirate goods, depending on whether they infringe a trademark or a copyright (or related right), respectively. We will use the term counterfeit in the rest of this paper to cover both counterfeit and pirate goods. Grey market activities are legal in the eyes of law and thus customs has no right (or interest) to intervene in these cases, but they can break contracts and thus be considered illicit by the IPR owners.

This paper is organized as follows. Section 2 introduces the customs role in anti-counterfeiting, how RFID is used in customs today and the import process in the European Union. In section 3 we propose ways to improve the current process by making use of RFID and NFC technologies. Section 4 analyzes the impact of the proposed improvements and we finish with conclusions.

2 Customs

Customs is a critical institution for protecting the interests of a society and its citizens. It manages the physical movement of goods and people across borders and is responsible of collecting customs duties at import. National customs administrations cooperate with industries and with each other and their work is governed by the World Customs Organization (WCO). The role of WCO is to increase the efficiency and effectiveness of customs administrations and, in order to achieve this goal it provides guidelines [14] for modern customs principles.

2.1 Customs and Counterfeiting

Customs are responsible for about 70% of all seizures of counterfeit products in the world [2]. The role of customs is especially important in protecting the European Union and the U.S. because the vast majority of counterfeit products in those markets are imports [4] and, after entering the market, subject to free circulation within the community. However, in anti-counterfeiting customs role is more supportive than proactive, which means that customs mostly provide help to trademark owners to protect their IPRs when this is requested.

Customs authorities fail to seize large amounts of counterfeits either because they do not know how to recognize the fakes or because the process of gathering statements from trademark owners is too time-consuming. Proper labelling, overt anti-counterfeiting technologies and training in recognizing counterfeits would, for example, assist officials in enforcing the IPRs of affected trademark owners. Furthermore, the lack of information sharing is often perceived to be one of the main obstacles in the fight against counterfeiters [9].

Finding counterfeit products is one part of customs responsibility to control the trade. Other control objectives include finding various kinds of dangerous materials. While controlling the trade, however, customs also work to facilitate the trade not to disturb import and export. These two objectives conflict and thus customs always have to balance between control and facilitation. Given also that the vast majority of goods that pass through customs are legal and should not be disturbed, finding counterfeit goods is not among customs top priorities.

2.2 RFID in Customs Today

Customs are using RFID in many ways today. In the busy Shenzhen customs in Hong-Kong, RFID has been used to speed and facilitate the flow of low-risk traffic since 2002 [11]. The Shenzhen customs use passive RFID tags to identify vehicles and their drivers in the traffic lanes. Vehicle ID, driver ID and weight of the consignment are compared to the information that is sent in pre-hand into the customs department's computer system. If any discrepancies occur, the consignment is subject to closer inspection.

Customs use RFID also to strengthen the security of consignments. To guarantee the integrity of cargo, shippers install electronic seals, or *e-seals*, into their containers. The e-seal consists of an active tag and a mechanism that can detect if a container's door has been opened without authorization, so that it can communicate whether the container's integrity has been guarded or not. E-seals are used to secure cargo arriving to the U.S. from many foreign ports [10, 14]. The concept of an *e-container* was set forth in [29] as a risk mitigation technology against the risks associated with global container transport. The suggested e-container uses real-time monitoring of a container's physical status acquired from an array of embedded RFID-enabled sensors. It is shown that by selecting a suitable set of sensors (e.g. those measuring ambient temperature, light, air humidity, radioactivity etc.), a number of container transport threats can be mitigated, such as smuggling of drugs, weapons or humans and theft during inspections. The role of RFID in the e-container is to provide connectivity and real-time telemetry.

Also tracking of tagged containers increases cargo security, and it has been used for example in automated border program of the U.S. customs [10]. During the recent years, the movement of international cargo has become more and more regulated. Though these regulations do not demand the use of RFID, they oblige companies to provide more accurate and timely data on shipments [15]. One consequence of this trend is the emerging of *green lane* programs where shipping companies gain lighter inspections when they conform to certain additional regulations, such as in the Smart & Secure Tradelanes (SST) initiative [13] or the Customs-Trade Partnership Against Terrorism (C-TPAT) [28]. In addition to the control of the flow of goods, RFID chips in electronic passports allow for biometric identification of travellers [17].

2.3 The Import Process

Due to limited resources and size of the workload, it is impossible to search every consignment entering the country; in practice, only about 1–4 percent of imported goods are inspected. Therefore customs success in anti-counterfeiting depends on how the scarce resources are allocated, which in practice means deciding which consignments are inspected. Customs conduct *risk analysis* to identify high-risk consignments in pre-hand [5]. The risks are estimated by combining the likelihood and consequence of an event [5] and the analysis is based on information in the *freight papers* that are used in processing the flow of imported and exported

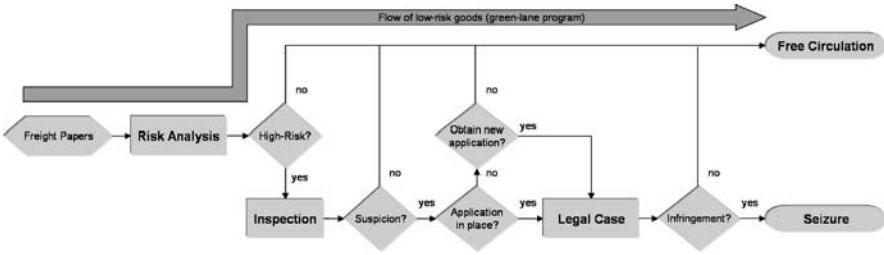


Fig. 1 Simplified diagram of the European customs import control process for seizing counterfeit goods

goods. Freight papers like the *air waybill* specify, for example, the shipping route and the cargo manifest of a consignment and they can exist both in physical and electronic form. Though the risk-analysis can be partially automated, interviews with customs officers reveal that the experience of the officers plays a very important role in recognizing suspecting consignments. Regarding counterfeiting, the country of origin is the most important criteria in the risk-analysis and, consequently, often attempted to be disguised by the carriers of counterfeit goods [9]. Careful selection of inspected containers can provably provide considerable improvements in the detection rates of counterfeit products [3].

Our focus is on how counterfeit goods are seized in the European customs import process, illustrated in Figure 1. The legal aspects of this process are defined by the European Council regulation 1383/2003 [1]. To protect their IPRs, the right holders have to lodge an *application for action* where they provide the customs with the information to authenticate original products. The application for action is an important form of cooperation between customs and industry and the customs rarely seek for counterfeit goods without an application in place.

3 The Improved Processes

In this section, we describe new micro processes that can be used to improve the existing customs import process to find and seize more counterfeit goods. The enabling technology of the proposed processes is any hand-held NFC device with a network connection, such as already available NFC mobile phone [19]. This device allows the customs officers to read tagged items in the field work. We take into account that in a modern customs process, the flow of information and the flow of goods are separated and therefore the customs officers need to move to the warehouse to conduct the physical inspections. In a very lean and automated import process, the time that the products spend in the customs warehouse can be very small and measured in tens of minutes, which can set rigid time-constraints for the inspections.

3.1 Facilitated Manual Authentication

Inserting RFID tags on logistic units allows for novel ways to obtain information for manual authentication[1] of goods during the inspection. This facilitates the inspection by making product authentication faster and more reliable by providing the inspectors with accurate and timely data. The process steps are following:

1. Obtain the product-class ID number (i.e. two first fields of standard structure company *prefix.item reference.serial number*) of the product under inspection. Three possible ways are distinguished:
   - When the product is tagged, read the product tag to obtain the product-class ID.
   - Read the consignment (e.g., pallet or case) tag that contains aggregated data structure of product ID numbers (e.g., electronic freight papers) (see Figure 2).
   - Query a database that links product classes with their ID numbers.
2. Find the network address of an authorized server for the product class using a network address resolution mechanism (e.g., Object Naming Service [6]).
3. Establish a secure connection with the authorized server (e.g. EPC Information Service [7]).
4. Download the information that supports manual authentication.

Customs officers would benefit from the described process by having an automated way to obtain information for authentication of tagged, but also non-tagged products. This information could contain descriptions of features and pictures of the original products as well as descriptions of common counterfeit features. In practice, a product-class ID database could be gathered from applications for action (subsection 2.3) to assist authentication of goods that are not tagged.

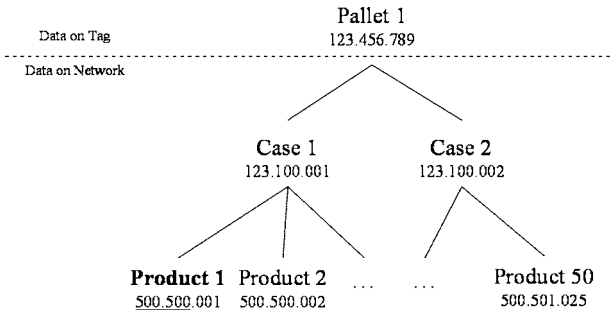


Fig. 2 An example of how the aggregated data structure of a logistic unit can be used to obtain the product-class ID number (500.500) of one specific product (Product 1). The pallet ID number acts as a link to data on network.

[1] In this work, authentication is defined as answering whether something original or not

This micro process relies on right-holders upkeeping the authentication data on online servers and letting the customs to access it. Having the authentication data on the right-holder's server guarantees the timeliness of the data and allows it to be updated and published to all stakeholders without delays. In addition, the technology allows the customs officers to take pictures of counterfeit consignments and upload them, and other data, to right-holder's server to share information of actions of counterfeit players.

3.2 Automated Authentication

Although RFID is primarily an identification technology, it also supports for many ways of authentication. The starting point of RFID based product authentication is to insert a tag on the product and to authenticate the tag or use it to collect data for authentication. Even low-cost tags can be used for product authentication (e.g. track and trace based plausibility check [22]). Higher levels of security can be achieved by using more expensive tags that support cryptographic operations. A review on RFID product authentication techniques can be found from [18]. Currently there are no standardized RFID product authentication protocols and, furthermore, multiple protocols are needed to balance the trade-off between cost and security of the solution. For these reasons, it is necessary to establish which authentication protocol, if any, the inspected product (tag) supports, before the authentication can take place. The process steps are following

1. Identify the product by reading the tag.
2. Obtain the network address of the authentication server using a network address resolution mechanism (e.g., Object Naming Service [6]).
3. Establish a secure connection with the authorized server (e.g. EPC PAS [22]).
4. Establish which authentication protocol, if any, the tag supports.
5. Automatically authenticate the product (tag) using the supported protocol.
6. Verify the tag-product integrity.

The presented process makes the authentication almost completely automatic. Support for multiple authentication protocols does not pose specific hardware requirements because authentication protocols can be made transparent for the interrogator. In particular, the authentication can be processed on the back-end server so that the reader device only needs to be informed about the result. It should be kept in mind that usually it is actually the tag that is authenticated and not the product itself. Therefore verification is required to make sure that the authenticated identity really matches the physical product (step 6). Omitting this verification makes the system vulnerable to simple attacks where fake goods are equipped with any authentic tags.

For cases where no network connection is available (e.g. inside a container or a vessel), an offline authentication protocol is required. Feasible solutions for offline authentication, however, have not yet been proposed [18]. In some cases the problem can be overcome through batch mode authentication [24] where the reader device initiates the authentication in offline mode, for example during inspection

inside the steel container, and finishes the authentication when a network connection becomes available.

3.3 Machine-Readable Freight Papers

RFID enables also automatic identification and authentication of tagged freight papers. The carrier company could insert RFID tags with consignment identifier numbers into the physical freight papers to allow for automated document handling process at customs. Making use of the dual existence, the tag could provide a link to electronic freight papers or to an aggregated consignment data structure illustrated in Figure 2, for example, to enable customs officers to access all data of a consignment by scanning the freight papers. The main motivation to use RFID with freight papers is to have the same infrastructure to identify containers, pallets, cases, and freight papers. Even though RFID is neither the only nor always the optimal technology to make documents machine-readable [16], we believe that benefits of having only one infrastructure would outweigh the privacy- and cost-related shortcomings. Furthermore, the RFID tag could be used to strengthen the security of the freight papers [32].

4 Discussion

In this section, we analyze the impact of the proposed micro processes on the existing customs import process and on illicit trade. First, the use of RFID and NFC supports the modern customs principles given by WCO [21], which include continuous development of control techniques and maximum use of information technology. Second, RFID enables new means of communication between the right-holder of goods and customs. Currently customs receive the information of the content of a merchandised consignment only from the carrier company's freight papers. Tagging cargo would enable the customs officer to access the manufacturer's online server that maintains data about the tagged products, which could make the customs process less dependent on the data provided by the carrier company. Most notably, this link could provide means to inform the right-holder about grey market activities such as parallel trade. However, this could need changes in legislation because currently customs do not have legal basis to interfere in these kinds of activities.

Our micro process concepts show how RFID and NFC can facilitate information sharing and help customs officers to recognize fakes, given that right-holders tag their consignments. This helps customs to seize bigger amounts of counterfeit products, which the again provides economic benefits for the affected right-holders. In addition, RFID brings value also in other supply chain applications and so the cost of tagging the cargo doesn't need to be justified only by the increased supply chain security.

Interoperability between different kinds of tags, however, remains a technical obstacle to be solved. NFC tags operate at 15.56 MHz HF band, while RFID tags common in supply chain applications, such as EPC Class1 Gen2, use the 860–960 MHz UHF band. These tags cannot be read with the same device. As a result, manufacturers might have to consider using multiple tags for different applications. However, it is not impossible to produce an integrated tag that would support both these standards, and first such tag is planned to be released during 2007 [31].

The proposed processes contribute primarily to faster, easier and more reliable authentication of tagged consignments, while the product-class ID number database (subsection 3.1) would also facilitate manual authentication of non-tagged products. We have shown how also the case or pallet level tags, together with an aggregated data structure of the logistic unit, can be used to facilitate the authentication of single products inside the unit. It is important to note that RFID won't remove the need for human oversight; at minimum the customs officers have to verify that the tags are attached to right products.

Finally, we argue that RFID enabled authentication does not considerably increase the number of counterfeit products customs find if only a small ratio of cargo supports it. Furthermore, we can assume that counterfeit products are normally not tagged. As a consequence, the underlying problem will be how technology that can authenticate some of the original products can be used to distinguish more counterfeit ones. This problem is illustrated with the following example.

4.1 Example

Let's consider the import process described in subsection 2.3 and denote the percentage of imported goods that are counterfeits by P_1. When customs officers inspect a consignment, the probability to find counterfeit goods is $C \cdot P_1$, where C is an internal performance factor. Assuming that customs ability to select suspicious shipments (risk analysis) cancels out the chance of not detecting counterfeit goods even though they have been inspected we can estimate that $C=1$, though there is no substantial data to support this estimation. We can now formalize that by inspecting y percent of imported cargo, customs find $y \cdot C \cdot P_1$ percent of imported counterfeits.

Let's further assume that x percent of imported consignments are tagged and thus support for easier authentication (subsection 3.1 and 3.2). Assuming that counterfeit consignments are never tagged, customs can direct the tagged ratio of imported goods into a green lane program where they are not inspected, upper path in Figure 1 (in reality, also these goods need to be inspected, but since they are tagged the inspection is considerably faster). By doing so, counterfeit goods can be searched among smaller amount of consignments and, if other factors remain unchanged, the probability to find counterfeit goods in one random inspection increases to:

$$P_2 = C \times \frac{P_1}{100 - x} \tag{1}$$

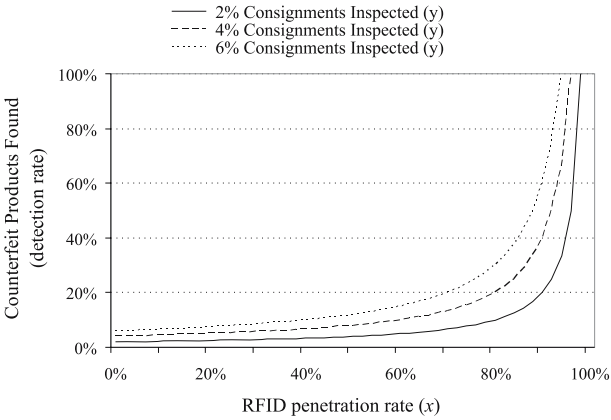


Fig. 3 An evaluation of how the inspection rate and RFID penetration rate of imported goods affect the detection rate of counterfeit products, Equation 1 (*C=1*).

As a consequence, the detection rate of counterfeits increases to $y \cdot P_2$. This is demonstrated in Figure 3 which illustrates the effect of RFID penetration rate *x* to the seizure rate.

4.2 Analysis

The example illustrates how increasing RFID penetration rate in the imported cargo helps customs to find increasing number of fakes with constant amount of inspections. In the optimistic case, the lack of RFID tags in a consignment would act as an indicator of counterfeit origin. In practice, however, also tagged cargo needs to be inspected because also counterfeit goods can be tagged. Therefore the automated authentication process (subsection .2) becomes important. Ultimately, this could lead to a situation where the lack of tags indicates counterfeit origin while the forged tags can be easily found.

Furthermore, the presented detection rate model does not consider all effects of the technology. In addition to the mentioned dependency between RFID penetration rate and counterfeit detection rate, the adoption of RFID in product inspections also decreases the time and effort needed to inspect a consignment, increasing the inspection rate *y* (given a constant time and effort for inspections), as well as increases the chances of detecting counterfeits in an inspection, increasing the internal performance factor *C*. Both these effects contribute to a higher counterfeit detection rate.

5 Conclusions

Though RFID is already used in customs logistics in different ways today, it still has unused potential to help customs in the fight against illicit trade. In this paper we have presented how, together with NFC enabled mobile reader devices, RFID enables product authentication applications that make inspection of tagged cargo faster and more reliable. The benefits of RFID and NFC in anti-counterfeiting are discussed and the impact of the improved import process is analyzed. Overall, we have shown how RFID and NFC enable new technological countermeasures for customs and intellectual property right holders that contribute towards safer supply chains.

References

1 European Commission: Regulation (EC) no. 1383/2003. Official Journal of 2 March (2003) L 196, page 7
2 European Commission: Counterfeiting & piracy: Frequently asked questions. MEMO/05/364, Brussels, 11 October (2005)
3 European Commission: International customs operation ”FAKE”. IP/05/1383. Brussels, 8 November (2005)
4 European Commission: Community-wide counterfeit statistics for 2004. Available from: http://ec.europa.eu/taxation_customs/customs/customs_controls/counterfeit_piracy/statistics/index_en.htm (26.3.2007)
5 European Commission: Standardized framework for risk management in the customs administrations of the EU (2006)
6 EPCglobal: Object naming service (ONS) specification version 1.0. EPCglobal public document, October (2005)
7 EPCglobal: EPCglobal architecture framework version 1.0. EPCglobal public document, July (2005)
8 U.S. Food and Drug Administration: Combating counterfeit drugs – a report of the food and drug administration. February (2004)
9 Organization for Economic Co-operation and Development: The economic impact of counterfeiting (1998)
10 RFID Journal: E-Seals smooth border crossings. News Article, September 3 (2002) Available from: http://www.rfidjournal.com/
11 RFID Journal: RFID speeds border crossings. News Article, October 15, 2002. Available from: http://www.rfidjournal.com/
12 RFID Journal: Wal-Mart draws line in the sand. News Article, June 11, 2003. Available from: http://www.rfidjournal.com/
13 RFID Journal: Interest grows for tagging cargo. News Article, February 2 (2005). Available from: http://www.rfidjournal.com/
14 RFID Journal: Colombian shipper to use RFID. News Article, May 15 (2006). Available from: http://www.rfidjournal.com/
15 RFID Journal: Coping with regulations. Perspective Article (2006). Available from: http://www.rfidjournal.com/
16 RFID Journal: DHS subcommittee advises against RFID. News Article, May 22 (2006). Available from: http://www.rfidjournal.com/

17 Juels, A., Molnar, D., Wagner, D.: Security and Privacy Issues in E-passports. In Conference on Security and Privacy for Emerging Areas in Communication Networks – SecureComm, Athens, Greece (2005)
18 Lehtonen, M., Staake, T., Michahelles, F., Fleisch, E.: From identification to authentication – a review of RFID product authentication techniques. Printed handout of Workshop on RFID Security – RFIDSec 06, July (2006)
19 Nokia: Nokia unveils the world's first NFC product – Nokia NFC shell for Nokia 3220 phone (2004). Available from: http://press.nokia.com (29.6.2006)
20 International Chamber of Commerce: The fight against piracy and counterfeiting of intellectual property. Policy Statement. Submitted to the 35th ICC World Congress, Marrakesh, 7 June (2004)
21 World Customs Organization: The Kyoto convention: Customs contributing to the development of international trade. Fact Sheet (2006)
22 Staake, T., Thiesse, F., Fleisch, E.: Extending the EPC network – the potential of RFID in anti-counterfeiting. In 2005 ACM symposium on Applied computing, pages 1607–1612, New York (NY), ACM Press (2005)
23 NFC Forum, http://www.nfc-forum.org/home (2006)
24 Tsudik, G.: YA-TRAP: Yet another trivial RFID authentication protocol. In International Conference on Pervasive Computing and Communications – PerCom, March 2006, Pisa, Italy. IEEE, IEEE Computer Society Press (2006)
25 United Nations: The Review of Maritime Transport, 2003. United Nations Publication UNCTAD/RMT/2003 (2003). ISBN 92-1-112582-0
26 United Nations: The Review of Maritime Transport, 2005. United Nations Publication UNCTAD/RMT/2005 (2005). ISBN 92-1-112674-6
27 World Trade Organization: International Trade Statistics (2004)
28 U.S. Customs and Border Protection: Securing the Global Supply Chain: Customs-Trade Partnership Against Terrorism (C-TPAT) Strategic Plan (2004). Available from: http://www.customs.gov/linkhandler/cgov/import/commercial_enforcement/ctpat/ctpat_strategicplan.ctt/ctpat_strategicplan.pdf (29.8.2006)
29 Schlesinger, A.: Mitigating Container Security Using Real-Time Monitoring with Active Radio Frequency Identification and Sensors. Master's Thesis, Massachusetts Institute of Technology, June (2005)
30 Ecma International: Near Field Communication. White Paper Ecma/TC32-TG19/2004/1 (2004)
31 RFID Journal: TwinLinx Proposes to Marry NFC and EPC. News Article, Dec. 19 (2006). Available from: http://www.rfidjournal.com/
32 Lehtonen, M., Staake, T., Michahelles, F., Fleisch, E.: Strengthening the Security of Machine Readable Documents by Combining RFID and Optical Memory Devices. In Ambient Intelligence Developments Conference (AmI.d), September (2006)

Chapter 12

Improving the Safety and Security of the Pharmaceutical Supply Chain

Mark Harrison[1], and Tatsuya Inaba[2]

[1] Auto-ID Lab, University of Cambridge, Cambridge, UK. mark.harrison@cantab.net
[2] Auto-ID Lab, University of Keio, Japan. tinaba@autoid.sfc.keio.ac.jp

Abstract. This paper discusses various techniques that may be used to combat counterfeiting in the pharmaceutical supply chain. These include the use of electronic pedigrees (to ensure the integrity of the supply chain), together with mass-serialization (to provide for a unique lifecycle history of each individual package) and authentication of the product (to check for any discrepancies in the various attributes of the product and its packaging are as intended for that individual package). Management of the pedigree process and product authentication is discussed in some detail, together with various other learnings from the Drug Security Network, including identification of some remaining vulnerabilities and suggestions for tightening these loopholes.

Keywords: Supply chain, Electronic pedigree.

1 Introduction

The Drug Security Network (DSN) was formed as a forum for a number of major players in the pharmaceutical industry to consider the major changes and challenges to business practices which will result from the enforcement of pedigree legislation [1] and introduction of mass-serialization, which are being introduced imminently in order to make the pharmaceutical supply chain safer and more secure.

The DSN was led by Cap Gemini and SupplyScape Corporation, with participation from GSK, Roche, Amerisource Bergen and members of Auto-ID Labs at MIT and Cambridge (UK), together with technical contributions from Hewlett-Packard and Verisign.

Unlike other initiatives such as Jumpstart, (led by Accenture), the focus of the DSN activity was not on creating or supporting an industrial field trial – but rather in developing pro-active thought leadership on three major issues – pedigree, serialization and data sharing and security.

The approach taken was to define, identify and prioritize supply chain use cases, using storyboarding, scripts and activity diagrams, to consider not only the processes which are required or impacted in meeting forthcoming regulations, but to go beyond that and consider what additional measures could be introduced to achieve a more safe and secure supply chain, then finally, consider other drivers which could add business value, both in terms of greater efficiency or protection of brand, product integrity and reputation.

Following an initial plenary kick-off meeting in December 2004, the members of the Drug Security Network met for three 2-day face-to-face meetings in January, March and May of 2005, using the Cap Gemini Accelerated Solutions Environment (ASE) to facilitate a large amount of clear thinking within each meeting. Furthermore, a practical DSN laboratory was set up at the Boston offices of Cap Gemini, to demonstrate an end-to-end practical example of how electronic pedigree could be managed between a manufacturer, distributor, pharmacy and returns processing company.

The motivation of DSN was to undertake focussed brainstorming among major players in the pharmaceutical market, identify a number of the open issues which either need to achieve consensus or require further research, and to publish the output of the activity, also contributing it as input to regulatory bodies such as the U.S. Food and Drug Administration (FDA) and the U.S. Drug Enforcement Administration (DEA) and to standards development processes at EPCglobal and elsewhere. The primary deliverables of the DSN activities consist of three papers:

1. The first paper [2] is entitled 'Serialization Options for Tracking of Pharmaceuticals using Radio-Frequency Identification', authored by Dr. Mark Harrison of Auto-ID Labs at Cambridge, UK.
2. A second paper [3] is entitled 'Technical Issues of Electronic Pedigree Inter-organizational Transactions', authored by Dr. Tatsuya Inaba, formerly of Auto-ID Labs at MIT, now with Auto-ID Labs at Keio University in Japan. This is summarized in Sections. 5–6 of this paper.
3. The third paper [4] provided an overview of the DSN activities and a summary of the two other papers, as well as a discussion of many of the remaining open issues to be addressed.

2 Electronic Pedigree

The purpose of a pedigree is to provide legal proof of a secure chain of custody from the originator of the pharmaceutical package (usually the manufacturer or wholesaler) through to the organization that sells or dispenses the pharmaceuticals. Three key issues need to be considered:

- Pedigree Data Content/Format
- Pedigree Processing
- Pedigree Transmission Mechanism

2.1 Data Content Format

A number of key requirements can be identified for a standardized format for electronic pedigrees:

- **Completeness**: It is in the best interests of everyone that a 'highest common multiple' pedigree format emerges, with the complete superset of information required by the various US state and federal legislation, as well as any other information which may be required by traceability legislation in other countries, such as Belgium [5] and Italy [6, 7], where traceability initiatives have already begun.
- **Global Scope**: It is also very important to maintain a global perspective rather than being US-centric. For example, rather than have a data field for the US National Drug Code (NDC), have one field for product code and another field for product code type, such that in the USA, the product code type may be set to 'NDC' – while it may be set to other values in other regions of the world, such as the AIC code for the drug, issued by the Italian Medicines Agency (AIFA).
- **Suitability for legal or government audit**: The scope of the information present in the pedigree format should be carefully considered, since it will be a legal document. Information that is not required by legislation nor essential to the implementation of pedigree security should be contained in a separate information document or wrapper – but not in the individual pedigree document format. Government agencies may require that pedigree information systems and pedigree management applications should be audited, to ensure the security of the information and to ensure that it is not possible to falsify, alter or delete the information which constitutes the legal pedigree document. In particular, it is very important that when the pedigree is stored in electronic format, that adequate provisions are made for data backup and recovery and that records which form part of a legal document cannot be modified or deleted within the legal lifespan of that document.
- **Integrity, Authentication and Non-Repudiation**: Digital signatures [8] provide document integrity, authentication and non-repudiation. The authentication checks that the information has not been altered from that which was signed and that the signer signed the information. The signed content must include the original hash and a reference to the public key of the signer. This allows each transaction to be electronically authenticated by the recipient's system.

With some input and refinement from the Drug Security Network members, SupplyScape Corporation have proposed an Open Universal Pedigree Interchange Format [9]. This was subsequently contributed as an input to the EPCglobal Healthcare and Life Sciences Pedigree Task Force, together with other contributions of schema from Cyclone Commerce, Raining Data and Verisign. The result was a blend of all four contributions. In terms of data content, it provides not only a superset of what is required by state laws in Florida [10], California, etc. and by the National Association of Boards of Pharmacy, but also additional product information fields such as Item ID, Pedigree ID and Parent Pedigree ID and

transaction information such as transaction type (sale/transfer/return), license state and other digital signature information (key information, signature information, meaning associated with signature, timestamp of signature). Furthermore, an Advance Pedigree Notice (APN) was proposed as a wrapper or envelope for transmitting a collection of pedigrees. The APN can also contain additional business data to be shared with the trading partner, while keeping the business information segregated, so that it is neither mixed with regulatory data in the pedigree documents nor propagated further down the supply chain beyond the specific trading partner for whom it was intended. The APN therefore consists of three elements:

1. Order / Trading partner information
2. Shared Business Data
3. Pedigree Information (a collection of one or more pedigree documents)

2.2 Pedigree Processing

Processing of electronic pedigrees for pharmaceutical packages requires the following three steps as a minimum requirement:

1. Authentication of the pedigree, including verifying transactions for all previous custodians before the product arrives
2. When receiving product, verification that the incoming product matches the authenticated pedigree
3. When shipping product, sign the outgoing pedigree and transmit to the next custodian before shipping the product.

Figure 1 illustrates the various stages of pedigree processing for both receiving and shipping processes. It also indicates the responsibilities for manufacturers,

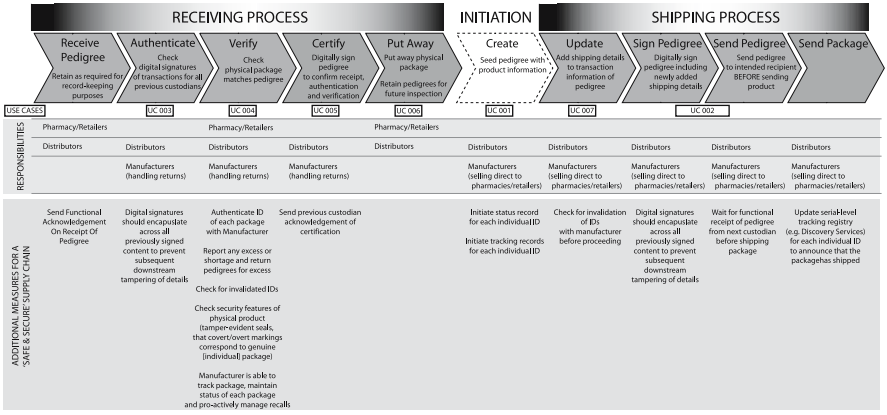


Fig. 1 Stages of pedigree processing, roles and responsibilities and additional measures for a safe and secure supply chain.

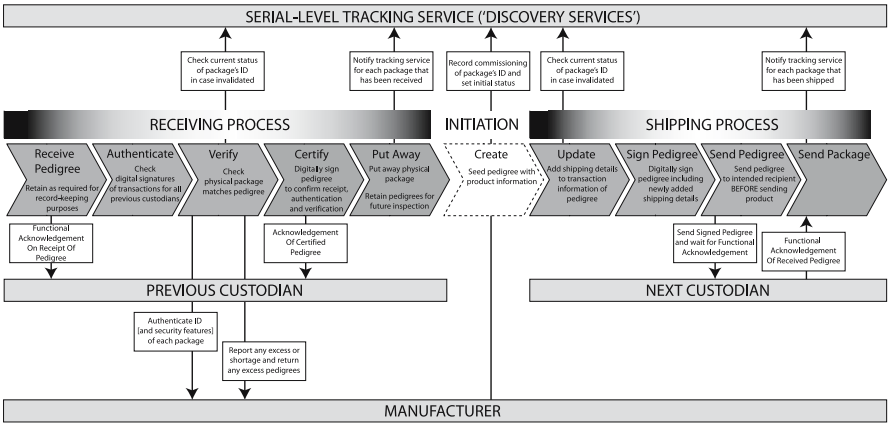


Fig. 2 Stages of pedigree processing with enhancements to improve safety and security of the supply chain.

distributors and retailers/pharmacies. The figure also indicates which of the DSN use cases represents each processing stage.

Figure 2 illustrates the stages of pedigree processing when additional measures are implemented to move closer towards a safe and secure supply chain, including various acknowledgement messages and potentially also updating of a serial-level tracking service such as the EPC Discovery Services in future. The acknowledgements and message choreography is discussed in much greater detail in [3] and Section 4 of this paper provides a summary of the paper.

It must be remembered that a pedigree is a document of record, which is subject to record-keeping, record retention and record availability requirements. Furthermore, electronic systems for managing pedigree documents are subject to regulatory requirements to provide computer systems security and control in order to protect against tampering with computers or electronic records.

It is optional whether manufacturers create and provide a pedigree to their customers, unless the manufacturer is selling direct to a retail store, in which case it needs to provide a pedigree. In 2005, the legislation in the USA did not require pharmacies and hospitals to authenticate received pedigrees, although they were required to retain them.

2.3 Pedigree Transmission Mechanism

A number of key requirements can be identified for the transmission mechanism for electronic pedigrees:

– **Timely access to data for verification and certification processes**: It is essential for the efficient operation of business that verification of all previous custodians and transactions can take place rapidly, without significant network delays or outages.

- **Robust access to data for verification and certification processes**: It is essential for the legal audit, that the verified trace of all previous custodians and transactions can be completely retrieved, whenever required, whether from a locally stored copy or from a distributed system of information services.
- **Authentication, Integrity and Non-Repudiation**: The Pedigree format or Pedigree access mechanism should provide for the highest technically achievable degree of security to ensure that each successive custodian can authenticate the pedigree and verify the trace of previous custodians and transactions, as well as appending and certifying the pedigree information when they in turn propagate the pedigree with goods shipped downstream.
- **Suitability for legal/government audit**: This may have an impact on the decision about how closely to integrate the pedigree into more general-purpose software, such as legacy EDI applications or the EPC Network [11, 12] components (specifically EPC Information Services), since doing so may result in these entire systems falling within the scope of government auditors.

There are two principal mechanisms by which pedigree information may be transmitted forwards down the supply chain and by which it may be subsequently retrieved. In the propagating document approach, the pedigree data is contained within a document, which is appended, re-signed and forwarded by each successive party in the supply chain. In the fragmented data approach, the pedigree data is stored separately by each party in their own information systems or those of a third-party provider, rather than being propagated down the supply chain. The relative merits of the two approaches are discussed below.

2.3.1 Propagating Document Approach

In this approach, each subsequent custodian verifies the signed content of previous custodians, then amends and re-signs the data, before transmitting the pedigree to the next custodian when the goods are shipped onwards. As the pedigree document moves across the supply chain, additional outer layers are added. As a consequence, the length of a propagating pedigree document can rapidly grow from a few kb to around 1 Mb per consumer-level package if each 'layer' of the pedigree document includes a full digital signature. This process is illustrated in Figure 3.

This approach offers a double-linked chain of security, since each custodian can verify all the inner layers of the pedigree document, then signs to confirm that they have done so (the reverse link). At the time of shipping, they then add additional data about the next recipient and sign this (the forward link). The double-linked chain is intended to ensure that each successive custodian is the one whom the previous custodian intended as the next recipient of the package.

A further major security benefit of the propagating document approach is that as soon as the goods pass further down the supply chain, the despatching party no longer has complete control over all copies of the data, since all subsequent receiving parties will also obtain copies of the data. If the despatching party fails to produce the required data when requested to do so, there are other copies of the data in circulation further down the supply chain. As well as providing some additional

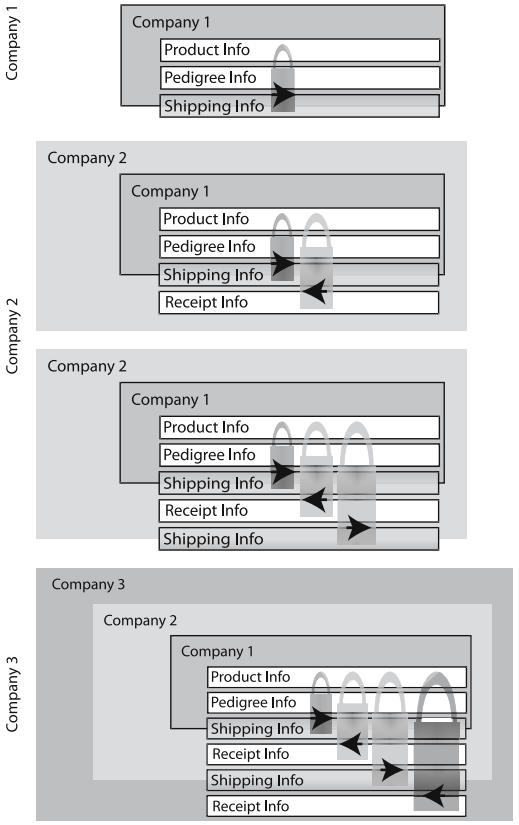


Fig. 3 A simple fragmented data approach to linking of pedigree data. Each company sends the next custodian a link to the pedigree data they hold for the product but retain the data themselves rather than embedding into a pedigree document.

robustness against accidental deletion, this approach also provides some protection against deliberate falsification of the records after the event, since a discrepancy with the data held by downstream recipients will be apparent upon investigation.

XML markup [13] is a standard method of communicating structured data in a way that is both human-readable and machine-readable and can be readily reformatted (e.g. using technologies such as XSLT) into alternative formats. The methodology of constructing digital signatures over parts of XML documents is already standardized by W3C [8].

2.3.1.2 Fragmented data approach

In this approach, the pedigree information is not sent forward from one custodian to the next. Instead, each company hosts its own electronic pedigree records on a networked database or information service, which is secured but to which trading partners and regulatory agencies are granted appropriate access.

Subsequent custodians are merely sent a hyperlink to the information, rather than being sent the data itself. This is shown schematically in Figure 4.

The obvious advantage is that much smaller amounts of data are being transmitted across the network, since the hyperlink is typically much smaller than the amount of data that it represents.

A potential disadvantage of this approach is that the receiver will need to contact each of the previous custodians independently in order to authenticate the package. This may actually result in an increased burden on the internet and local network and may halt the authentication stage if any of the upstream parties is temporarily unreachable, just because the full information required for authentication has not been transmitted in a self-contained way. Indeed, this type of distributed pedigree mechanism was discussed at the EPCglobal Healthcare and Life Sciences (HLS) meeting in Chicago – but it was considered that it did not meet robustness of available information because of the number of remote servers which needed to be contacted for verification and the cumulative probabilities of downtime over the set of relevant database servers. The buyers felt that the potential delays involved were unacceptable.

The major vulnerability in this approach is that potentially each company retains the only authoritative copy of their data – and would technically have the ability to either delete or amend and re-sign modified data, if the company were under investigation, even though such deletion or amendment and re-signing would be unlawful. A potential solution to this vulnerability would be for a requirement that when a company registers their involvement in the pedigree chain for a particular individually serialized package, they provide not only a network address to the data but also a digital signature of the data to the next receiver (see Figure 5a) and/or to a central registry (see Figure 5b), to which they are granted only one-time write access for each individual package.

Even if the data they hold is subsequently falsified and re-signed, the new digital signature will not match the value that was transmitted to the receiver (and retained by the receiver) and/or stored earlier with the serial-level tracking registry. If the shipper sends a digital signature regarding their information to the receiver, then it is important that the receiver retains the digital signature they received in addition to any hyperlink information to the data, since this independent digital signature may be required by government inspection if subsequent falsification of data by the shipper is suspected. The retention of received digital signatures is also shown in Figure 5a, 5b.

Fig. 4. A simple fragmented data approach to linking of pedigree data. Each company sends the next custodian a link to the pedigree data they hold for the product but retain the data themselves rather than embedding into a pedigree document.

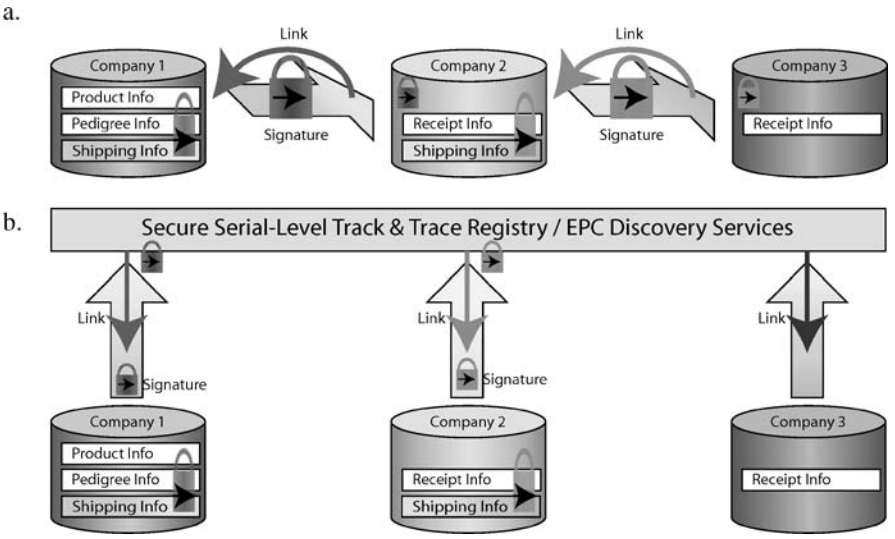


Fig. 5 A more robust mechanism for linking distributed pedigree information. In (a), the link to the pedigree information is sent from the shipper to receiver and is accompanied by a digital signature, which is retained by the receiver. In (b), the link to the pedigree information is sent to a secure serial-level track and trace registry or EPC Discovery Service, together with a digital signature, which is retained by the registry; each company is only allowed one-time write access for posting the signature.

Figure 5b introduces the concept of Discovery Services or registries holding serial-level pointers to information across the supply chain. This approach raises a number of issues regarding administration, operation and financing of such registries, all issues which need to be seriously considered by the regulators.

If a distributed approach to pedigree management were implemented, the issues of both record loss and access delay would need to be fully considered in the architectural design.

3 Authentication of Identity and of Products

To the members of the Drug Security Network, pedigree is only one aspect of the safe and secure supply chain. It provides a legal trace of the chain of custody of a product. However, as well as being able to verify the custody history of a package, an equally important aspect is the ability to track where a package is at the current time, especially in a product recall scenario. Pedigree by itself does not provide this, since there is no current requirement for information to be sent back upstream in the supply chain, towards the manufacturers – only for the pedigree information to be passed downstream. Even then, a pedigree document primarily records a chain of transactions. It does not warrant that the package itself is the genuine product. For this, authentication is required. One can think of two kinds of authentication:

– authentication of the identity, since the identity provides the 1–1 link to the pedigree data
– authentication of the product itself, in case the identity of the package has been copied or the details about the product have been falsified

The manufacturer typically holds information about which identities or serial numbers have been 'commissioned' for genuine products they have released. This might also include correlations between the package ID and the original hard-coded ID built into an RFID tag, in order to make it more difficult for counterfeiters to simply copy the package ID onto duplicate RFID tags. The manufacturers also hold data about dates of manufacture, date of expiry and other information that may also be recorded in the pedigree document or printed on the label of the package. They might also retain records of any mass-customized specialized security features that were used for a particular serial number, as well as details of what tamper-evident seals or packaging should be expected.

A key feature of the Safe and Secure Supply Chain is the emphasis on authenticating the object, as well as the pedigree trail, as shown conceptually in Figure 6.

If downstream supply chain parties authenticate the identity and product with the original manufacturers for each individual serialized package, then the manufacturers will gain much greater downstream visibility about the current locations of their products, which in turn should enable them to track them efficiently, if a recall needs to be triggered.

A networked information system, such as one complying with the future EPC Information Services (EPCIS) standard, would provide a mechanism for a manufacturer or labeller (or other authoritative party) to be able to validate a number of properties specific to a particular serial number. These might include an independent hard-coded read-only tag ID, the product class and/or details of customized security features, either covert or overt.

Clearly such information must only be provided to authenticated authorized parties, in order to prevent counterfeiters from abusing the system. In some cases, it may be practical or even preferable for the networked information system to simply respond with a Boolean (Yes/No, Pass/Fail) response to a challenge from an authenticated authorized client.

i.e. the system is allowed to respond to a query such as:

'Does this Tag ID / Product Type / Combination of security features correspond to this Object ID?'\ (Answer is Yes or No).

However, the following type of query might be forbidden to prevent counterfeiters from mining such a Product Authentication Service:

'Tell me the Tag ID / Product Type / Combination of security features for this Object ID'.

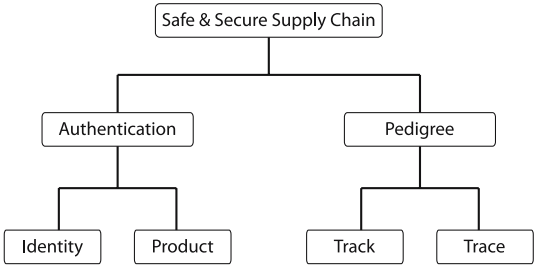


Fig. 6 DSN concept diagram to illustrate the fundamental elements of a safe and secure supply chain.

At present, one way in which this sort of Challenge / Boolean Response type query might be implemented in EPCIS is for the provider of the EPCIS service to allow access to a query whose input parameters are the Package ID or EPC and the Product Type or Combination of security features detected. An empty result set or a count of 0 events indicates no records – i.e. there is no match between the Package ID/EPC and the specified Product Type or security features – i.e. authentication failed, whereas a non-empty result set or a count of 1 event indicates successful authentication.

The pharmaceutical industry needs to consider whether this type of query approach is sufficient for object authentication purposes or whether custom types of query are needed – and whether there needs to be greater access controls regarding who is allowed to receive records of the 'commissioning event' when an EPC is first created for a package, since this event might normally hold details of attributes about the object, such as which Tag ID was used – or which combination of security features are present. i.e. when a pharmaceutical manufacturer implements an EPC Information Service, it may have its security controls configured to exclude the commissioning events from general EPCIS queries – and only provide them to a restricted group of clients and only when two or more parameters are supplied which match the commissioning event, i.e. only on a challenge-response basis.

When validating the authenticity of the product, it may be necessary to check the following criteria:

Authenticity of the tag

- Was the tag being read the same original tag which the original manufacturer or labeller applied? (i.e. do the EPC and TagID match the manufacturer's records about the association between a particular Tag ID and the corresponding EPC?)

Authenticity of the pedigree ID

- Is the number of pedigree IDs greater than the number allowed for a given lot?
- What is the structure of the pedigree ID?
- Was the pedigree ID actually issued by the manufacturer or labeller?

Authenticity of the serialized identifier

- Is the serialized identifier or EPC programmed into the tag a valid one?
- Has that particular serialized identifier or EPC been issued by the manufacturer?
- Does the serialized ID or EPC match the one specified in the Pedigree?

Authenticity of the product's packaging

- Are there security features (microprinting, holograms, watermarks, iridescent inks, UV inks)?
- Have the security features been mass-customized (i.e. not always the same combination for all products or all serial numbers within a product line)?
- Do the information services have a record of the security features to expect (and where to find them) – and those not to expect? (which if present, indicate that the packaging is suspect)
- Does the mass-customization of security features (both present and absent) agree with what is observed?

Checking the current state

- Is that particular serialized identifier or EPC still available for distribution, sale or dispensing or has it already been decommissioned, marked as sold, recalled, returned, destroyed, etc.
- Is the information record corresponding to that serialized identifier or EPC now closed?
- Is the serialized object still in circulation beyond the expiry date assigned by the manufacturer?

Authenticating the trail

- Can the pedigree trail be verified for all previous custodians?
- Has the object followed a permissible supply chain path, without irregularities? (How can irregularities be defined?)
- Where are the events signifying cross-border transportation and customs clearance?
- Is the serialized object travelling along the forward supply chain or the reverse supply chain? Is this consistent with the last recorded state and intended destination region for that object? (What are the possible states and permitted state transitions?)

4 Data Sharing and Security

This section provides a summary of the DSN paper in [3]. The paper is primarily concerned about the messages that are exchanged between businesses in order to conduct transactions, once the requirements for pedigree are in force. The choreography of messages is documented in terms of Unified Modelling Language

(UML) [14] activity diagrams, together with tables of descriptions. Functional acknowledgements, transactions, timeouts and retries are also considered.

The paper also discusses various aspects of security, transport protocols and includes an analysis of the network bandwidth requirements which will be required for processing of electronic pedigrees and uses queuing theory to estimate the waiting times and number of items in queues to be processed.

4.1 Use Cases

Three groups of use cases are considered:

Base Case

- sufficient to comply with Florida pedigree law
- implemented in DSN Lab

Safe and Secure

- goes further, contains use cases useful in realizing the vision of a safer, more secure drug supply chain
- specifically identifies the following (currently optional) steps as being characteristics of a safe and secure supply chain:
- shipment confirmation messages,
- confirmation messages of the order from the buyer,
- termination or closure of the object's identifier and the associated pedigree document

Business Value

- realizing business value for companies employing an e-Pedigree application

Use cases are considered from an inter-organizational perspective, rather than an intra-organizational perspective. The use cases documented have clearly defined goals, scope/level, preconditions, description and successful end conditions and fail end conditions. Tables list the primary actors (described by roles (buyer/seller) rather than as manufacturer/wholesaler/retailer), triggers, frequency and extensions, issues and notes.

The paper clearly distinguishes between the different impacts of electronic pedigree on retailers and wholesalers/distributors; whereas wholesalers must receive goods from their suppliers and verify the pedigree, then authenticate and certify the pedigree document, before selling the goods, a retailer is only required to receive and verify, but is not currently required to authenticate and certify the pedigree document, or to 'close' the pedigree document. This is illustrated in Figure 1 of this paper.

When mass-serialization is introduced, there will be significant changes to receiving processes:

- It will no longer be sufficient to check bulk quantities and product types against a purchase order.
- For each item, there will need to be a check for a 1–1 match of serial numbers between:
- Pedigree documents with purchase order
- Pedigree documents and received items

A further complication is that the buyer might not necessarily receive all pedigree documents at the same time, even though the buyer also needs to control the relation between pedigree documents and purchase orders. The paper considers the message choreography in terms of the following:

- Offer documents (e.g. Purchase Orders (PO), Shipping Notices)
- Acceptance documents (response to offer – need not be electronic)
- Functional acknowledgements (a message from seller to verify syntax or confirm correct transmission/format, not necessarily acceptance of deal)

The timing between messages, acceptable delays and time lapses before retries are acknowledged is an issue which must be considered and may have impacts on the design of e-pedigree application software, although the actual policies and actual values of time to retry, number of retries etc. are matters for trading partners to agree upon. Many of these parameters are already handled in existing EDI standards, such as the X12 series [15] – but pedigree management software will need to be able to be configurable with these policies, ideally in a machine-readable way. The paper also considers revocation documents used to cancel an offer document before an acceptance document is received.

4.2 Security

In the discussion on security, the paper [3] identifies five key security requirements:

Authentication

- establishes trust regarding the identity of two partners exchanging messages

Authorization

- does the other partner have appropriate authorization to send a business document / deal?

Confidentiality

- is the communication channel private? (e.g. encrypted documents / channel)

Integrity

- is it certain that the business document is not garbled or tampered?

Non-Repudiation

– receiving partner has proof of the receipt of the original business document – and the initiating partner has a proof of the receipt that the receiving partner successfully received the business document.

The paper then discusses how specific existing EDI and internet technologies can be used to cover each of these aspects of security. These are summarised in Table 1 below.

Table 1 Summary of technology solutions.

Technology solution	Security feature offered
EDI-INT AS2 + SSL + S/MIME	Partner authentication and authorization
SSL + S/MIME	Confidentiality of message exchange
Digital Signature embedded in S/MIME packet	Data integrity and non-repudiation of the original business document
Message Disposition Notification (MDN) sent back from receiver to initiator	Non-repudiation of message receipt
Public Key Infrastructure (PKI) + Digital Signatures	Guarantee authorization of the business document

4.3 Pedigree documents – Information content

The concept of a Pedigree Business Document is introduced. This serves as a wrapper or envelope to consolidate several individual pedigree documents when multiple packages are shipped together. However, the individual pedigree documents remain intact within the Pedigree Business Document, which makes it easier to send them forward when shipments are split further downstream.

The format of the individual pedigree document could either be a common format agreed by all states – or a composite of the separate pedigree formats that individual states decide to use, which contains a superset of all the information which is required, even if some of it is not required by each state.

Information in the pedigree document includes:

– Information which is unique to a particular pedigree document (ID, version of format, timestamp)
– Information which is unique to an individual product package (Drug name, Manufacturer/Distributor, Object ID, NDC, Manufacturing date, Expiry date, Dosage form, strength, container size, lot number, parent package object ID)

When the package is about to be sold or transferred to the next custodian, the following is added:

- Information about the shipper
- Transaction data (sales invoice no, date of purchase, quantity by lot number)
- Shipper information (business name, address, licence number, name, title, address of person certifying pedigree, timestamp of signature, meaning of signature etc.)

The shipper then digitally signs the pedigree.

Upon receipt, the receiver validates the digital signatures (authentication) and after matching received products with the pedigree document (verification) then signs the pedigree to confirm receipt (certification). At this point, the packages can be put away until needed.

When the receiving party is ready to dispatch the packages, they take on the role of the shipper and append the pedigree with:

- Information about the shipper
- Transaction data (sales invoice no, date of purchase, quantity by lot number)
- Shipper information (business name, address, licence number, name, title, address of person certifying pedigree, timestamp of signature, meaning of signature etc.)

Finally, they digitally sign the Pedigree documents and send the pedigree information in advance of sending the package. These processing stages are also shown schematically in Figure 1 of this paper.

The paper also considers the following pedigree-related documents or messages:

- Pedigree Document Acceptance – also considered as a type of pedigree document, with a similar structure
- Revocation document – refers to original offer document – but does not contain pedigree info.
- Functional acknowledgement – generic, refers to original document, plus status and reason for error.
- Pedigree business document – wrapper to carry multiple pedigree documents and related business info. This may be either an Advance Pedigree Notice or a Pedigree document acceptance.

The paper also includes a comparison of how e-business technologies such as AS1 [16], AS2 [17] and AS3 and ebMS [18] can handle security aspects (confidentiality, integrity, authentication, authorization, non-repudiation), functional acknowledgement, revocations, retries, payload types and synchronous vs asynchronous communication. The paper [3] also provides a comparison of the AS1/AS2/AS3 specifications used in e-business.

4.4 Other issues

The problem of managing identifiers is not overlooked; the paper identifies the need to maintain associations between Purchase Order (PO) numbers and the

number of the Advance Pedigree Notice (APN) – and between the Advance Pedigree Notice (APN) and the Advance Shipping Notice (ASN), in the case where an ASN is used. It is expected that the Advance Pedigree Notice would list the unique serialized IDs of each package. This highlights a problem when no ASN is sent; the buyer does not have advance notice of which shipment contains which order or pedigree.

A section of the paper also considers use cases for Less-Than-Truckload (LTL) (e.g. consolidated shipments of mixed cases). The use case involving third-party carriers is also considered.

The paper also discusses how wholesalers could use an Advance Shipping Notice to construct a pedigree document. Appendix F of the paper considers the following use cases:

- Normal Buyer/Seller in response to purchase order
- Vendor managed inventory
- Handling returns, handling chargebacks/proof of sales

There will be a need to not only design new documents such as Pedigree document and Pedigree business document (wrapper/envelope) – but also assess the impact of the Pedigree application on existing inter-organizational transaction standards, specifically in terms of links with the pedigree documents.

4.5 Risks of paper pedigree

Appendix G of the paper [3] discusses the use of paper-based pedigree from wholesaler to retailer, to cope with retailers who cannot receive electronic messages, authenticate electronic Pedigree documents – or read object identifiers. Although this practice is currently allowed, it carries the following risks:

- Lack of digital signature technology. As discussed in Section 6 of this paper, in comparison with handwritten signatures, digital signatures provide a much higher level of confidence that the data was not corrupted or tampered with – and that the digital signature was not forged by someone else.
- Retailers cannot authenticate all the previous digital signatures of previous trades and handovers, nor the authenticity of the paper itself. Wholesalers may need to use overt anti-counterfeit measures to the paper-based Pedigree document.
- Retailers cannot check the authenticity (trade history) of the product before the product is shipped from wholesalers. Wholesalers may not execute granular status control without confirmation messages from retailers.
- Retailers can only verify received products by counting number of items and number of paper-based Pedigree documents. Human-readable object identifiers on the items and the paper-based Pedigree documents are necessary to verify the shipment. This may be quite labour intensive.
- Retailers may need to use handwritten signatures – but these do not have robust verification mechanisms, so once Pedigree document is printed out rather than being handled electronically, the pedigree (and the associated drug package) is

not transferable (or has a potential risk of counterfeiting). This may also impact the legitimate returns process.
- Retailers do not terminate the object identifiers and associated Pedigree documents when the packages are dispensed. Even if they can terminate the pedigree, paper-based Pedigree documents may be illegally reused assuming that some of the retailers just count the number of both drug packages and papers without checking the object identifiers.

The paper identifies a number of potential loopholes of paper-based pedigree documents:

- Wholesalers can print out paper-based pedigree documents of items sold to retailers with electronic pedigree – or sold to other retailers with paper-based pedigree documents. Wholesaler can have more paper-based pedigree documents than saleable items. Then, a fraudulent wholesaler can sell counterfeit items with legitimate paper-based Pedigree documents.
- Retailers can sell paper-based Pedigree documents to wholesalers. Retailers may be able to sell the object identifiers with paper-based Pedigree documents to a wholesaler. A fraudulent wholesaler may forge paper-based Pedigree documents and sell counterfeit items saying they are returns from the retailer.

In summary, using paper-based Pedigree documents increases the risks of entry of counterfeit drugs. One of the major issues here is that those who receive paper-based Pedigree documents cannot validate the trade history of the items, which is a purpose of implementing electronic Pedigree application. This also means that once a Pedigree document is printed out, the item should be transferred in the limited area. If it is allowed to re-convert paper-based Pedigree document into electronic Pedigree document, the next buyer should be notified of the risk associated with the products.

The paper also considers conversion between different transport protocols – and the need to ensure that also errors and functional acknowledgements are correctly translated between protocols.

One must also consider the legal issue of intermediate companies – is confidentiality guaranteed? Does the intermediate company also assume liability for the transaction? Are third party logistics companies also expected to comply with regulations?

5 Vulnerabilities

There are still a number of potential loopholes in the security of the proposed pedigree legislation. Regulatory bodies such as the FDA or DEA may be advised to review the vulnerabilities identified here and consider whether further guidelines or legislation needs to be issued.

5.1 Pedigrees initiated by the wholesaler rather than the manufacturer

In the USA, it is current practice for distributors to break down bulk product from manufacturers into smaller packages for shipment to retailers. In addition to the potential risks associated with paper pedigrees, there is the need for some clarification about how repackaged products should be identified and who is responsible for them. When mass serialization is introduced, it would be inappropriate for the distributor to re-use the serialized identifier of the bulk product for each of the smaller packages broken down from it, since each needs to be uniquely identifiable. Having said that, the pedigree record needs to provide the traceability all the way back to the source, so it should at minimum record the identity of the bulk product. Ideally, a new pedigree document should be created for a new package ID, which includes and extends the pedigree of the bulk product from which it was obtained. The new package ID may also be used for lookup purposes, to find authoritative information services about the package. There may be a legal issue about whether the distributor or the original manufacturer is the authority for that package and accepts the liability that accompanies this role. It is likely that wholesalers or distributors would only accept the liability and workload of applying RFID tags only for repackaged items – and in this case, the package ID should probably indicate the wholesaler or distributor, rather than the original manufacturer Finally, if it is allowed to use the original manufacturer's labeller code for the new package ID, there must be close co-ordination between distributor and manufacturer about allocation of serial numbers, in order to ensure that the manufacturer 'commissions' that particular serial number for that particular product class and records that it is a valid serial number, i.e. one which they have allocated. Ultimately, the organization that is the authority for the package ID (in this example, the manufacturer) would also be responsible for keeping track of when the package ID is ultimately decommissioned or 'closed', e.g. on dispensing, return or invalidation.

5.2 No requirement for closure – of the pedigree record or the serialized ID

At the point of sale or dispensing, when the package reaches the end of its normal supply chain, it is advisable to require that the corresponding pedigree document should be formally 'closed' or 'terminated' in order to avoid any opportunity of genuine pedigree documents re-circulating to provide an alibi for counterfeit products being introduced into the supply chain.

By the same reasoning, it is also advisable for an authoritative record of the serialized ID to be formally 'closed' or 'terminated'. This does not mean deletion of records tied to that serialized ID – but rather that termination or closure should trigger an alert if the serialized ID is subsequently detected in the normal forward supply chain, since this may be an indication of a counterfeiter attempting to reuse discarded genuine packaging or serialized IDs read from genuine packaging to

introduce counterfeit product. Within the architecture of the EPC Network [10], appropriate places to record 'closure' or 'termination' of an individual serialized ID are either in the EPC Information Service provided by the manufacturer or labeller – or as a 'flag' or field in the appropriate 'EPC Discovery Service' for the records for that individual serialized ID.

However, it is unlikely that most retail pharmacies would ever install the necessary infrastructure and staff training for closing out each serialized ID unless there is a legislative mandate requiring them to do so.

5.3 Conversion of paper pedigrees to electronic pedigrees

Digital signatures provide a much higher degree of security than handwritten signatures, since they are much more difficult to fake. A digital signature is essentially constructed from the data to be signed by algorithmically computing a message digest or summary of the data, then encrypting this with the signer's private key. In this way, the signature is different for each block of data, whereas a handwritten signature is expected to be approximately the same for each block of data. A change to a single bit of the data results in a completely different signature. Furthermore, because the signature is encrypted using the signer's private key, it is possible for anyone to use the signer's public key to verify that only they could have signed it – i.e. it provides a high degree of non-repudiation, so long as the private key is kept confidential. Knowledge of the signer's public key does not allow a third party to reverse engineer the signer's private key (at least not on a practical timescale with computing technology available today or in the near future) – so they cannot forge the signer's digital signature over data which they falsify.

Because handwritten signatures are easily forged and are not inextricably tied to the data being signed, there is a vulnerability if a pedigree in paper format (using a handwritten signature) is ever allowed to be converted back into electronic format, because the handwritten signature offers a much lower guarantee of authenticity.

Pedigree documents in which any of the signatures is not entirely digital should not be regarded as first-class genuine electronic pedigree documents and the conversion of electronic pedigrees to paper formats should be avoided if at all possible – and ideally, pedigree legislation should dictate that paper pedigrees are not acceptable. Indeed, the pedigree legislation in California does not allow for paper pedigrees, even though the state of Florida does allow for both to co-exist. In practice, paper pedigrees are unworkable because of the high volumes of units involved at item-level and the way in which products are broken down into single unit quantities, which are then combined for distribution purposes.

However, a theoretically possible transmission of an electronic pedigree via paper is described below,

- An entirely electronic pedigree document is printed out or faxed onto paper.
- The recipient scans the document and performs optical character recognition (OCR) to regenerate the text file that was originally sent.

- The text file should be canonicalized, to ensure that no additional white spaces or line break characters have been inadvertently introduced and to eliminate any syntactic variability in how an XML document is formatted. W3C has already specified how to represent an XML document in canonical form [19].
- All previous digital signatures must be verified successfully. If any of these fail, then there may be an error in the OCR process or the canonicalization. Return to step 2.
- At this stage, the recipient is effectively in possession of an electronic pedigree document and should then sign, add the shipping information, then re-sign.

This approach is not recommended in practice for normal operations, since it is clearly very time consuming and inefficient to scan and perform optical character recognition to reconstruct the XML document. Furthermore, this approach is only applicable between immediate nearest neighbour trading partners within the supply chain. It is not possible to reconstruct a secure first-class XML electronic pedigree if a handwritten signature has been included while the pedigree was being transmitted via paper.

In practice, once pedigree legislation is fully in effect, it is much more advisable for companies to implement a fully electronic pedigree management system and also to ensure that they (and their trading partners) have reliable network connectivity between them, with sufficient redundancy (e.g. via pre-positioning and local caching of data and possibly also the use of secondary internet service providers for backup) to ensure that the pedigrees conform to the highest available security, while also ensuring that their distribution/processing operations do not experience any downtime due to network connectivity outages.

If at any stage, any of the digital signatures fails to verify – or if any exchange is accompanied only by a handwritten signature, rather than a digital signature, then the pedigree can no longer be regarded as a first-class electronic pedigree for security purposes.

The current EPCglobal standard for pedigree allows for conversion from electronic pedigree to paper, although the paper signatures would simply serve a notice that the electronic signatures had been validated prior to printing. Further validation would require considerable manual methods allowed by law (e.g. e-mails, phone calls, etc.), all of which are considerably more time-consuming and labour-intensive than automated validation of electronic signatures.

5.4 The need for certification authorities

Certification authorities such as Verisign, Thawte and TRUSTe already act as trusted third parties who issue digital certificates that vouch for the correspondence between an individual or organization and their public key. This is routinely used for electronic commerce on the internet.

For electronic pedigrees for pharmaceutical packages, the public/private key is required to belong to a named individual within the organization, rather than belonging to the organization itself. It may also be appropriate to require that the certificate should contain the individual's licence number as approved by the relevant

government agency for pharmaceuticals, e.g. US FDA. In this case, a standard web-trader digital certificate may not be acceptable – instead the government agency may require that certification authorities verify additional data such as licence number etc.

Some government agencies may even consider very close involvement in the process. Indeed, the Florida regulations on certificate authorities for self-authenticating pedigrees [20] impose many additional requirements beyond those that are usual for certificate authorities issuing certificates for ordinary e-commerce purposes.

5.5 Enforcing a change of serial ID and labeller code on repackaging

When a pharmaceutical package is broken down into smaller packages, it is essential that new serial IDs are created for each of the sub-packages, so that each is independently traceable for pedigree purposes. The new serial IDs should reflect the labeller code of the distributor, rather than the labeller code of the original manufacturer of the bulk product unless there is agreement and communication between the distributor and manufacturer about which serial IDs should be allocated, in order to ensure that the new serial IDs of the sub-packages can be correctly resolved to the appropriate information records.

5.6 Cross-border shipments and diversion

Diversion is a major issue for the retail sector, but even more so for the pharmaceutical industry, since pharmaceuticals are sometimes sold to developing countries at a discounted price compared with the prices charged to the developed world. Unfortunately, these pharmaceuticals often fail to reach those in developing countries who so desperately need medical treatment. Instead, they are often intercepted by criminals or corrupt regimes and re-exported to the developed world, for sale at the regular price, resulting in a profit for the criminals or corrupt regimes involved, at the expense of the people suffering in the developing countries as well as a financial fraud perpetrated on the pharmaceutical industry.

Effective pedigree records provide an opportunity to greatly reduce such diversion activities, provided that the pedigree includes details of cross-border shipments, including details of export licences, customs clearance, etc.

The pedigrees for pharmaceuticals that are intended for shipment to developing countries should perhaps contain information about the country or at least regions[1] of the world in which they are intended for use. This designation should be irrev-

[1] For example, the United Nations International Strategy for Disaster Reduction has classified the world into three regions, not on a geographic basis – but in terms of economic development. Seehttp://www.unisdr.org/disaster-statistics/pdf/Classification-countries-UNDP-report-2004.pdf

ocable in the sense that customs clearance officials should be actively involved in the pedigree process and required to check that re-importation is not taking place for discounted pharmaceuticals intended for a developing country or region.

If region-specific restrictions can be applied to consumer products such as DVDs primarily for the commercial purposes of market fragmentation, then surely a similar mechanism could be used for the far more worthwhile purpose of ensuring that pharmaceuticals which are intended for developing countries actually reach the people whose suffering could be alleviated by those pharmaceuticals and finally put an end to interception and diversion activities by criminals and corrupt regimes in those countries.

Acknowledgments

The authors wish to thank the reviewers from Cardinal Health and Cyclone Commerce for additional insights and recent updates, which have now been included within this paper.

References

1 The Prescription Drug Marketing Act: Report to Congress. Available from: http://www.fda.gov/oc/pdma/report2001/default.htm.
2 Harrison, M.G.: Serialization options for tacking of pharmaceuticals using radio-frequency identification. In: Drug Security Network – White Papers (2005)
3 Inaba, T.: Technical issues of electronic pedigree inter-organizational transactions. In: Drug Security Network – White Papers (2005)
4 Harrison, M.G.: The drug security network – an overview and discussion of remaining issues. In: Drug Security Network – White papers (2005)
5 Royal decree modifying the royal decree of 21 December 2001 laying down the procedures, deadlines and conditions for intervention by the obligatory insurance for health care and benefits in the cost of proprietary medicinal products. Kingdom of Belgium (2003)
6 Bollini Legislation. Law 39 – 1st March 2002 (art. 40), Republic of Italy (2002)
7 Bollini Legislation. Law 14 – 3rd February 2003, Republic of Italy (2003)
8 XML and Digital Signatures. Available from: http://www.w3.org/Signature/
9 Open Universal Electronic Pedigree Interchange Format (2005). Available from: http://www.epedigree.org.
10 Florida Prescription Drug Protection Act (2003). Available from: http://election.dos.state.fl.us/laws/03laws/ch_2003-155.pdf
11 EPCglobal Network. Available from: http://www.autoidlabs.org/whitepapers
12 EPCglobal Architecture Framework Version 1.0. Available from: http://www.epcglobalinc.org/standards_technology/specifications.html
13 XML – Extensible Markup Language: W3C – World Wide Web Consortium. Available from: http://www.w3.org/XML
14 Unified Modelling Language (UML). Available from: http://www.uml.org/
15 The Accredited Standards Committee (ASC) X12

16 Harding, T., Drummond, R., Shih, C.: MIME-based secure peer-to-peer business data interchange over the Internet. In: RFC, IETF – Internet Engineering Task Force (2002). Available from: http://www.ietf.org/rfc/rfc3335.txt
17 Moberg, D., Drummond, R.: MIME-based secure peer-to-peer business data interchange using HTTP, Applicability Statement 2 (AS2). In: RFC, IETF - Internet Engineering Task Force (2005). Available from: http://www.ietf.org/rfc/rfc4130.txt
18 ebXML Messaging Service, OASIS – Organization for the Advancement of Structured Information Standards. Available from: http://www.oasis-open.org/committees/ebxml-msg/
19 Canonical XML Version 1.0. W3C. Available from: http://www.w3.org/TR/xml-c14n
20 Certification Authority and Digitial Signatures for Self-Authenticating Pedigree. Florida Department of Health. In: Florida Administrative Weekly (2.06.2006)

Part IV
Cryptographic Solutions

Chapter 13

Product Specific Security Based on RFID Technology

Thorsten Staake[1], Zoltan Nochta[2], and Elgar Fleisch[1]

[1]Auto-ID Lab St.Gallen, University of St.Gallen and ETH Zurich, Switzerland
tstaake@ethz.ch, elgar.fleisch@unisg.ch
[2]SAP Research, CEC Karlsruhe, Germany. zoltan.nochta@sap.com

Abstract: The following contribution introduces an approach to avert removal-reapplication attacks of RIFD transponders. This is achieved by tightly coupling security tags with individual objects using object specific data.

1 Why object specific security

As outlined in chapter 2, security solutions that are based on tagging technologies have a system specific drawback: when checking an object, it is the security tag (in this context, the RFID transponder) which is authenticated and not directly the object the tag is attached to. The link between tag and object is often only provided by an adhesive bonding and may not be strong enough to deduce that the object is authentic solely based on the validity of the transponder. In theory – and also in practice if the solution is not designed properly – a tag can be removed from an original article and attached to another object, thereby compromising the security system. This deception is referred to as **removal-reapplication attack**. Systems are particularly susceptible if only the packing of an article is equipped with a security feature, as it is often the case when using holograms, micro printings, or current RFID transponders.

In contrast to most other tagging technologies, RFID can efficiently overcome this shortcoming. Even low-cost RFID tags with a rather limited amount of memory can store product-specific data – e.g. the product's very accurate weight, form-factor, surface described by spectrographic analysis – in order to ensure that the tag is really attached to the corresponding product. The approach resembles the use of personal pictures in passports that logically bind the documents to their

holders. As a result, illicit actors are detained from simply removing a tag from a legitimate product and reapplying it to a counterfeit article in a way that the fake is not detected during product validation.

2 System description

In the following, a system is proposed, termed object specific security, which provides the required binding between an RFID transponder and the object the transponder is attached to. It consists of four main components: the transponder with the object specific data, the branding machine, the data management system, and the user terminal. Figure 1 shows the main components of the architecture.

Transponders with Object Specific Data
For the proposed solution,[1] passive RFID transponders with approximately 32 to 64 Byte of storage capacity are sufficient. No cryptographic functionalities are required on the side of the tag; a transponder only has to store the object specific information. An exemplary data set, termed product validation data, is given below.

Product Validation Data := {

Unique Tag ID,

Unique Product Serial Number,

Product Specific Data,

Signature Method,

Signature Value};

Unique Tag ID: The RFID tag contains a unique number (Tag ID), which is programmed by the tag manufacturer during the production process.

Unique Product Serial Number: This number is assigned by the brand owner. It may base on the EPC or any other numbering system which facilitates unique object identification.

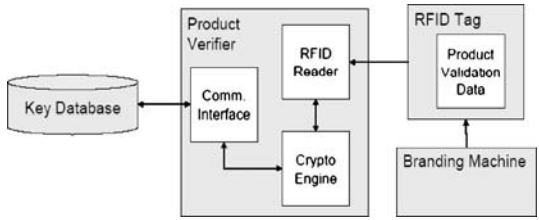


Fig. 1 Architecture overview

[1] A similar but rudimentary approach has been introduced by Nochta, Staake, and Fleisch (2004).

Product Specific Data: This information resembles the picture or fingerprint in the passport analogy. The data has to be characteristic for an individual object, stable over time, and easily measurable during inspection; it has to be unique in the way that two different instances of the same product can be distinguished by the feature the data describes. Which properties may be selected depends on the measurable specific physical, chemical, electrical, etc. characteristics of a given object and the available test environment. Examples characteristics are – either altogether or a subset thereof – weight, physical dimensions, a serial number printed on the product itself or its packaging, etc. This data will typically be written on the tag by the product's vendor before product delivery, for example during packaging. It is also possible to store a reference to the data on the tag, such as an URI that specifies an entry in a remote database. This may help to save tag resources and thus allow for the application of cheaper transponders, but will make product validation dependent on the availability of network connectivity.

Signature Method: In this section, a bit sequence identifies the combination of cryptographic methods that were used when computing the Signature Value. This information is required by the user terminal to apply the correct cryptographic algorithms during product validation.

Signature Value: The product vendor computes the Signature Value by applying a cryptographic hash function h with a asymmetric key encryption method SPr, such that

$$\text{Sig\_Value} = \text{SPr}\,(\,h\,(\,\text{Unique Tag ID, Unique Product Serial Number, Product Specific Data, Signature Method})).$$

Here, SPr indicates the usage of the vendor's private key (a.k.a. signing key) when computing the Signature Value. Note that the private key must be exclusively known to the entity (e.g. product vendor). During product validation, the corresponding public key called *Validation Key* is used to check of Signature Value.

Branding Machine

The actual binding between an object and a passive RFID tag residing on the object is performed by the component termed Branding Machine. This component is responsible for measuring the set of object specific characteristics, computing the hash value, and writing the Product Validation Data to the tag and to the manufacturer's database.

Data Management System

The data management system stores the Product Validation Data and facilitates the access management. Its functionality is similar to systems which solely check for the validity of a serial number, except that each data entry is extended by the corresponding product specific information.

User Terminal

The User Terminal (or Product Verifier) consists of a reader component, a device to measure the product specific data, a computational unit, a user interface, and a means of establishing network connectivity. The reader component can be a standard

device, as it only has to read out the Product Validation Data from defined memory addresses. The computational unit is responsible for checking the integrity of the Product Validation Data, i.e. it checks if the information on the tag has been altered (to describe a different, potentially counterfeit article). Therefore, the communication interface may retrieve the public key from the manufacturer via a trusted source. In an alternative setting, the user terminal could also store the required private keys which would allow for offline product checks.

An advantage of the approach is that low-cost tags with approximately 32 to 64 Bytes of memory can be used. It does not rely on cryptographic functions implemented in the transponders, which would require more expensive tags. The approach can also be combined with plausibility checks that are based on track and trace or secure tag authentication principles to avert cloning attacks.

3 Conclusion

In this chapter, we proposed a novel anti-counterfeiting security solution based on passive, low-cost RFID transponders. The exceptional feature of the approach is that it tightly couples transponders to the object they authenticate. The tags contain verifiable, item-specific information. Thus, a tag that is applied to a product is tightly bonded to that item, providing a measure to avert cloning attacks. The solution is also adaptable for offline checks if no network connection is available. One drawback of the approach is that it is only suitable if the object to be protected has a unique, specific property that can be inspected in a cost-efficient manner. Application scenarios include not only individual products but also electronic freight papers and customs documents where the weight of the consignment, information on the source, destination, the shipping date, etc. may serve as object specific data in order to bind the documents to the actual products.

References

1 Nochta, Z., Staake, T., and Fleisch, E. 2006. Product specific security features based on RFID technology. In Proceedings, International Symposium on Applications and the Internet Workshops – SAINTW ’06: 72–75

Chapter 14

Strengthening the Security of Machine-Readable Documents

Mikko Lehtonen[1], Thorsten Staake[2], Florian Michahelles[1], and Elgar Fleisch[1,2]

[1]Information Management, ETH Zurich, 8092 Zurich, Switzerland, {mlehtonen,fmichahelles}@ethz.ch
[2]Institute of Technology Management, University of St.Gallen, 9000 St.Gallen, Switzerland, {thorsten.staake,elgar.fleisch}@unisg.ch

Abstract. There is an on-going trend towards turning paper documents that store personal information or other valuable data into machine-readable form. An example of this trend is the electronic passport that will become common in the near future. In this paper we show how the security of these machine readable documents could be improved by combining RFID with optical memory devices. We propose integrating an optical memory device into the RFID enabled smart document and present methods how these two storage media can be combined to secure the document against threats like illicit scanning, eavesdropping and forgery. The presented approaches make use of the optical document-to-reader channel which is more secure than the radio-frequency communication interface. To demonstrate the potential of our approaches we show how they could overcomea number of existing security and privacy threats of electronic passports.

Keywords: Electronic passport, machine-readable document, optical memory, RFID, security

1 Introduction

Radio frequency identification (RFID) is an important enabling technique of ambient intelligence. In applications like supply chain management it is used as a mere labelling technique [23], while in anti-counterfeiting its role is for example to implement cryptographic challenge-response authentication protocol [21]. As RFID technology becomes more and more pervasive and closer to our everyday life, also the discussion of the relating security and privacy risks increases. Indeed, addressing the security and privacy threats is of great importance for the acceptance and adoption of RFID [28, 27].

Integration of RFID transponders into physical documents has lead to evolution of machine readable documents. The best known application of this field is the electronic passport, or *e-passport*, where an RFID transponder is used to store biometric data of the passport's holder. Millions of e-passports are already in the circulation today [12] and the number will keep increasing – the U.S. alone will issue more than seven million e-passports each year starting from October 2006 [18, 19].

There are numerous applications where tagging physical documents would be interesting: besides e-passports and other travel documents, also for example customs freight papers, security papers (e.g. gift certificates, jewellery appraisals), driver's licenses and vehicle registration papers would benefit from being machine readable through radio-frequency (RF) communication[1]. A common factor of these documents is that they all relate to a physical entity that is not very well suited to be tagged to become a data carrier itself: integrating RFID chips into expensive jewellery, for example, might conflict with its classical, non-technical nature. Also, besides some extreme cases[2], tagging of human beings is not likely to happen. Therefore, even though objects are turning into data carriers through integration of ambient intelligence technologies, there is and will be a need also for separate data carrier documents.

In this paper we propose new ways of combining RFID with optical memory devices to increase the security of machine readable documents. Our goal is to evaluate and show different approaches of the combined use of these devices. It should be noted that throughout this paper we refer to RFID devices in a broad sense that comprises also contactless smart cards. We propose and evaluate four different approaches how this combination could be used to overcome existing security threats of machine readable documents in terms of more secure communication protocols and resistance against forgery and cloning. Instead of establishing security based on sharing secrets between the reader device and document before the communication, we make use of optical memory devices which cannot be read or eavesdropped without a line of sight.

This paper is organized as follows. In section 2 we discuss machine readable documents in general. A general model of the technical infrastructure for RFID enabled machine readable documents is presented in subsection 2.1 and an overview to travel documents in subsection 2.2. The security and privacy of machine readable travel documents is discussed in subsection 2.3. Section 3 presents optical memory devices and four approaches how we combine them with RFID in physical documents to achieve specific security objectives. In section 4 we discuss the security of the proposed communication models and we finish with conclusions.

[1] Most passports and driver's licenses of today are machine readable through optical character recognition.

[2] http://amal.net/rfid.html

2 Machine Readable Documents

Physical documents can be made machine readable by integrating RFID transponders into them. This creates a link between the physical world and the virtual world and can extend the role of the documents. Within this paper we denote all physical documents that carry a digital memory device as machine readable documents. The typical instance of these kinds of documents is an RFID tagged paper. Another way to make documents machine readable is to use optical character recognition (OCR) to read data printed on the document.

In existing and proposed applications RFID tags are integrated into documents to enable automated document tracking [20], to increase the security of the documents [22, 11] and in general to improve the document handling processes, like the biometric authentication using e-passports [8]. The possible applications of machine readable documents are as manifold as those of normal documents, and more. This is made possible by the digital storage and, optionally, by the logic of the integrated circuits.

The benefits of having RFID transponders in physical documents come from the simple and fast read processes that does not demand a line of sight connection. Depending on the grade and price of the chip, the contactless memory device can also support for re-writable memory and logical functions like cryptographic primitives. Therefore machine readable documents can also provide high level of security and counterfeit resistance.

The following subsection presents the general technical infrastructure of machine readable documents application. Because travel documents and especially e-passports are the most discussed application of machine readable documents within the scientific community, we concentrate on them in subsection 2.2 and on their security and privacy threats in subsection 2.3.

2.1 Technical Infrastructure

The considered components of an RFID enabled machine readable document application are the document itself, the reader device and the reader's control and crypto unit. These components and their mutual communication channels are illustrated in Figure 1. The document is a physical entity that contains an integrated RFID transponder that serves as a contactless memory device. Typically the transponder stores at least a unique identifier (UID) number. In addition, the transponder can provide logical functionalities like access control (through key comparison), random number generation and data encryption. Thus, the transponder serves as more than a mere barcode label.

The two-way communication in the air interface between the contactless memory device and the reader is indicated as a two directional arrow in Figure 1. Without specific addressing, RFID air interface is not secure and the transponder is vulnerable to clandestine scanning (or *skimming*) and eavesdropping. These two security threats are denoted as dashed lines in the illustration. A commonly used standard for

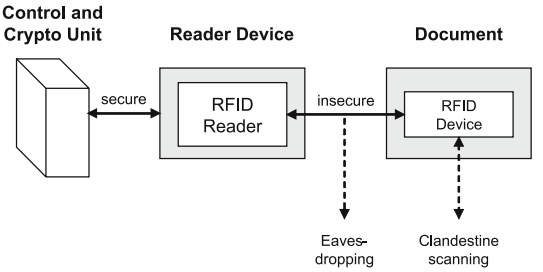


Fig. 1 Technical infrastructure and communication channels of RFID enabled machine readable documents. Security threats of eavesdropping and clandestine scanning are illustrated with dashed lines.

RFID air interface for machine readable documents like e-passports is the ISO 14443 for proximity cards.

The reader device is responsible of the wireless communication. It is connected to the control and crypto unit through a closed, secure channel. The last component in the general infrastructure is the online database that represents data on the network. Though this database is not used in the proposed approaches, it is presented to complete the general view.

2.2 Travel Documents

Machine readable travel documents (MRTD) comprise e-passports, visas and special purpose ID/border-crossing cards [8]. Because of their similar nature, we include also driver's licenses within this group. Most public discussion around MRTDs has been around the electronic passports. The first e-passports were issued in 1998 by Malaysia, followed by other early adopters [5]. The Malaysian e-passports use an RFID transponder to store a fingerprint image of the passport's holder, which enables automated border checks with less human oversight. E-passports will be a prominent and widespread form of identification within a couple of years [12] as its adaptation is fuelled for example by the U.S. Visa-Waiver Program [3] that involves twenty-seven nations. Not all MRTDs of today use RFID technology. Currently the vast majority of U.S. states use or have plans to use 2-D barcodes to store personal data on driver's licenses [1].

The role of the digital memory devices in the authentication processes of travel documents is twofold: on the one hand they help authenticating the traveller and on the other hand they help proving the authenticity of the document itself. Current e-passport (de-facto) standards are given by the ICAO guidelines [8]. They define only one mandatory security measure that is digital signature. Verifying the integrity of biometric features is of primary importance for passports, but addressing only data integrity leaves the system open to various security and privacy threats. The ICAO guidelines do define other cryptographic features that make use of public-key infrastructure, but these are optional.

E-Passport design has to address needs for individual privacy and national security and thus it poses severe security and privacy requirements. These requirements are discussed in [12] and [9]. First of all, the integrity and authenticity of the data the passport stores has to be guaranteed. Second, the data has to be kept confidential from non-authorized parties. Third, the passport must not pose privacy threats for its carrier and, furthermore, all these have to be fulfilled in a public system during up to 10 year long life-span of the passport.

2.3 Security and Privacy Threats

Most discussion about security and privacy of machine readable documents comes from the field of e-passports. Because of their rigid security requirements listed in subsection 2.2, also we concentrate on e-passports in order to provide a short overview of common security threats of machine readable documents in general.

Juels et al. [12] have discussed the security issues of e-passports and the following four threats, among others, were brought into light: clandestine scanning, clandestine tracking, eavesdropping, and cryptographic weaknesses. Moreover, the authors concluded that the e-passports do not provide sufficient protection for their biometric data. Threats do not only concern the functionality of the system but the security and privacy of its users as well. Also Pattison [17] has listed his concerns about the security of e-passports, concerning the baseline ICAO guidelines. These concerns comprise: unprotected data, unprotected wireless transmission, and missing connection between the chip and the paper. The last of these concerns is relevant regarding forgery because without this connection, the system can be fooled for example by putting a valid transponder into a fake paper.

The security of e-passports clearly needs careful addressing – compromising the system would threaten individual and national security. The U.S. State Department has already altered its e-passport design due to privacy concerns [10]. Various proposals for addressing the security and privacy issues of RFID do exist, most often based authentication protocols that use public or symmetric key encryption [2, 14, 9]. Scarce resources on the chip limit the use of cryptographic primitives and the goal of the design is often low-cost low-security features.

In the following section we present how optical memory devices can be combined with RFID to overcome some of the security threats of machine readable documents. The addressed security issues comprise:

- No connection between chip and paper
- Data integrity
- Clandestine scanning
- Clandestine tracking
- Eavesdropping

The first two aforementioned issues relate to the security of the overall system and the latter three to the unsecured wireless communication.

3 Combining RFID and Optical Memory Devices

We propose integrating an optical memory device into the RFID enabled machine readable document. What is common to all optical memory devices is that they need a line of sight connection for reading, making them resistant against clandestine reading and eavesdropping. Therefore we can assume that this channel is secure. The optical memory devices we refer to work normally with write-once-read-many (WORM) principle. Figure 2 illustrates how the addition of optical memory device extends the communication channels between machine readable document and reader device.

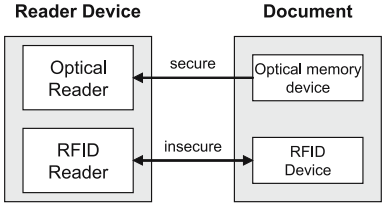


Fig. 2 The communication channels between reader device and document using RFID and optical memory

Before we present our approaches, a short introduction to optical memory devices is provided. Even though we do specify what types of memory devices should be used, basic understanding of possible technologies is necessary for the following discussion.

3.1 Optical Memory Devices

Optical memory device refers to numerous technologies, ranging from barcodes to holographic memories with storage capacities on the order of 1 byte to 100 GB, respectively. The optical memory devices are characterized by their data density (e.g. bytes/mm^2) and can support for error correction coding so that data from partially damaged devices can be successfully recovered. The reader or scanner devices of optical memories use photo sensors, laser and charge-coupled devices (CCD). One interesting advantage of optical memory devices, concerning especially printed codes, is that they are easy to integrate in documents in a rather permanent way as the ink that makes the code is inside the paper. Permanent integration of the memory device is important because it contributes to the security of the overall document.

The simplest low capacity optical memory devices are printed one and two dimensional barcodes. Memory capacities of typical barcodes vary from 95 bits of EAN. UCC-12 barcode to maximum data density of about 850 bits per cm^2 of PDF-417 2-D code. In general, the maximum memory capacity of 2-D barcodes is defined by the accuracy of printing and scanning and the redundancy of the code. The printed high-density 2-D codes can provide data capacities up to about 1,250 KB per cm^2 [26].

Holograms and holographic memory form another type of optical memory devices. Since the early seventies, it has been seen as the high capacity storage solution of the future, promising data densities on the order of hundreds of megabytes per square centimetre and remarkable improvements to the data transfer rate [29]. Hologram based optical memory devices are currently being used for example as anti-counterfeiting labels [24]. Also other promising optical memory technologies emerge, for example photoaddressable polymers (PAPs) which offer re-writeable (RW) data storage capabilities [6]. PAPs are promising recording materials for optical data storage applications such as high-capacity DVDs and holographic memory.

In the following subsections we present four approaches how the combination of RFID and optical memory devices can be used to increase the security of machine readable documents. First two approaches address data integrity and bind the chip and the document, while the two other approaches aim at securing the communication. For the sake of simplicity, we use the term reader in the rest of this paper for the combination of optical and RFID reader devices and their control and crypto unit. Because machine readable documents often relate to a physical entity, we assume that data of interest that the document stores relates to this entity. We denote this data as *object specific data* and it can be used for example in authentication. In addition, the documents can store any other application specific data which is merely referred to as *other data*. This other data can be static or dynamic.

3.2 Storing the Object Specific Data in the Optical Memory

In this basic approach, the static object specific data is stored on both the RFID transponder and on the optical memory. This is illustrated in Figure 3. Mirroring the data of interest helps to maintain redundancy and thus increases the reliability of the overall document. Redundancies on other media than contactless memory devices may be important as the electronic devices can be destroyed without visual effect. Moreover, the additional use of optical storage devices may help overcoming problems resulting from limited storage capacities of RFID devices.

The advantage of mirroring the data is to have a mechanism – indicated by the *result* block in Figure 3 – that can tell if one of the two devices has been tampered with. This increases the integrity of the data, though an identical tampering of both

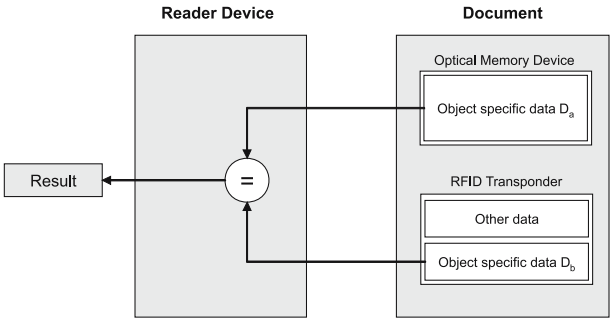


Fig. 3 An illustration of storing the object specific data in the optical memory device

devices cannot be detected. In addition, the object or person specific data also interlinks the two devices in an unquestionable way, since this data can be used as a unique identifier of the physical entity. A disadvantage of this approach is that a relatively large size optical memory is needed. Furthermore, the optical memory doesn't provide access control and thus offers another medium where the data is vulnerable to skimming.

3.3 Storing Hash of the Object Specific Data in Optical Memory

The basic step for taking advantage of the additional data integrity and bind between paper and chip of the first approach while guarding the object specific data from optical access is to save only a hash value of the data on the optical device. This comes with the expense of some extra computations and losing the optical backup of the data of interest. The data structures of the memory devices and the data integrity check of the document using this approach are illustrated in Figure 4.

The used hash function needs to be known by the party performing the data integrity check so the specification of the hash function is stored on the chip. Moreover, this gives an additional binding between the two storage devices. Including the transponder UID number in the hash-calculation can also strengthen this linkage. The reader has to calculate the hash value of the object specific data loaded from the contactless device in order to perform the integrity check. Compared to the previous approach, another advantage of storing the hash value is that smaller optical storage space is needed.

3.4 Storing Access Keys in the Optical Memory

While the two previous approaches in subsection 3.2 and 3.3 define data integrity checks, this approach aims at protecting the data. In this approach the RFID transponder does not reveal the object specific data if no correct access key (e.g. a PIN

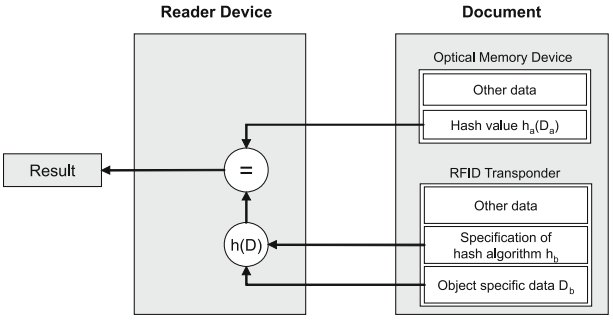


Fig. 4 An illustration of storing hash of the object specific data in the optical memory device

code) has been transmitted in advance, which prevents clandestine reading. Moreover, the data on the electronic tag is read-protected. Also the emission of transponder serial number can be protected in the way described above, protecting the document and its bearer against clandestine tracking. The access key is stored on the optical device and thus can only be read with a line of sight connection. In the context of e-passports, for example, this means that the passport has to be opened for reading and therefore its owner can control who can have access to the contactless memory.

Figure 5 illustrates how the reader can access the object specific data using this approach. Now the requirements of the tag include an access control unit which is capable of generating random indexes and comparing keys. The communication model works as follows: The RFID reader initiates the session by asking the transponder for an index i between 1 and N, where N indicates the number of access keys stored in the document. For the sake of simplicity, this message is not illustrated is Figure 5. After the index value is transmitted by the transponder, the reader can obtain the corresponding key Ki from the optical memory and send it to the RFID transponder. We denote the length of K_i as M (bits). The transponder access control unit verifies if the received access key matches the one stored in its memory and can grant the access for the reader.

In this approach the link between the paper and transponder is strong and both optical and contactless devices are needed for successful communication. In order to access the object specific data, an attacker has to obtain or successfully guess the access key that matches the requested one. Because single access keys can be still obtained by eavesdropping the radio channel between the reader and the transponder, the number of access keys N needs to be large enough to make the malicious use of compromised keys infeasible and spoofing access keys difficult. More precisely, N should be chosen in such a way that the probability of getting access with a compromised key in a single random challenge, $\Pr = 1/N$, is not significantly greater than the probability of guessing an access key, $\Pr = 2^{-M}$.

Spoofing access keys can be also countered by temporarily locking the tag when anyone tries to unlock it using a false access key. However, we leave this to be addressed by the more detailed level protocol design.

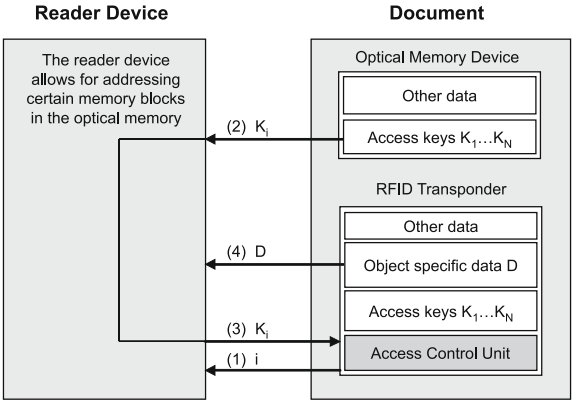


Fig. 5 An illustration of storing access keys in the optical memory device

3.5 Storing Session Keys in the Optical Memory

This fourth approach is similar to the previous one presented in subsection 3.4 where the transponder challenges the reader for a response to be read from the optical device. However, instead of using access keys which can be eavesdropped from the reader-to-tag radio channel, in this approach the optical memory device stores session keys that are used to encrypt the communication.

How the combination of RFID and optical memory is used for protecting the communication in this approach is illustrated in Figure 6. The access control is established by a challenge response pair which is initiated by the tag by transmitting a pseudo-random challenge *ch* and an index *i* between 1 and *N*. After having received the index, the reader accesses the optical memory of the document to obtain the corresponding session key *Ki*. To authenticate itself to the transponder, the crypto-unit of the reader device calculates and sends the response *resp* which is the challenge encrypted using the session key, denoted by K_i(ch) in Figure 6. Last, the session key *Ki* is used to encrypt the following wireless communication (e.g. transmission of the object specific data) which takes place between the reader and the transponder, to provide protection against eavesdropping.

In this approach the successful authentication means that the reader has optical access to the document. Storing session keys in the optical device provides comparable benefits than the approach in subsection 3.4 where access keys are stored on the optical device. Most notably, the passport has to be opened for reading and the two storage media provide strong interlink between the devices. In addition, the use of session keys and encryption provides protection against eavesdropping. Most importantly, the session key is never transmitted in the insecure radio channel as this key is only optically accessible, which overcomes the weakness of the previous approach regarding compromised access keys. On the other hand, this approach requires the transponder to support data encryption.

The presented design has still cryptographic weakness regarding the use of session keys. These keys are not chosen in a truly random manner but taken from

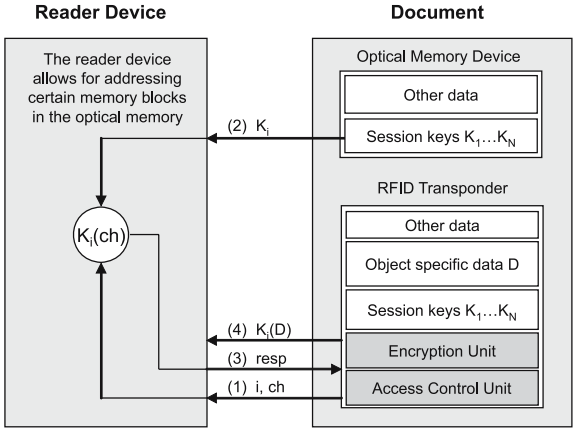


Fig. 6 An illustration of storing session keys in the optical memory device

a restricted list which makes the authentication protocol more vulnerable for attacks. Nevertheless, even the use of only one session key ($N = 1$) provides good protection for the machine readable document.

4 Discussion

Providing machines with the capability to communicate with documents through RFID could dramatically change the way we see and use them: besides precisely knowing where each document is, data on the tag and on the network could be used to manage the pedigree of the documents, to provide digital signatures etc. However, when the documents are of great value or contain personal information, the upcoming security and privacy threats need to be adequately addressed to protect the systems and their users.

We have proposed four approaches to increase the security of machine readable documents that make use of the different properties of optical and contactless memory devices. For the optical memory device, these properties are resistance to clandestine scanning and eavesdropping. The benefits of RFID are their support of logical functions like cryptographic primitives that can be used for example for authentication protocols. Optical devices, especially printed codes, are easy to integrate in paper documents. Also RFID chips, however, can be integrated inside paper; for example the Hitachi μ-chip [22] is designed to be attached to paper documents. In the future, the development of printable polymer electronics [7] may provide novel and interesting ways to seamlessly integrate transponders in paper.

As discussed in subsection 2.3, the ICAO standards for e-passports may not be secure enough. This does not necessary mean, however, that ISO 14443 based proximity smartcards are inappropriate technology for secure machine-readable documents in general. Because RFID tags are generally cheaper than smartcard chips, we can see a trade-off between the proposed approaches that combine RFID and optical memory devices which always require a line-of-sight, and the more expensive contactless smartcards which don't.

We see the main benefits of combining RFID with optical memory devices in the field of document security. These benefits are discussed in the two following subsections. Besides strengthening the security, also other benefits occur, for example the optical memory device can be used to ease the memory capacity requirements of the RFID transponder. This might become especially interesting in the future with the radical data capacity improvements of emerging optical memory technologies. Furthermore, the optical memory device on the document gives a visual cue of the existence of the tag for the users and holders of the document, which can contribute for the acceptance of RFID technology in general.

4.1 Increased Communication Security

The optical channel can be used to overcome threats relating clandestine scanning and eavesdropping in ways presented in subsection 3.4 and 3.5. These approaches require no pre-distribution of shared secrets between the reader and the document, which is favorable in simpler systems; indeed, key distribution is seen as one of the future challenges of RFID security [13]. On the other hand no mutual authentication between the devices is provided. Here the security is established by the assumption that a party who has free visual access to the document is trusted – an assumption that we consider quite feasible regarding for example passports, because, tagged or not, they are to be kept safe and presented to authorized personnel only.

In any case, especially concerning public systems, high level of security needs to be established through secret keys. The presented approaches do not limit the use of public or symmetric key cryptography and so they can be used inside the communication models. Furthermore, the proposed approaches can be combined with each other, namely by selecting one of the first two approaches (subsection 3.2 and 3.3) to guarantee the data integrity and one of the latter two approaches (subsection 3.4 and 3.5) to secure the communication, while taking into account the hardware constraints of reader and memory devices.

With regard to the security and privacy threats of e-passports listed in subsection 2.3, the increased communication security of the approaches presented in subsection 3.4 and 3.5 would help overcome concerns about clandestine scanning, clandestine tracking and eavesdropping.

4.2 Increased Security of the Overall System

Other security contributions of the proposed approaches include increased data integrity, as presented in subsection 3.1 and 3.2. Also a strong bind between the paper and the chip is provided, which answers to the e-passport security concern of missing connection between the paper and the tag, as discussed in subsection 2.3.

The use of two memory devices adds complexity to the system and thus makes the documents harder to be cloned or forged. Even though this conflicts the fundamental security doctrine of Kerckhoffs which says that the security of a system should depend on its key, not on its design obscurity [16], it can provide effective ways to combat counterfeits. For example, the approaches allow for selection of a proprietary optical memory type which cannot be read using devices that are publicly available, if a closed-loop application is preferred.

Also the link between the paper and the tag contributes to cloning resistance: Tags can be made hard to clone by using read protected memories or factory programmed unique transponder ID numbers. An example of how the read protected KILL password of EPCglobal Class-1 Generation-2 tags [4] can be used to strengthen the transponder against cloning can be found in [15]. In addition, special efforts have been made towards anti-clone tags [25]. Consequently, when the seamlessly integrated optical memory device binds the document to an anti-clone tag, cloning and forging the document becomes even harder.

4.3 Related Work

The potential of RFID to secure physical documents is well established. Takaragi et al. [22] have discussed how to prevent forgery of security papers by integrating small RFID tags into the physical documents. The authors compared RFID to barcode and concluded that RFID provides better counterfeit protection due to the fact that it is harder to be copied. However, no link (expect for the physical integration) between the transponder and the document was provided.

The optical channel can be used in RFID also in other ways. Raskar et al. [30] have developed photosensing wireless tags called radio frequency identification and geometry (RFIG) transponders that support for geometric functions. Making use of the photo-sensors, RFIG allows for tags being read only when properly illuminated, which could be used to solve the problem of clandestine scanning of machine readable documents. However, on the contrary to our approaches, no information is transferred in RFIG from the transponder to the reader through this optical channel and, furthermore, no optical memory devices were used. Because of these two facts, our approaches have more potential to increase the security of machine readable documents.

Combining RFID with data printed on the object is not novel in security applications and it has been used for example for privacy protection of tagged banknotes [11]. In this approach the printed serial number (and a signature value) of the banknote is read, namely using OCR, and used to bind the banknote and its RFID chip. Printed data is also being used for the security of travel documents in the advanced security methods defined by optional ICAO e-passport guidelines [9]. In the so called basic access control protocol the e-passport has to be opened and clear-text data like passport number and data of birth of the bearer is used to derive secret cryptographic keys. The purpose of this protocol is to prevent skimming and eavesdropping but according to [12] the scheme fails due to too small entropy of the keys and the fact that only one key is provided for the lifetime of the passport.

This review shows that the previously proposed approaches differ from the contribution of this paper in the way that we make use of the secure optical channel from the document to the reader that cannot be eavesdropped. Furthermore, the existing approaches that use machine readable optical data on documents are normally based on OCR of printed clear-text data, whereas we propose using dedicated optical memory devices which support for much larger storage spaces.

5 Conclusions

In this paper we have shown different approaches to combine RFID and optical memory devices in order to increase the security of machine readable documents. In particular, we have presented how the proposed approaches could overcome existing security threats of electronic passports concerning eavesdropping and clandestine scanning and tracking. Instead of establishing security based on sharing secrets between the reader device and the document before the communication, we make use of the optical channel between the document and the reader

which cannot be read or eavesdropped without a line of sight. Even though strong security in communications always needs secret keys, security of RFID enabled machine readable documents will also depend on a strong connection between the transponder and the paper. We have illustrated how optical memory devices can be used to provide this connection. In conclusion, we believe that the interlinked co-existence of RFID and optical memory devices can play an important role for strengthening the security of smart documents of the future.

References

1 American Association for Motor Vehicle Administrators: Current and Planned Technologies for U.S. Jurisdictions. Available from: http://www.aamva.org/standards/stdUSLicenseTech.asp (22.3.2006)

2 Dimitriou, T.: A Lightweight RFID Protocol to protect against Traceability and Cloning attacks. In Conference on Security and Privacy for Emerging Areas in Communication Networks – SecureComm, Athens, Greece, IEEE (2005)

3 U.S. Department of State: Visa-Waiver Program (VWP). Available from: http://travel.state.gov/visa/temp/without/without_1990.html (20.4.2006)

4 EPCglobal: Class-1 Generation-2 UHF RFID Conformance Requirements Specification v. 1.0.2. EPCglobal Public Document, February (2005)

5 RFID Gazette: E-passports. News article, November 8, 2005. Available from: http://www.rfidgazette.org/airline/index.html (28.3.2006)

6 Hagen, R. Bieringer, T.: Photoaddressable Polymers for Optical Data Storage. Advanced Materials Volume 13, Issue 23 (2001) 1805—1810

7 Hammerschmidt, C.: Polymer electronics yet to realize promise. EETimes, November (2004). Available from: http://www.eetimes.com/ (28.3.2006)

8 ICAO: Document 9303, Machine Readable Travel Documents, October (2004). Available from: http://www.icao.int/mrtd/publications/doc.cfm (28.3.2006)

9 ICAO: PKI for Machine Readable Travel Documents offering ICC Read-Only Access, Technical Report, version 1.1, October (2004). Available from: http://www.icao.int/mrtd/publications/doc.cfm (28.3.2006)

10 International Herald Tribune: U.S. to alter passport design because of privacy fears. News Article, April 28, 2005. Available from: http://www.iht.com/articles/2005/04/27/news/passport.php (28.3.2006)

11 Juels, A. Pappu., R.: Squealing Euros: Privacy Protection in RFID-Enabled Banknotes. In: Wright R. (ed.): Financial Cryptography – FC'03. Lecture Notes in Computer Science, Volume 2742, Springer-Verlag, Le Gosier, Guadeloupe, French West Indies, IFCA (2003) 103—121

12 Juels, A., Molnar, D., Wagner, D.: Security and Privacy Issues in E-passports. In Conference on Security and Privacy for Emerging Areas in Communication Networks – SecureComm, Athens, Greece (2005)

13 Juels, A.: RFID Security and Privacy: A research Survey. IEEE Journal on Selected Areas in Communication, Vol. 24, Issue 2 (2006) 381—394

14 Juels, A. Weis, S.: Authenticating pervasive devices with human protocols. In Victor Shoup, editor, Advances in Cryptology – CRYPTO'05. Lecture Notes in Computer Science, Vol. 3126. Springer-Verlag, Santa Barbara, California (2005) 293–308

15 Juels, A.: Strengthening EPC Tags Against Cloning. In: Jakobsson, M., Poovendran, R. (ed.): ACM Workshop on Wireless Security (2005) 67–76

16 Kerckhoffs, A.: La Cryptographie Militaire. In Journal des Sciences Militaires (1883) 5—38. Available from: http://www.petitcolas.net/fabien/kerckhoffs/ (21.4.2006)
17 Pattison, N.: Securing and Enhancing the Privacy of the E-Passport with Contactless Electronic Chips. (2004) Contact: pattison@axalto.com.
18 RFID Journal: U.S. Tests E-Passports. News Article, November 2 (2004). Available from: http://www.rfidjournal.com/article/articleview/1218/1/1/ (19.4.2006)
19 RFID Journal: United States Sets Date for E-Passports. News Article, October 25, 2005. Available from: http://www.rfidjournal.com/article/articleview/1951/1/1/ (19.4.2006)
20 RFID Journal: Roman Lab to Offer Commercial Services. News Article, March 28, 2006. Available from: http://www.rfidjournal.com/article/articleview/2223/1/1/ (19.4.2006)
21 Staake, T., Thiesse, F., Fleisch, E.: Extending the EPC Network – The Potential of RFID in Anti-Counterfeiting. In Proceedings of the 2005 ACM symposium on Applied computing, ACM Press, New York (2005) 1607—1612
22 Takaragi, K., Usami, M., Imura, R., Itsuki, R., Satoh, T.: An Ultra Small Individual Recognition Security Chip. IEEE Micro, November–December (2001)
23 Tellkamp, C., Angerer, A., Fleisch, E., Corsten, D.: From Pallet to Shelf: Improving Data Quality in Retail Supply Chains using RFID. Cutter IT Journal – The Journal of Information Technology Management, Vol. 17, No. 9 (2004) 19—24
24 Tesa AG: Protection system by Tesa Scribos marks spare part packs (2006). Available from: http://www.tesa.com/corporate/211628.html (28.3.2006)
25 Tuyls, P., Batina, L.: RFID-tags for Anti-Counterfeiting. In: Pointcheval, D. (ed.): Topics in Cryptology – CT-RSA – The Cryptographers' Track at the RSA Conference. Lecture Notes in Computer Science, No. 115–131, Springer Verlag, San Jose (2006) 3860
26 Veritec Inc: VSCode®. Technology Overview (2006). Available from: http://www.veritecinc.com/vs_code.html (28.3.2006)
27 Weis, S.: RFID Privacy Workshop: Concerns, Consensus, and Questions. IEEE Security and Privacy, Vol. 02, No. 2 (2004) 48—50
28 Wong, K., Hui, P., Chan, A.: Cryptography and authentication on RFID passive tags for apparel products. Computers in Industry (2006)
29 Wickett, M.J.: Memories of the future. Emerging replacements for semiconductor memory, optical and magnetic disks. Multimedia Systems – MMS, South Hampton, UK (2002)
30 Raskar, R., Beardsley, P., Baar, J., Wang, Y., Dietz, P.H., Lee, J., Leigh, D., Willwacher T.: RFIG lamps: Interacting with a self-describing world via photosensing wireless tags and projectors. ACM Transactions on Graphics (TOG) SIGGRAPH, Vol. 23, No. 3 (2004) 406–415

Chapter 15

Enhancing Security of Class I Generation 2 RFID against Traceability and Cloning

Dang Nguyen Duc[1], Hyunrok Lee[1], and Kwangjo Kim[1]

[1]CAIS Lab (R504), Information and Communication University (ICU). 119 Munjiro, Yuseong-gu, Daejeon, 305−732, Republic of Korea. {nguyenduc, tank, kkj}@icu.ac.kr

Abstract. In this whitepaper, we present a synchronization-based communication protocol for EPCglobal Class-1 Gen-2 RFID devices. The Class-1 Gen-2 RFID tag supports only simple cryptographic primitives like Pseudo-random Number Generator (PRNG) and Cyclic Redundancy Code (CRC). Our protocol is secure in a sense that it prevents the cloned tags and malicious readers from impersonating and abusing legitimate tags, respectively. In addition, our protocol provides that each RFID tag emits a different bit string (pseudonym) or meta-ID when receiving each and every reader's query. Therefore, it makes tracking activities and personal preferences of tag's owner impractical to provide the user's privacy.

Keywords: RFID, authentication, anti-counterfeiting, CRC.

1 Introduction

Radio Frequency Identification (RFID) technology is envisioned as a replacement for Barcode counterpart and expected to be massively deployed in the coming years. The advantages of RFID system over barcode system include many-to-many communication (i.e., one tag can be read by many readers and one reader can read many tags at once), wireless data transmission (versus optical communication, thus requiring light-of-sight, in case of Barcode) and its computing nature. Those major benefits enable much wider range of applications including: supply chain management, library management, anti-counterfeiting banknotes, smart home appliances, etc.

Despite many prospective applications, RFID technology also poses several security and privacy threats which could harm its global adoption. Ironically, the security weakness of RFID technology comes from the most basic operation of a RFID tag, that is to wirelessly release a unique and static bit string known as Electronic Product Code (EPC for short) identifying the object associated with the

tag upon receiving a query request from a reader. Using the unique EPC as a reference, one (equipped with a compatible reader) can track the moving history, the personal preferences and the belongings of a tag's holder. Even worse, absence of secure authentication results in revealing EPC to malicious readers (referred to as skimming attack). Once capturing EPC, an attacker can duplicate genuine tags and use the cloned tags for its malicious purposes. A natural solution to the aforementioned security problems is to employ cryptographic protocol in the RFID system. Unfortunately, the cost of manufacturing a tag has to be extremely low, *e.g.*, less than 30 cents (according to RFID journal, one RFID tag is expected to cost 5 cents by 2007). Therefore, the computationally intensive security protocols widely known in cryptographic literature cannot be incorporated into a small chip with tightly constrained computational power (at least in the foreseeable future).

To foster and publicize RFID technology, standardization is certainly important in order to allow interoperability at large scale. And the most viable standard is proposed by EPCglobal Inc. The latest RFID standard ratified by EPCglobal is named EPCglobal Class-1 Gen-2 RFID specification version 1.09 (Gen-2 RFID for short). We briefly summarize properties of a Gen-2 RFID tag as follows:

- Gen-2 RFID tag is passive, meaning that it receives power supply from readers.
- Gen-2 RFID tag communicates with RFID readers in UHF band (800–960 MHz) and its communication range can be up to 2 ~ 10 m.
- Gen-2 RFID tag supports on-chip *Pseudo-Random Number Generator* (PRNG) and *Cyclic Redundancy Code* (CRC) computation.
- Gen-2 RFID's privacy protection mechanism is to make the tag permanently unusable once it receives the kill command with a valid 32-bit kill PIN (*e.g.*, tag can be *killed* at the point-of-sale).
- Read/Write to Gen-2 RFID tag's memory is allowed only after it is in secure mode (*i.e.*, after receiving access command with a valid 32-bit access PIN).

We would like to note that that privacy protection mechanism suggested in the specification is arguably over-killed. In many scenarios like tracking animal, smart home appliances, *etc.*, the tag should never be killed. Furthermore, in case of supply chain management, the tag is likely to be helpful in many ways after items being purchased (*e.g.*, for warranty purpose). Therefore, in designing a new protocol, we should avoid this kind of mechanism and, hopefully make a new use for the *kill* PIN. For the *access* PIN, we want to stress that it is useless from security point of view since the 16 bits of PIN is XORed with a 16-bit pseudo-random number sent by the tag in a session. Just by eavesdropping the 16-bit pseudo-random number and the XORed PIN, an attacker can easily recover the access PIN. Losing the access PIN is very dangerous because it allows a malicious reader to read/write the entire memory of a tag.

Lots of researchers have proposed several lightweight cryptographic protocols, designed specifically for low-cost RFID tags, to defend against security and privacy threats. Most of the proposed solutions make use of a hash function [7, 8, 9, 10, 13]. Even though the hash function can be efficiently implemented on low-power hardware, it is still beyond current capability of low-cost RFID tag. In particular, current EPCGlobal Class-1 Gen-2 RFID specification does not ratify

any cryptographic hash function like MD5 and SHA-1. Thus, we need to look for another solution which should use only the available functionalities of current RFID standards. In fact, Juels [3] suggested such a scheme to prevent the legitimate tags from being cloned. However, his protocol does not take eavesdropping and privacy issues into consideration, and thus provides no protection against privacy invasion and secret information leakage. In this paper, we present another scheme targeting most of security features for a RFID system including authentication, traffic encryption, and privacy protection as well. Last but not least, we think that a RFID reader should never be fully trusted (but a legitimate one acts honestly) because it is a portable device and would be used by many people. The only trusted party is a RFID system would be the backend server and all the secrets are kept only at tags and backend server's database. In addition, RFID reader should not be able to learn any secret information including PIN and EPC itself from data called *meta-ID* sent by a tag. The meta-ID should be forwarded to the backend server and backend server can retrieve detail object information keyed by that meta-ID. The advantage of this approach is as follows:

- *Accountability* and *Access Control*: The approach enables easy accountability and access control because the backend server is in charge of looking up object information so it can decide who can get which information as well as some statistics (*e.g.*, how many times of object is queried).
- *Reader-to-Tag Authentication*: It is obvious that tag querying will happen most frequent. And because reader needs to contact the backend server in order to learn useful information about an object, there is no need for Reader-to-Tag authentication in this case. Instead, we can require reader to authenticate to the backend server before sending a meta-ID.

2 Some Background

In this section, we will discuss briefly two primitives, namely *Pseudo-random Number Generator* (PRNG) and *CRC Checksum*, which we are going to use in our authentication protocol.

2.1 Pseudo-random Number Generator

When designing a security protocol, one often faces problems of the following forms:

- A value should not used more then twice, *e.g.*, a challenge in a challenge-response authentication protocol.
- A value should not be predictable, *e.g.*, a secret key.

In such cases, we need a randomly chosen number, more specifically a random number generator. Ideally, a random number is a number which is drawn from a pool of n numbers with probability exactly n^{-1}. In other words, each number

among n numbers has equal chance of being chosen. However, it is impossible to realize such truly random number generator. Instead, people come up with close approximation ones (in computational sense) and call them pseudo-random number generator. In a common setting, a PRNG is modelled as a deterministic function whose next output is computed from previous outputs (usually the last output). The output sequence starts from a (randomly chosen) seed number. The security strength of a PRNG depends on the period and probability distribution of the output sequence. A popular class of PRNG has the congruential form of $x_i = ax_{i-1} + b \bmod N$ where x_0 is the seed number and a, b and N are PRNG's parameters. Another popular class of PRNG is based on *Linear Feedback Shift Register* (LFSR) which could be efficiently implemented on a low-cost RFID tag. In this paper, we will use a PRNG to share a new session key between RFID tag and reader for each and every session. In the EPCglobal Class-1 Gen-2 specification, a Gen-2 RFID tag is capable of generating 16-bit pseudo-random number with the following properties (although the detail algorithm is not given):

- The probability that a single 16-bit number j is drawn shall be bounded by $0.8 \times 2^{-16} < \mathbf{Prob}[j] < 1.25 \times 2^{-16}$.
- Among a number of 10,000 tags, the chance that any two tags simultaneously generate the same 16-bit pseudo-random number is less than 0.1%.
- The probability of guessing the next pseudo-random number generated by a tag is less than 0.025% under the assumption that all previous outputs are known to an attacker.

Since Gen-2 standard requires only 16-bit pseudo-random number, the security margin, *i.e.*, success probability of adversary) of a security protocol using such PRNG is usually bounded by 2^{-16}. We suggest that Gen-2 standard should support 32-bit PRNG to take full advantage of 32-bit PIN currently supported by Gen-2 specification. Otherwise, XORing two halves of a 32-bit PIN with the same 16-bit nonce in one session provides no better security than using the full 16-bit PIN.

2.2 CRC Checksum

In our proposed protocol, we also make use of checksum code to provide security and resolve possible collisions at backend server's database. A checksum code is often used to check the integrity of data being sent or received. The popular cryptographic checksum codes are cryptographic hash function, MAC and HMAC. In this paper, we will make use of a well-known, efficient (yet less cryptographically strong) checksum algorithm, namely CRC. This kind of checksum code is currently ratified in EPCglobal Class-1 Gen-2 RFID specification, version 1.09. CRC algorithm treats binary data as a polynomial whose coefficients are in GF(2) (*i.e.*, 1 or 0). For n-bit CRC, an irreducible and primitive polynomial of n degree (called CRC polynomial) over GF(2) should be chosen. The CRC checksum is then computed as a remainder of the division of the original data by the CRC polynomial. For example, the polynomial $x+1$ is a CRC polynomial resulting in 1-bit CRC checksum equivalent to parity bit. In EPCglobal Class-2

Gen-2 specification, a 16-bit CRC checksum is used to detect error in transmitted data and the corresponding CRC polynomial of degree 16 is $x^{16}+x^{12}+x^5+1$. Even though calculating CRC checksum involves polynomial division, it actually can be implemented very efficiently by using a *shift register* in hardware and *look-up table* in software. Generally, if CRC is setup properly, we can expect that the probability of collision on n-bit CRC checksum is about 2^{-n}. Of course, we can always use cryptographically secure checksum algorithms as well.

Note that, CRC checksum of $0^*\|s$ where s is some bit string is the same as CRC checksum of s itself. To avoid this, we should specifically require a bit string s to start with a bit 1.

3 A New Authentication Protocol for Gen-2 RFID Specification

Main Idea. We first think of protecting data transmitted between the tag and reader against eavesdropping. The obvious way is to utilize encryption/decryption and the most simple encryption function that we are aware of is XORing which is popularly used in a stream cipher). The problem now turns to key management issue: that is to ensure that a new encryption key is used in every session. Solving this issue turns out to be a solution to privacy protection as well since RFID tag can XOR EPC with different key in every session, thus, prevent malicious readers from tracking the tag. And we suggest that the simplest, yet most efficient way of key sharing in this scenario is to use the same PRNG with the same seed at both RFID tag side and backend server side (see Figure 1). The session key can be computed by generating a new pseudo-random number from current session key after every session. This computation is required to be done at both RFID tag and reader/backend server in a synchronous way. Otherwise, subsequent traffic cannot be understood by both sides.

The next security problem that we need to solve is authentication. We argue that, in most cases, a reader just needs to know EPC stored in a tag and then eventually contact the backend server to get/update information about the object carrying the tag. Keeping this in mind, we propose that reader-to-tag authentication can be delegated to tag-to-backend server authentication. More specifically, reader can only receive EPC from RFID tag in an encrypted form. It needs to authenticate itself till backend server first, and then, depending on its privileges,

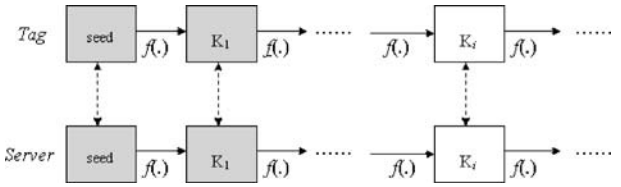


Fig. 1 Using PRNG $f(.)$ to generate session key.

backend server can decide what kind of information to send back to reader (for example, in case of a public reader, only information describing what the referenced object is; and in case of a manufacturer's reader, actual EPC and PIN associated with that tag can be sent). Actual reader-to-tag authentication needs to be carried out when reader wants to access (read/write) other sections of tag's memory bank. To do so, we can use PIN-based approach just like in the original Gen-2 RFID specification.

We also would like to note that, there exists another scheme that allows a reader to be able to decipher EPC without help from backend server for several sessions [10]. We have a different view in this respect. We believe that, in a ubiquitous environment, connectivity is abundant and exercising practical security and simplicity is a key factor to the successful adoption of new technology. In addition, we think that backend server's database can be partitioned in a hierarchical way, thereby reducing overhead at each backend server. This scenario naturally fits in both DNS-like hierarchical structures of EPCglobal Object Naming System (ONS) and real-life situations (for example, each department in a company manages its own inventories, thus, should have its own backend server). We want to stress that our proposed scheme is simple and provides reasonable security strength within the bound of the low-cost RFID tag's functionalities.

Notations. Before describing our protocol in detail, we give the definition of notations that we use in the description of our protocol.

- $f(.)$ – pseudo-random number generator
- CRC(.) – cyclic redundancy check function (produce checksum)
- Ki – secret key at the i-th session
- EPC – Electronic Product Code
- r – random nonce.
- PIN – "access" command password
- T – Tag
- R – Reader
- S – Backend Server

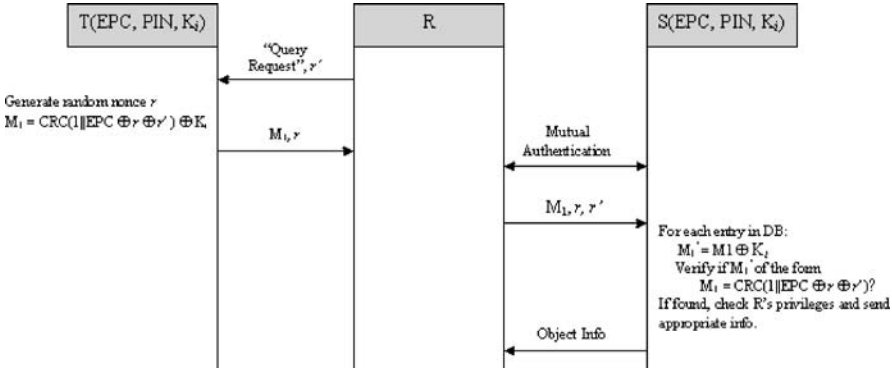


Fig. 2 Tag querying protocol.

The Protocol. There are three sub-protocols including: tag querying protocol, tag access protocol and key updating protocol. During the manufacturing time, a tag is initialized by assigning EPC and other parameters. Then, it chooses a random seed number *seed* and store $K_1 = f(seed)$ into tag's memory and backend server's database entry corresponding to matching EPC. A random PIN (say, *access* PIN defined in Gen-2 specification) is also stored in both tag's memory and backend server database in a similar way. The tag querying protocol is described in the Figure 2.

The tag access protocol is carried out when a reader needs to read/write to tag's memory. In this case, we need to perform actual Reader-to-Tag authentication. The approach is the backend server sends a one-time authentication token to the reader and it uses it to authenticate itself to the tag. The protocol is described as follows:

- $S \rightarrow R$: $M_2 = CRC(1 \| EPC \| PIN \| r) \oplus K_i$
- $R \rightarrow T$: forward authentication token M_2 to T.
- T: Verify $M_2 \oplus K_i = CRC(1 \| EPC \| PIN \| r)$?

The last step of a tag querying or accessing session is to update session key. This step can be optional for non-critical applications but for a security-sesitive application, this should be enforced to guarantee better security. The key updating protocol is described as follows:

- $R \rightarrow T$, S: 'End Session'
- T: $Ki + 1 = f(Ki)$
- S: $Ki + 1 = f(Ki)$

There is might be synchronization issue with the above protocol since a false 'End Session' message might be sent by a malicious reader. To prevent such interference, reader might announce 'End Session' with a token CRC(r' $\oplus$ PIN') where r' is a random nonce broadcasted to Tag with 'Query Request' messae and PIN' is another secret shared between T and legitimate R.

Protocol Analysis. In the querying protocol, the message M_1 is blinded with the session key K_i and therefore we can avoid weak one-way property of CRC checksum. However, CRC checksum is useful for the backend server to search through its database because of its collision-resistant property. If collision does occur due to short length of the checksum, backend server might instruct reader to request for another message M_1. With two different CRC checksums, it is likely that there will be no collision. On the other hand, without knowledge of EPC, PIN and K_i, it is very difficult to construct a message M_1 which can be recognized at backend server. Therefore, the protocol provides tag authentication. As we mentioned before, we require that every input to CRC function should starts with a bit 1. This is to avoid obvious collision attack on CRC.

For reader authentication, Reader-to-Tag authentication is delegated to Reader-to-Backend authentication (which we can use available cryptographic authentication protocol). Actual Reader-to-Tag authentication is achieved by using the token M_2. M_2 is a one-time token because of the session key K_i and the random nounce r.

Table 1 Comparision between Juels' and the proposed protocol

	Juels' Protocol	Our Protocol
Server's complexity	O(n)	O(n)O(CRC)
Reader's complexity	O(q)	O(1)
Tag's complexity	O(q)	2CRC+2PRNG
Tag authentication	YES	YES
Reader authentication	YES	YES
Eavesdropping protection	NO	YES
Privacy protection	NO	YES

Note: n – # of tags; O(CRC) – complexity of CRC; q – number of PIN-test round; Reader-to-Server authentication complexity not counted.

By a the same argument as above, it is difficult to duplicate this token for another session.

Lastly, for privacy concern, the proposed protocol provide privacy protection by randomizing tag's response to each and every query request. In addition, tag never gives out EPC in cleartext, therefore malicous readers have no reference to perform tracking.

A Comparison with Juels' Protocol. We compare our proposed protocol and the one by Juels in the Table 1.

4 Concluding Remarks

We have presented a simple communication protocol for RFID devices, especially EPCglobal Class-1 Gen-2 RFID devices. Our protocol achieves desirable security features of a RFID system including: implicit reader-to-tag authentication, explicit tag-to-reader authentication, traffic encryption and privacy protection (against tracking). Our scheme makes use of only PRNG and CRC which are all ratified in current Gen-2 RFID specification. Moreover, there should be little overhead to adapt our protocol into the Gen-2 RFID specification.

Comparing to Juels' protocol, our suggested protocol offers more security features and better performance at the tag and reader sides. While Juels' protocol requires a tag and a reader to invoke q rounds of communication and PIN testing, our protocol has only one round of communication. We think that reducing computational and communication burden on the RFID tag is very crucial for the sake of the low-cost RFID tag. From security point of view, our scheme is also a more viable solution to the security threats than Juels' scheme. It is because Juels' scheme does not solve the privacy invasion issue which is considered to be the most serious problem faced by RFID technology.

References

1 EPCglobal Inc. Available from: http://www.epcglobalinc.org/
2 EPCglobal Inc.: Class 1 Generation 2 UHF Air Interface Protocol Standard Version 1.09. Available from: http://www.epcglobalinc.org/standards_technology/specifications.html
3 Juels, A.: Strenthening EPC tag against cloning. In: Jakobsson, M., Poovendran, R. (eds.): ACM Workshop on Wireless Security (WiSe) (2005) 67–76
4 Juels, A.: RFID security and privacy: A research survey. In: IEEE Journal on Selected Areas in Communication (2006)
5 Weis, S.: Security and privacy in radio frequency identification devices. In: Master Thesis (2003) Available from: http://theory.lcs.mit.edu/~sweis/masters.pdf
6 Molnar, D., Soppera, A., Wagner, D.: A Scalable, delegatable pseudonym protocol enabling ownership transfer of a RFID tag. In: Preneel, R., Tavares, S. (eds.): Selected Areas in Cryptography (SAC), LNCS, Vol. 3897. Springer-Verlag, Berlin Heidelberg New York (2005) 276–290
7 Ohkubo, M., Suzuki, K., Kinoshita, S.: Efficient hash-chain based RFID privacy protection scheme. In: the Proceedings of International Conference on Ubiquitous Computing, Workshop Privacy (2004)
8 Avoine, G., Dysli, E., Oechslin, P.: Reducing time complexity in RFID system. In: Preneel, R., Tavares, S. (eds.): Selected Areas in Cryptography (SAC), LNCS, Vol. 3897. Springer-Verlag, Berlin Heidelberg New York (2005) 291–306
9 Yang, J.: Security and privacy on authentication protocol for low-cost radio frequency identification. In: Master Thesis (2005). Available from: http://caislab.icu.ac.kr/Paper/thesis_files/2005/thesis_jkyang.pdf
10 Avoine, G., Oechslin, P.: A scalable and provably secure hash-based RFID protocol. In: Proceedings of Workshop on Pervasive Computing and Communications Security (2005)
11 NIST: Random number generation and testing. Available from: http://csrc.nist.gov/rng/
12 RFID Journal: Available from: http://www.rfidjournal.com/
13 Dimitriou, T.: A lightweight RFID protocol to protect against traceability and cloning attacks. In: Proceedings of SecureComm'05 (2005)

Chapter 16

A Random Number Generator for Application in RFID Tags

Wenyi Che[1], Huan Deng[1], Xi Tan[1], and Junyu Wang[1]

[1] Fudan University, Auto-ID Lab, 825 Zhang Heng Road, Zhangjiang High-Tech Park, Shanghai, China 201203. autoidlab@fudan.edu.cn

Abstract. With the extensive use of RFID systems, the problem of information security becomes more and more critical. Cryptography can offer private communications between the RFID reader and tag by using elaborately generated cryptographic keys. These unpredictable and irreproducible secret keys determine the communication security, and they are normally created by a non-deterministic random number generator (RNG) [1]. In current RFID technologies, pseudo random number generators (PRNG) serve as random number sources. Owing to the mechanism of PRNGs, their output numbers show poor randomicity. These less random secret keys, with no doubt, reduce the security of data transmission. An oscillator-based Truly Random Number Generator application scheme in [2] provides a better solution. The TRNG exploits thermal noise of two resistors to modulate the edge of a sampling clock. The white noise based cryptographic keys prevent potential attackers to perform any effective prediction about the generator's output even if the design is well-known. A major topic of this paper is to discuss how to realize a TRNG in the RFID tag system.

Keywords: RFID, Random Number Generator.

1 Introduction

Due to the confidential nature of most cryptographic systems, relatively few hardware RNG designs have been published. Designs available in literature reveal three different IC-compatible methods for producing random sequences, summarized here as follows: direct amplification; oscillator sampling; and discrete-time chaos [3]. Several TRNGs were modeled, analyzed and compared with the method of direct amplification in [4]. The results show that the oscillator-based TRNG is almost free from $1/f$ noise and periodic influences of substrate and power supply. These advantages make the oscillator-based TRNG a desirable solution for RFID tag application.

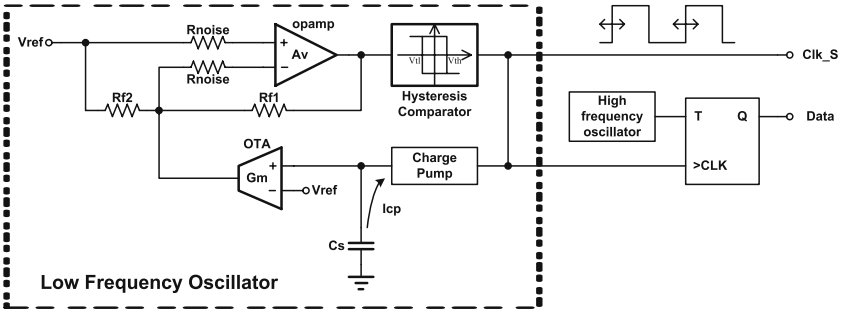


Fig. 1 Circuit structure of oscillator-based TRNG.

In oscillator-based TRNG, a jittered low-frequency clock is used to sample a high-frequency clock. Figure 1 shows the detailed structure of the proposed TRNG. Blocks in the dashed frame constitute a low-frequency oscillator where two resistors' thermal noise is amplified to dither the edges of the low-frequency clock. The high-frequency clock (Clk_F) is oriented from tag's analog front-end. According to EPCglobal RFID Class-1 Generation-2 Protocol, it should be n times of 1.28 MHz, where n is an integral.

1.1 The Low Frequency Oscillator

In Figure 1, an opamp is used as a noise amplifier. Its output is connected to a hysteresis comparator whose output signal Clk_S serves as the sample clock and the

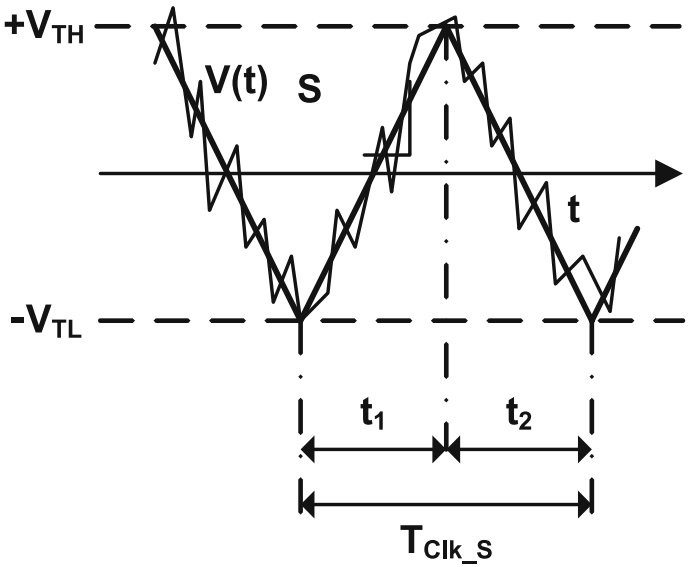


Fig. 2 Noisy triangular wave.

charge pump clock. When Clk_S is high, the capacitor C_S is discharged by a current I_{CP} flowing through the charge pump. The input voltage drop of the OTA due to the discharge of C_S causes a current flowing into the OTA's output. This current results in the voltage drop of the opamp's negative input. The opamp's output voltage keeps increasing until it reaches the high threshold voltage V_{TH} of the hysteresis comparator. Then the Clk_S is turned to be low. The converse process runs in a similar way. This interconversion between V_{TH} and V_{TL} generates a triangular wave at the opamp's output. The noise of R_{noise} is amplified by the opamp and affixed to the triangular wave. Its amplitude follows a Gaussian probability density and is proportional to the value of its resistance. Figure 2 shows the noisy triangular wave where V_{TH} and V_{TL} are the high and low threshold voltages of the hysteresis comparator and S is the wave's slope. The period of the triangular wave equals to the mean period of Clk_S.

In Figure 2, t_1 and t_2 are the rise time and fall time of the triangular wave respectively. They change with the amplified thermal noise of the resistors. A transient voltage of the triangular wave is defined as:

$$V(t) = -V_{TL} + St + V_n(t) \tag{1}$$

where $V_n(t)$ is the amplified noise of the resistors. Since the average value of $V_n(t)$ equals to zero, we have

$$E\{T_{Clk\_S}\} = \frac{2}{S}(V_{TH} + V_{TL}) \tag{2}$$

$$\sigma\{T_{Clk\_S}\} = \frac{\sqrt{2}}{S}\sigma\{V_n\} \tag{3}$$

where $E\{T_{Clk\_S}\}$ is the mean period of the low frequency clock, $\sigma\{T_{Clk\_S}\}$ and $\sigma\{V_n\}$ are standard deviations of Clk_S's jitter and the amplified noise voltage. In (2) and (3), we can see that $\sigma\{T_{Clk\_S}\}$ is proportional to $\sigma\{V_n\}$ and reverse proportional to S, while these two parameters are determined by the characteristics of the circuit,

$$\sigma\{V_n\} = \sqrt{8kTB_W R_{noise} A_V^2} \tag{4}$$

$$S = \pm\frac{I_{CP}}{C_S} G_m R_{f_2} A_V \tag{5}$$

where I_{CP} is the charge pump current, G_m is the OTA transconductance, k is the Boltzmann constant, T is the Kelvin temperature which equals to 300 K at room temperature, A_V and B_W are the close loop gain and bandwidth of the opamp.

1.2 The Sample Circuit

Duty cycle of the high-frequency clock is a factor which may affect the quality of the output random numbers. Since the rise time and fall time of the high-frequency

oscillator is hard to be precisely equal, the duty cycle of the high-frequency clock has a small deviation around 50 percent. A T-flip-flop is used as the sample circuit to overcome this problem. Assuming that the duty cycle of high-frequency clock is 40%, the probabilities of 0 sampling and 1 sampling equal to $P(0)=0.6$, $P(1)=0.4$, respectively. The output probability of the T-flip-flop equals to

$$P_{n+1}(0) = P(0)P_n(0) + P(1)P_n(1) \tag{6}$$

$$P_{n+1}(0) = P(0)P_n(1) + P(1)P_n(0) \tag{7}$$

where $P_n(0)$ and $P_n(1)$ are the n_{th} sample probabilities for 0 and 1, $P_{n+1}(0)$ and $P_{n+1}(1)$ are the $(n+1)_{th}$ sample probabilities for 0 and 1. By (6) and (7), we can have

$$\begin{aligned} P_{n+1}(0) = &\frac{1}{2}\left[\left(P(0)+P(1)\right)^n + \left(P(0)-P(1)\right)^n\right]P_1(0) \\ &+\frac{1}{2}\left[\left(P(0)+P(1)\right)^n - \left(P(0)-P(1)\right)^n\right]P_1(1) \end{aligned} \tag{8}$$

$$\begin{aligned} P_{n+1}(1) = &\frac{1}{2}\left[\left(P(0)+P(1)\right)^n - \left(P(0)-P(1)\right)^n\right]P_1(0) \\ &+\frac{1}{2}\left[\left(P(0)+P(1)\right)^n + \left(P(0)-P(1)\right)^n\right]P_1(1) \end{aligned} \tag{9}$$

Since $P_n(0)+P_n(1)=1$ and $0<\left|P_n(0)+P_n(1)\right|<1$

$$\lim_{n\to\infty} P_{n+1}(0) = \frac{1}{2}\left(P_1(0)+P_1(1)\right) = 0.5 \tag{10}$$

$$\lim_{n\to\infty} P_{n+1}(1) = \frac{1}{2}\left(P_1(0)+P_1(1)\right) = 0.5 \tag{11}$$

2 Design Considerations

In designing the oscillator-based TRNG, a main factor which influences the randomicity of the output sequence is the sample rate. In our finite bandwidth system, the highest sample rate is limited by the noise amplifier. Assuming the output noise of the opamp is pure white noise which follows a Gaussian distribution, the upper limit of sample rate f_s is determined by the equation below [5]:

$$r_x(T_s) = \exp(-2\pi f_0 / f_s) = E\{R_X(1)\} < 0.367(N)^{-\frac{1}{2}} \tag{12}$$

where $r_x(T_s)$ is the continuous-time autocorrelation function, f_0 is the amplifier's pole frequency which equals to the opamp's bandwidth, $E\{R_x(1)\}$ is the estimation of the Gaussian variable, N is the number of the output bits which equals to 16 in

RFID tag application. With a given bandwidth of the noise amplifier, we can calculate the upper limit of sample rate

$$f_s < 1.66 f_0 \tag{13}$$

In order to eliminate the correlation between the 16 bits of the output random number, the sample rate must be less than 1.5 times (roughly) the bandwidth of the noise amplifier. Because the bandwidth of an opamp is a function of power consumption, the highest sample rate is actually limited by the total power available.

The lower limit of sample rate is determined by the period of tag to reader communication cycle. The TRNG must provide 16 bits of truly random number within this period of time. According to EPCglobal RFID Class-1 Generation-2 Protocol, the minimum time of this period equals to 465 μs. Therefore, the lower limit of sample rate

$$f_s > 34 \text{ kHz} \tag{14}$$

2.1 Circuit Characterization

To implement TRNG in the RFID tag, there exist two main constraints: power consumption and chip area. In our scheme, the most important factor is to keep the total power consumption of the low-frequency oscillator at around 1 W. With the state of art, by setting the power supply voltage to 0.8 V, the total current consumption should be no more than 1.3 μA. Table 1 shows the proposed current distribution.

For low power consideration, the output white noise needs to be as small as possible. In respect that the oscillator-based TRNG shows good quality against 1/f noise and some periodic influences, the lower limit of the output white noise is the resolution of the hysteresis comparator. Therefore, it is recommended to set the noise magnitude higher than 3 mV. The difference between the threshold voltages of the hysteresis comparator should be big enough to overcome the input offset, but it may not be too big as to increase the power consumption. Here, we choose the value of 50 mV.

Considering the current consumption budget in Table 1 and the aforementioned restrictions, the circuit specifications of the low-frequency oscillator can be obtained through (2)–(5). Table 2 shows the calculated results.

Table 1 Power budget.

Heading level	Current consumption
Opamp	550 nA
OTA	500 nA
Hysteresis comparator	150 nA
Charge pump	100 nA
Total	1.3 μA

Table 2 Design specification.

Parameter	Value
$\sigma\{V_n\}$	3 mV
$V_{TH}+V_{TL}$	500 nA
B_W	150 nA
A_V	100 nA
PSRR	1.3 μ
R_{noise}	2 MΩ
Clk_F	5.12 MHz
Clk_S	40 kHz
G_m	10 μA/V

2.2 Trade-offs: Power Consumption and Chip Area

As illustrated in Section 2, the design of opamp is of vital importance to the low-frequency oscillator, because the opamp is not only power-consumptive, but also decisive to the system performance. In other words, with a desired noise level, both output data rate and the value of the noise resistors, which is a fatal factor to chip area, are related to the performance of opamp. We tried some new structures [6], [7] of low power opamp using subthreshold techniques and present four proposals by (4). Each combination of the opamp's close loop gain and bandwidth defines a value of the noise resistor. For comparison, the bandwidth of opamp is set to be around 50 kHz. With (13) and (14), we can know that the frequency of Clk_S needs to be in the range of 35 to 75 kHz, and here we choose it to be 40 kHz. Table 3 shows the simulation results of opamps and the corresponding value of noise resistors under SMIC 0.18 μm technology.

In Table 3, there exists a trade-off between power consumption and the resistance. In order to control the current consumption of opamp, more chip area is needed for a large resistance value. Fortunately, accurate absolute resistance is not a rigid requirement, so well resistors with relatively high resistance can be used.

Table 3 Opamp versus noise resistor.

	B_W	Av	I_{DD}	R_{noise}
opamp_1	49 kHz	31.5 dB	550 nA	3.5 MΩ
opamp_2	53 kHz	31.5 dB	614 nA	3.4 MΩ
opamp_3	51 kHz	33.9 dB	933 nA	2.0 MΩ
opamp_4	55 kHz	44.0 dB	2.3 μA	0.2 MΩ

2.3 Trade offs: Power Consumption and Output Data Rate

The other trade-off exists between the power consumption and the output data rate. In designing a comparator, a threshold difference of about 50 mV is needed as mentioned before. From (2) and the sample requirements [8], we can obtain:

$$V_{TH} + V_{TL} = \frac{1}{2} T_{Clk\_S} \times S \tag{15}$$

$$S = \frac{\sqrt{2}\sigma\{V_n\}}{\sigma\{T_{Clk\_S}\}} \tag{16}$$

$$\sigma\{T_{Clk\_S}\} = (10 \sim 20) T_{Clk\_F} \tag{17}$$

With the given values of $\sigma\{V_n\}$ and T_{Clk_F} in Table 2, the calculated $V_{TH} + V_{TL}$ is only 7 mV, which means we need to make $V_{TH} + V_{TL}$ 8 times greater. This can be done by decreasing the frequency of Clk_S or increasing the frequency of Clk_F. The former indicates the reduction of output data rate and the latter indicates larger power consumption in the high-frequency oscillator and the excessive frequency divider. In this way, we give three trade-off proposals. They are:

1. Decreasing Clk_S's frequency by 8 times while not changing Clk_F's frequency
2. Increasing Clk_S's frequency by 8 times while not changing Clk_F's frequency
3. Decreasing Clk_S's frequency by 2.8 times while increasing Clk_F's frequency by 2.8 times

Table 4 shows the clocks' frequencies and the increased current consumption of the aforementioned proposals. In our former design, the frequency of Clk_F used to be 1.28 MHz. In Table 4, we listed the extra current consumption of the high frequency oscillator compared with the case of 1.28 MHz. Notice that in row 2 and 3, the frequencies of Clk_S lower than 34 kHz are not allowed by (14), but they still make sense with a system level optimization illustrated in Section 4.

Table 4 Data rate versus additional current consumption.

	Clk_S	Clk_F	Extra current consumption
1	40 kHz	40.96 MHz	1.8 μA
2	5 kHz	5.12 MH	200 nA
3	14 kHz	14.3 MHz	670 nA

3 System Level Optimization

In Section III, we discussed design considerations of a real-time TRNG and the constraints of its implementation in RFID tag. Because of the low power consideration, we were about to have larger chip area and a lower random number output rate. In order not to do these sacrifices, two system level optimization methods can be employed to improve the overall performance.

3.1 Combination of TRNG and PRNG

For low power consideration, the frequency of Clk_S needs to be as slow as possible. The method of combining TRNG and PRNG may be a promising way to decrease the lower limit of f_s given by (14).

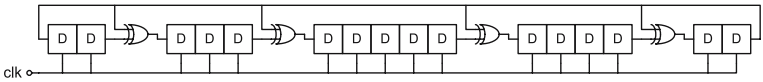


Fig. 3 A typical 16-bit LFSR.

In our former design, a typical 16-bit linear feedback shift register (LFSR) is implemented as a PRNG. Figure 3 illustrates the structure of the 16-bit LFSR. It exploits the initial states of the 16 D-flip-flops to generate random numbers with its cycle ring. A fatal disadvantage of the LFSR is that its output random numbers cycle after a certain period. If we add 1-bit truly random number in the cycle ring as a random number seed, which is generated by the aforementioned TRNG, the output sequence of the LFSR will also be unpredictable and irreproducible as a TRNG. Figure 4 is our proposal of the modified LFSR.

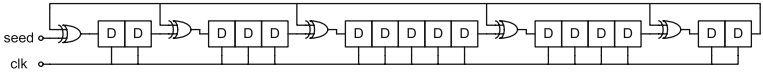


Fig. 4 A modified 16-bit LFSR.

By giving only a 1-bit truly random number as the random number seed instead of generating 16 bits within the time limit, the lower limit of sample rate can be decreased to 2.2 kHz, thus remarkably cut down the power consumption.

3.2 Power-on Generation

Power-on generation is another solution to have high quality random numbers with the limited power consumption. The basic idea of power-on generation is to generate all the random numbers that will be used according to security protocols

before other circuit blocks are awoken. Right after power on, the tag is set to random number generation mode. During this period of time, the TRNG is turned on, and most of the other circuit blocks in the tag are in sleep mode. The tag will not respond to the "Query" command sent by the reader until all the random numbers are prepared. Then the TRNG is turned off and the tag system goes into its natural working mode. Compared with real-time generation, the power-on generation method can provide the TRNG with a larger power budget, therefore providing the possibility of ever enhancing the randomness of the output sequences.

4 Conclusion

This paper introduced the principle of an oscillator-based TRNG. By characterizing the TRNG's power consumption, sample rate, chip area, and the output randomicity, the authors show that it is possible to implement a TRNG in the RFID tag system as a solution to security problems. Finally, two system level optimization methods were proposed to reduce the power consumption of the TRNG.

References

1 Schneier, B.: Applied Cryptography. JWiley, New York (1994) 1
2 Bucci, M., et al.: A High-speed oscillator-based truly random number source for cryptographic applications on a smart card IC. In IEEE Transactions on Computers, Vol. 52, No. 4, pp.403–409, 2003
3 Craig S.P., Connelly, J.A.: A noise-based IC random number generator for applications in cryptography. In: IEEE Transactions on Circuits and Systems, Vol. 47, No.5, (2000)
4 Craig S.P., Connelly, J.A.: Modeling and simulation of oscillator-based random number generators. In: IEEE Proceedings of ISCAS (1996) 4:324–327
5 Craig S.P., Connelly, J.A.: The sampling of noise for random number generation. In: IEEE Proceedings of ISCAS (1999) 6:26–29
6 Boni, A.: Op-Amps and startup circuits for CMOS band gap references with near 1-V supply. In: IEEE Journal of Solid-State Circuits, Vol. 37, No. 10 (2002)
7 Adalan, B., et al.: Low voltage low power operational amplifiers. In: IEEE proceedings of ICECS (2003) 822–825
8 Jun, B., Kocher, P.: The Intel random number generator. In: Cryptographic Research Inc., White Paper Prepared for Intel Corp. (1999). Available from: http://www.cryptography.com/resources/whitepapers/IntelRNG.pdf.

Chapter 17

A Low Cost Solution to Cloning and Authentication Based on a Lightweight Primitive

Damith C. Ranasinghe[1], Srinivas Devadas[2], and Peter H. Cole[1]

[1]School of Electrical and Electronics Engineering, The University of Adelaide, SA 5005, Australia. {damith, cole} @eleceng.adelaide.edu.au
[2]Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA. devadas@mit.edu

Abstract. This paper proposes a solution to address the issue of authentication to prevent counterfeiting in a low cost RFID based system based on using a lightweight primitive, Physically Unclonable Functions.

Keywords: RFID, Authentication, Counterfeiting, Physically Unclonable Functions

1 Introduction

In the implementation of Radio Frequency Identification (RFID) systems concerns have been raised regarding security and violations of end-user privacy. Those concerns may be alleviated using cryptographic primitives.

Despite the vast array of RFID systems, those that are at the low cost end pose the greatest threat to security and privacy due to the possibility of wide scale deployment and inherent constraints that place severe limitations on the number of possible solutions.

There is a large collection of literature available on efficient and inexpensive cryptographic engines suitable for smart card applications, but the use of such engines is an extravagant solution for low-cost RFID systems that are beginning to proliferate within global supply chains.

A primary concern with current low cost RFID systems is a cloning attack. The following sections will examine the vulnerabilities of low cost RFID systems to cloning attacks and the consequences of such an attack. A simple means of addressing

this issue is to implement a security service on the tag that can achieve the security objective of authentication. The issues are elaborated below.

1.1 Notation

It is appropriate, before proceeding any further, to discuss a number of notational aspects to improve the clarity of the discussions below.

An encryption function performed using a key K will be indicated by the expression $e_K(<plaintext>)$ while a decryption function using the same key will be given by $d_K(<ciphertext>)$. If the pair of keys used is public and private they will be distinguished as $K_{private}$ and K_{public}. However a hash function operation on a string of *plaintext* using key K will be expressed as $hash_K(<plaintext>)$.

The exclusive-or operator will be notated in diagrams and equations using the '$\oplus$' symbol throughout this paper, while its usage in a sentence will be termed as XOR.

A random number or a nonce will be denoted by *RN* where there is a series of random numbers used it will be denoted as *RN1*, *RN2*, *RN3*, … , *RNn*, while the notation *RN*(*i*) will note the *i*th random number chosen or used.

In order to distinguish commands specified in the EPCglobal C1G2 protocol specification, or commands proposed for usage in a protocol, the usage of these commands will be in Courier characters. For instance the command issued by a reader to write data to a tags memory will be expressed as `Write`.

1.2 Cloning

Cloning genuine RFID tags to impersonate tags (imitating the behaviour of a genuine tag) presents a serious threat to an RFID system. Using cloning attacks to impersonate tags will add a new dimension to thieving as attackers are able to write EPC data onto devices that function like RFID tags.

A direct consequence of cloning is the possibility for counterfeiting, where a genuine article tagged with an RFID label, may be reproduced as a cheap counterfeit and tagged with a clone of the authentic RFID label. The 'track and trace' concept outlined in [1] is one possible solution to detecting such a counterfeited product in a supply chain application.

At the time of writing there is no mechanism for a reader to verify that it is communicating with a genuine RFID label and not a fraudulent label. Thus a thief may replace a tag of a valid item with a fake tag or replace the tag of an expensive item with that of a fake tag with data obtained from a cheaper item. Hence the lack of a means for authentication allows an adversary to fool a security system into perceiving that the item is still present or this may fool automated checkout counters into charging for a cheaper item. Such fake labels may also be used to create imitation items.

Since there is presently no mechanism for a reader to authenticate itself to a label or a label to authenticate itself to a reader, labels and readers are constantly

in an un-trusted environment where the integrity of messages is doubtful and there are no means for establishing the legitimacy of a reader by a label or the legitimacy of a label by a reader.

Clearly more expensive RFID system implementations are also not immune from cloning as shown by a more recent cloning attack published in [2] where a cloned tag was used in the purchase of fuel at a service station and to start an automobile locked with a RFID based car immobilizer. A similar example of cloning of proximity cards is given in [3] while the possibility of cloning the VeriChip [4] in a discussion of its possible use in tagging employees was outlined in [5].

1.3 Authentication

In an RFID context authentication simplifies to the corroboration of the identity of a tag or a reader. Authentication is an important RFID security measure for preventing counterfeit manufacture or substitution by cloning authentic RFID labels. It is also important for controlling access to label contents. Use of authentication may also be required in other applications of RFID technology such as baggage reconciliation or secure entry systems.

The goal of an authentication scheme in RFID is to prevent an adversary from creating a fake tag to misrepresent the legitimate tag (and hence the authenticity of the object associated with the tag) by a carefully planned attack on the RFID system. There are a number of possible ways in which a low cost RFID system may be attacked to obtain the necessary information to clone a tag. In present systems based on EPCglobal Class I and Class II tags, passive eavesdropping or a scan of an RFID tag is enough to carry out a cloning attack [6].

1.4 Challenge-Response Protocol

Practically all identification schemes or authentication schemes use a challenge-response protocol as illustrated in Figure 1. Other identification schemes such as the Schnorr Identification Scheme [7] and the Okamoto Identification Scheme [8] are examples of more complex challenge and response mechanisms. The mechanism for authentication using challenge and response is described below [9].

1. Reader choses a challenge, x, which is a random number and transmits it to the reader.
2. The label computes $y = e_K(x)$ and transmits the value y to the reader (here e is the encryption rule that is publicly known and k is a secret key know only to the reader and the particular label).
3. The reader then computes $y' = e_K(x)$.
4. Then the reader verifies that $y' = y$.

Fig. 1 Challenge-response protocol

In the context of an RFID system, where there is no secure channel for communication, the security of the mechanism relies on the secure storage of the key k and the inability of an adversary to compute the key k given both the ciphertext and plaintext.

1.5 Constructing a Challenge-and-Response Protocol

It is possible to construct a challenge-and-response protocol using a variety of cryptographic tools. Most symmetric key encryption algorithms, such as AES, are suitable candidates. However, in terms of silicon they present expensive solutions, while at the same time the security provided by such schemes remains vulnerable to various invasive and non-invasive physical attacks [10].

Attacks such as micro-probing, laser cutting, glitch attacks and power analysis attacks along with reverse engineering techniques used to reconstruct the layout of circuits have enabled adversaries to extract digital keys stored in the memory of integrated circuits. Security systems based on keeping a key a secret have thus been broken as a result.

While various tamper-proofing methods have been developed over the years to counter such physical attacks they might be considered to be an extravagant solution for RFID applications. Such an example is the tamper sensing technology [10]. Using a sensor based on additional metallization layers allows interruptions and short circuits to be detected in the event of an attempt to tamper with the IC. However, such sensors only work while the IC is powered and such a sensor technology can only cause a degree of difficulty to an adversary attempting to obtain the key while the IC is powered as the key can still be extracted when the IC is powered off.

Alternatives to storing keys on insecure hardware devices have been developed. Such an alternative is the introduction of physical one-way functions (POWFs) in [11] and [12]. The solution presented used a laser beam as an input to a transparent optical medium with 3D microstructure and the output was a quantification of the resulting interference pattern. The resultant output is dependent on the frequency and the angle of the laser beam entering the optical medium and the optical characteristics of the medium. The developments in POWFs are related to the Physical Unclonable Functions (PUFs) discussed below. The following Sections will then look at utilizing a PUF as a lightweight primitive to construct a challenge response protocol for low cost RFID systems.

1.6 Physically Unclonable Functions

The concept of using physical unclonable functions (PUFs) is published in [13] and is a result of the early work on POWFs. Below is a general definition of a PUF.

Definition 1.1. A Physical Unclonable Function (PUF) maps a set of challenge inputs to a set of responses utilizing some physical characteristic incorporated in an object. A PUF should also satisfy properties listed below.

1. *Easy to compute*: The time taken for generating the response set should be acceptable or be computable in polynomial time.
2. *Difficult to model*: The amount of information that can be obtained about the response to a randomly chosen challenge by an attacker without access to the physical device based on a polynomial number of measurements conducted previously using only a polynomial amount of resources, such as time, is negligible [15].

The ability to construct a PUF on silicon was outlined in [13, 14, 15]. A PUF structure that can be easily fabricated into an IC using standard CMOS fabrication processes has far reaching consequences. Below is a definition of an integrable physically unclonable function.

Definition 1.2. An Integrable Physical Uncloneable Function (IPUF) is a PUF as identified in Definition 1.1 and also satisfying the following properties.

1. *Inseparability*: An IPUF is integrated and fabricated as part of an ASIC design such that any physical attack would lead to the destruction of the IPUF.
2. *Secure communication*: It is not possible to tamper with the measurement data from an IPUF.

The idea is based on using process variations, which are beyond a manufacturer's control, in wires and transistors on an IC to obtain a characteristic response from each IC when given a certain input. The IPUF circuit is able uniquely to characterize each IC due to manufacturing variations [13]. These individual characteristics then become similar to the secret keys used in a symmetrical encryption scheme. Thus, it is possible to identify and authenticate each IC reliably by observing the IPUF response. The observation of IPUF results reveal that a string of challenge bit sequences can be used to generate a response string unique to each IC.

The particular advantage in this technique lies in the fact that it is difficult for an adversary to construct a model or a device to clone an IPUF as there can be a number of possible challenge-response pairs, exponentially dependant on the number of bits in a challenge. Hence building a model based on an exhaustive search is impractical. More sophisticated attacks that use machine learning methods also fail on PUFs with nonlinear elements [15]. However, the IPUF based structure in [13] is sensitive to noise, especially thermal noise, as wire latencies and gate delays depend on the operating temperature of the device. This leads to reliability issues when trying to obtain consistent responses for a given challenge.

Unreliability due to such environmental variations have been addressed in an IPUF configuration given in [14, 15], wherein a challenge response pair is created using an IPUF circuit based on a differential topology, using only 100s of gates. The design of such an IPUF is considered in the following section where the word PUF will always refer to an IPUF.

1.6.1 Circuit Implementation

The block diagram in Figure 2 depicts the structure of a PUF circuit which is based on the arbiter-based PUF in [15, 16]. The circuit accepts an n bit challenge $b_0, b_2, b_3, \ldots, b_n$ to form two delay paths in 2^n different configurations. In order to generate a response bit, two delay paths are excited simultaneously to allow the transitions to race against each other. The arbiter block at the end of the delay paths determines which rising edge arrives first and sets its output to 0 or 1. The actual implementation of arbiter-based PUFs in [15, 16] uses 64 bit challenges. The details of the switch component are given in Figure 3.

The switch component indicated in Figure 2 is implemented using a pair of two-to-one multiplexers (refer to Figure 3). Depending on the select bit C_i, the switch either allows the signal to travel straight through or swap the delay paths. The arbiter is constructed using a simple transparent latch with an active-low enable input. The arbiter favors the path to output zero since it is preset to zero and requires a setup time constraint to switch to a logic one. Fixing a small number of most significant challenge bits can compensate for this skew by effectively lengthening one delay path. The layout was carefully done to ensure that both paths are symmetrical and arbiter responses are not biased to 0 or 1. However, subsequent implementations has allowed the use of arbiters without skew and thus eliminating the necessity for compensation.

The chip used in testing was built in TSMC's 0.18 micro m, single poly, 6-level metal process with standard cells [16]. The chip contains eight sets of the arbiter-based PUF circuits capable of generating an 8 bit response for a given challenge and a JTAG-like serial interface for communication. The total area of the eight PUF

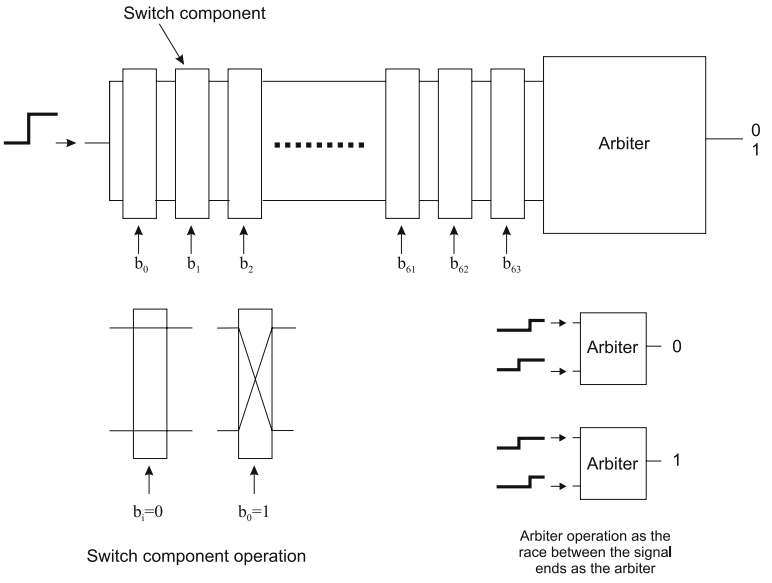


Fig. 2 Arbiter-based PUF circuit implementation.

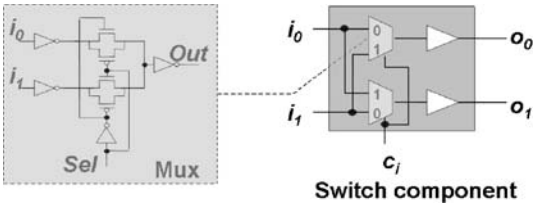


Fig. 3 Switch component implementation using two, two-to-one multiplexers to interchange the two delay paths [15].

circuits and other ring oscillator circuits is 1212 micro m × 1212 micro m and the chip can be operated 100 MHz [15, 16].

1.6.2 Design Analysis

Manufacturers attempt to control process variations to a great degree however, these variations are largely beyond their control and hence it is not possible for an adversary to fabricate identical PUF circuits. It is estimated in [17] that there is a strong enough variation between chip to chip fabricated from the same silicon wafer for a sufficient number of random challenges to identify billions of chips. The probability that the first measured response bits to a given challenge (set of bits) on a chip is different from the measured response for the same set of bits (challenge) on a different chip is estimated to be 23% to 40% depending on the PUF circuit architecture [17]. A symmetric layout will increase this probability to 50 %, and subsequent FPGA implementations have increased this probability further. It has been estimated that about 800 challenge response pairs are sufficient to distinguish 10^9 chips with the probability $p \sim 1 - 5 \times 10^{-10}$ [17]. Such an identification scheme can be implemented with less than 1000 gates on an RFID silicon design.

The input and output functions of the generator are responsible for most of the power consumption in the PUF and the power consumption of the generator core is relatively small. The total power consumption of a the ASIC (Application Specific Integrated Circuit) chip consisting of: 8 arbiter based PUFs (each of which is 64 stages in length), an alternative PUF circuit design based on a ring oscillator circuit, along with logic for interfacing with the JTAG, is about 130 microW in our implementation. This is largely because of the external circuits used for feeding input values to the PUF and obtaining results from the PUF; nevertheless it is relatively a small value.

2 Difficulties of Implementing Authentication on Low Cost Tags

There are a variety of reasons, other than that of tags being in insecure environments for long term secret key storage, behind the difficulties faced by scientists in implementing existing authentication mechanisms in RFID systems. Such reasons follow from the discussion in Section 1.4 of this chapter, and have been addressed in

publications such as [18, 27]. Perhaps the most important of all the issues considered and the most relevant to the current discussion of using PUF circuits for authentication is to consider the fact that the communication between and tag and a reader is constantly exposed to eavesdropping. The aspects of eavesdropping and other vulnerabilities have been dealt with great detail in [6]. This problem is further highlighted by the fact that the current air interface protocol ratified by EPCglobal for Class I tags (C1G2) [19] has provision for establishing a secure communication layer, and it is left up to RFID IC developers to implement such, as perhaps a proprietary solution.

3 Application to Low Cost RFID Authentication

The sections above have detailed the design of a low cost primitive that is suitable for implementation on a RFID IC. The following sections will illustrate the incorporation of a PUF in an RFID IC to provide an authentication service based on a challenge-response protocol.

3.1 A PUF Based RFID IC

The idea of using a PUF in a low cost RFID label was first described in [20]. There are varieties of ways in which such a secret key extraction technique can be incorporated on to a low cost label due to the security it provides to the long term storage of secure keys on an RFID label. Prior to discussing the proposed schemes the following sections will briefly outline the architecture of a low cost passive RFID label IC with a built-in PUF circuit block.

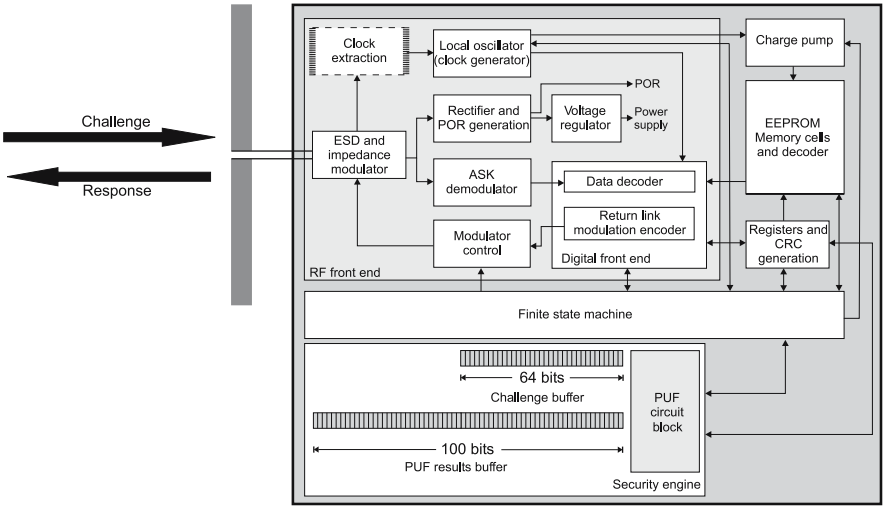


Fig. 4 Block diagram of a passive UHF/HF RFID label with a PUF circuit block.

Figure 4 is simplified block diagram of a typical passive RFID label with a PUF circuit block, where the distinction between UHF and HF is the fact that in a HF system the local clock is generated by dividing down the carrier frequency (13.56 MHz) in steps while for an UHF chip such a division is not possible and thus a low power local oscillator is used. Current fabrications of Class I labels consist of around 1000 to 4000 logic gates while Class II labels may have several thousand more gates. An RFID microcircuit can be subdivided into three primary sections: RF front-end, Memory circuitry, and Finite State Machine (label logic circuitry).

3.1.1 RF Front-end

The RF front-end consists of antenna pads for attaching the terminal of the antenna to the label IC. The antenna input passes through circuits for ESD (electrostatic discharge) protection. The ASK (Amplitude Shift Keying) demodulation circuits extract the modulation dips from the received signal while the Rectifier, rectifies the received signal to generate power which must be regulated using a voltage regulator to avoid voltage surges due to variations in RF field intensities.

Passive RFID chips consist of a relatively large capacitor following a rectifier for storing charge to power the circuit in the absence of a battery. It is important to note here that the capacitor occupies a relatively large portion of the silicon area and RFID chips consuming larger amounts of power will need higher capacity capacitors and thus will cost more.

3.1.2 Memory Circuit

The IC has memory capacity in the order of hundreds of bits. Class 1 labels have only read only memory while Class II labels may have some read-write memory. Read write memory, at the time of writing is implemented using EEPROM and thus requires a large voltage before information can be written to memory. Thus a charge pump, consisting of a series of capacitors is required to achieve a voltage of about 17 V for writing to the tag's memory.

The CRC circuits are used in the validating the CRC in the received data and commands from an interrogator. The CRC generation unit is also used in the computation of the CRC for data sent from the tag to an interrogator before being encoded for modulation by the Return link modulation encoder.

In the implementation of an EPC tag the EEPROM will store the EPC number of the tag, and the rest of the memory (generally of the order of a few kilobytes) is available to the users.

3.1.3 Finite State Machine (Logic Circuitry)

The logic on board the chip will define the label functionality. Primarily, chip logic will execute reader commands and implement an anti-collision scheme that allows the reading of multiple labels by a reader. These logic circuits are highly specialized and optimized for their tasks. The logic circuits also control read and write access to the EEPROM memory circuits.

3.1.4 Security Engine

The security engine consists of a PUF circuit block with an input and an output buffer. The length of the input buffer is capable of storing the next challenge to be sent to the PUF circuit while the output buffer will buffer up to 100 response bits. The set of response bits are transmitted 100 bits at a time under the control of the finite state machine.

3.2 Tag Authentication: A Simple Implementation

Figure 5 illustrates the use of a PUF based RFID system. The discussion below using PUF security engines will assume using 800 challenge-response pairs as a sufficient number of challenges in a single set, as discussed in Section 1.5.2 of this chapter.

Building a symmetric key engine is still not a cost effective solution. Although certain advances have been made towards the development of hardware optimized encryptions engines in [21, 22, 23], they still present a performance hindrance to current RFID systems. Hence instead of using the PUF to obtain a secret key, PUF can be directly utilized as illustrated in Figure 5.

It is clear that once a challenge has been used it cannot be used again since an adversary may have observed it. However it is possible to have a list of challenges and responses or use an encrypted communication link to deliver challenges and obtain the responses. Then there remains the question of delivering a secure communication channel between a reader and the tag. A protocol for obtaining such a communication layer encryption scheme is proposed in [20]. However, not all the challenges need to be discarded; the following method allows a limited reuse of CRPs (challenge-response pairs).

A simple alternative to the mechanism discussed above requires the reader to randomly alter the order in which a challenge set $[C_1, C_2, \ldots, C_{800}]$ is sent to the tag. The tag is then required to store the response, $RES = [R_{001}, \ldots, R_{800}]$ in a buffer (involves buffering the 800 bit long response). The *RES* to the challenge set is then subdivided into four different blocks, *RES1*, *RES2*, *RES3* and *RES4* of length 100 bits each. The resulting blocks are then XORed together as *RES1* XOR *RES2* XOR *RES3* XOR *RES4* as illustrated in Figure 6. Thus a third party observing the communication between a tag and reader is unable to formulate the correct challenge response pairs.

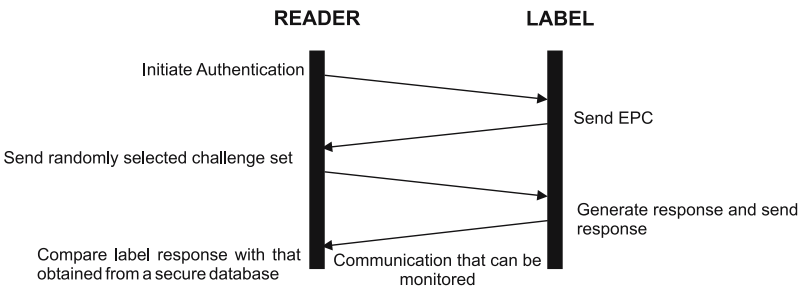


Fig. 5 Message exchange between a reader and an RFID label during an authentication process.

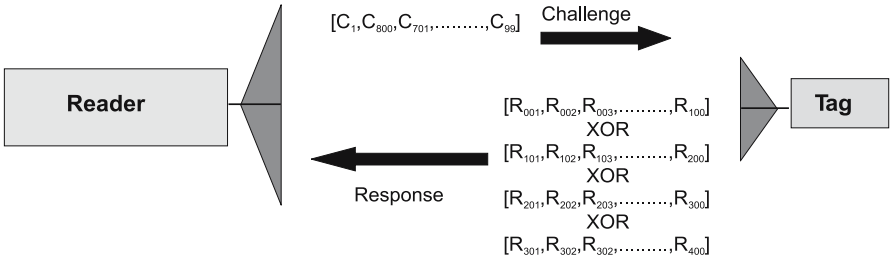


Fig. 6 Using randomised challenges and XORed responses to allow the re-use of challenges.

This random organization of the challenge string will allow the challenges to be reused even over an insecure communication channel as shown in Figure 6.

It is not possible to use the CRPs indefinitely without recharging PUF responses periodically in a secure location as an adversary collecting the challenge and XORed response pairs can construct a set of equations and solve the Boolean satisfiability problem to discover the RES vector with minimal effort. The effort required to collect data sets related to a single tag is also another deterrent to a passive eavesdropper due to the mobility of the tags.

The above scheme will allow a label to authenticate itself to a reader before any sensitive information passes between the devices, but the fact remains that a reader still needs to identify a tag by requesting its unique identifier (such as the EPC in case of Class I tag implemented using the C1G2 protocol). The scheme also implies that the RFID tags be characterized with a number of challenge response sets. Thus in a supply chain environment a manufacturer might have to perform individual tag characterizations using randomly selected challenges in a secure environment without eavesdroppers.

3.3 Tag and Reader Authentication (Mutual Authentication)

It is possible to extend the above scheme to enable a tag to authenticate a reader and for a reader to authenticate a tag (mutual authentication). This involves sending a randomly selected challenge set for which a tag generates a response string. The reply string will be used to authenticate the tag while the reader is authenticated using a one time pad that is updated at the end of the session.

This scheme requires that the tag stores a one-time pad *RN(i)* and a secret key *K* (refer to Figure 7) for encryption and that the reader has access to the tag related information stored in a secure database. The message exchange protocol is outlined in Figure 8. The key *K* is unique to each tag and is stored in a tag's memory at the time of manufacture. It is possible to construct the key *K* so as to provide product authentication by replacing the key *K* with an EPAC (Electronic Product Authentication Code), either encrypted or not, based on measurable or observational product specific features or feature, so that a third party can verify the EPAC independently to establish the authenticity of the product. Such a mechanism is a possible extension that can be adapted to the current protocol. A detailed description of the use of EPACs are described in [27].

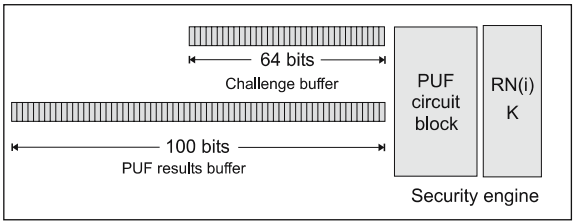


Fig. 7 RFID label with a PUF and additional memory for storing a secret key and *RN(i)*.

Once the tag is identified using its EPC, a reader can request the tag related data set [$RN(i) \oplus K$, $RN(i+1) \oplus CRC\_K$] from a secure database to complete the authentication process. The reader does not require any information regarding the key K, the number $RN(i)$ or the number $RN(i+1)$ in the entire process. The use of a CRC of the key K and the number RN is achieved at no additional cost, since CRC operation can be performed by the CRC block of a tag (refer to Figure 4). The use of the CRC generation hardware extends the usefulness of K and RN, especially K from what will otherwise be a 'use once only' number requiring an extended protocol with greater overheads to refresh two tag related keys. Thus, the CRC generator effectively aids the creation of a lightweight protocol.

The above scheme will require that at least a sufficiently large (about 800 challenges) *CHAL* set be used to be able to effectively identify billions of chips. However measurement noise may cause errors in the response vector and readers will need to access a sequence of redundant information along with a set of CRPs to be able to correct the *RES* vector prior to authenticating a tag. Alternatively a reader can transmit the *RES* vector to the secure database where it can be corrected for errors.

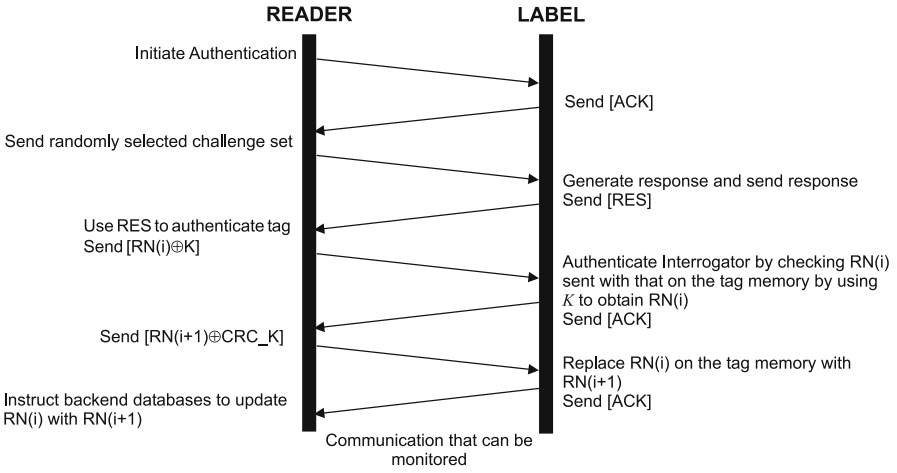


Fig. 8 Protocol for mutual authentication using a PUF.

3.4 A Hash Based Tag Authentication

While the nature of a PUF is ideal for the secure generation of a private key, without having to store the key in the memory of a device in an untrusting environment, its direct application in low cost RFID is limited by the fact that implementing a hash function or another asymmetric encryption engine requires valuable silicon area. Until the optimized cost of hardware implementations of symmetric key cryptosystems becomes a reality such an application context is not possible. However, Figure 9 illustrates a tag authentication scenario when the physical implementation of a hash function achieves the critical cost effectiveness required for low cost RFID. Currently there are no suitable candidates for such a hash function that meet the cost and resource constraints of a low cost RFID label.

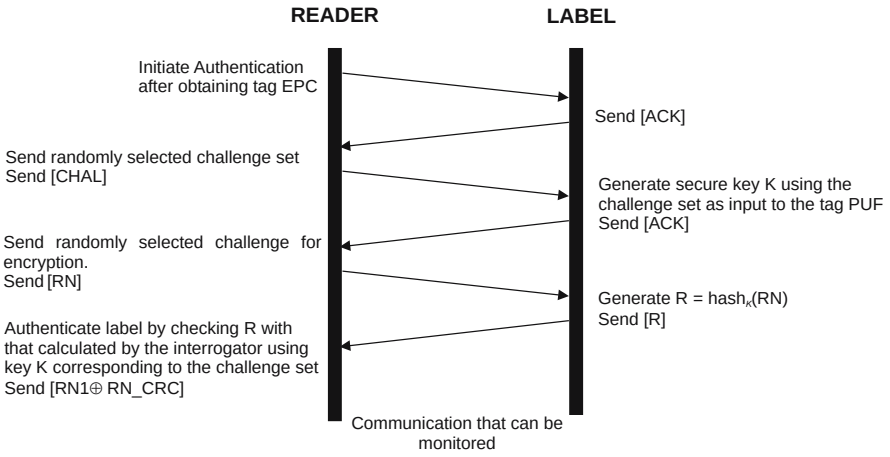


Fig. 9 Challenge-response protocol using a hash function.

3.5 Evaluation of the Mechanisms

The scheme outlined above based around a PUF is evaluated as outlined in Table 1 using a matrix developed for analyzing security mechanism developed for low cost RFID detailed in [27]. Outcomes summed up in Table 1 do not consider the network delays associated with accessing backend databases as such problems may be solved by using adequate bandwidths, and caching tag related data onsite. Clearly the memory space required to store at least 128 challenges incurs a far higher gate cost than that suitable for low cost RFID. Thus, if the number of challenges required is in excess of 128, the challenges almost always need to be transmitted. The transmission delay bottlenecks will then reduce the performance of the system and limit the number of tags that can be authenticated on average to around 2.5 tags per second.

Table 1 Evaluation of authentication mechanisms.

Achieved Security Objectives	Tag authentication Reader authentication
Tag Cost	PUF block: 856 gates Buffer (for storing 164 bits): 1968 gates If the CHAL set is to be stored in memory Memory cost of storing 128, 64 bit challenges: 12,288 gates
Performance	Transmitting 800, 64 bit challenges remain the bottleneck. The transmission from an interrogator to a tag will take 400 milliseconds at the highest possible speed. Thus using 800 challenges will only allow on average the authentication of around 2.5 tags per second. However, it is possible to use perhaps a 128 challenges instead of 800 as the tag EPC will be enough to uniquely identify the tag. The reduced overhead will allow the authentication, on average, of about 16 tags per second.
Backend Resource Requirements	Scheme requires online resources such as access to secure backend databases containing tag profiles.
Overhead Costs	Tags need to under go a verification phase prior to deployment to generate an adequate number of CRPs for each tag.
Power Consumption	No power consumption estimates are available but it is not expected to be greater than that required for writing to EEPROM memory.

The need to transmit a large number of challenges from the reader to the tag remains the primary obstacle, especially given that that the maximum possible transmissions speed from a reader to a tag possible in the C1G2 specification is about 126 kbps given equi-probable ones and zeros. However, higher inter-chip variations of around 50% [24] discussed in Section 1.5.2 and the fact that the EPC can be used for unique identification suggest that for the scheme outlined in Section 2.4 and Section 2.5 a smaller challenge set (of around 128 challenges to generate a 128 bit response) may be used. Using a smaller challenge set will significantly increase the number of tags that can be authenticated to over 15 tags per second.

4 Removing Barriers to Performance

The primary performance obstacle in the tag authentication and mutual authentication protocol presented previously is the excessive overhead of transmitting a large number of challenges, where each challenge consisted of 64 bits. There are several methods by which performance of the above protocol can be improved.

Instead of choosing to transmit the challenges, it is possible to use a linear feed back shift register (LFSR) to generate the challenges once initialized with a seed. Then only the seed to the LFSR needs to be sent to a tag as a challenge C, from a reader. This arrangement is outlined in Figure 10. The additional hardware of

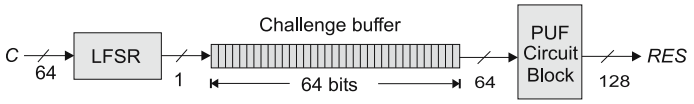


Fig. 10 Improved security engine design of the lightweight primitive to reduce overhead.

a LFSR will allow the protocols in Figure 5, Figure 8 and Figure 9 to be executed with greater efficiency.

4.1 Evaluating Performance

The schemes outlined above can be evaluated to consider the overall performance of PUF based authentication mechanism with the alteration to the security engine discussed in Figure 10. The results are detailed in Table 2. Unlike the authentication schemes outlined previously, where a large number of challenges need to sent to the tag, the current transmission requirement is that of a single challenge C, from which other challenges are derived.

Table 2 Evaluation of authentication mechanisms using the redesigned hardware.

Achieved Security Objectives	Tag authentication Reader authentication
Tag Cost	PUF block: 856 gates Buffer (for storing 64 bits): 768 gates Total cost of hardware: 1624 gates
Performance	Transmitting C, a 64 bit challenge, from an interrogator to a tag will take about 0.5 milliseconds at the highest possible speed. Using 128 challenges in the tag authentication scheme to uniquely identify a tag. The reduced overhead will allow the authentication, on average, of about 2000 tags per second, ignoring the delay on tag for evaluating the PUF responses for the 128 challenges generated on the tag. The mutual authentication protocol execution time will be slightly greater due the requirement for transmitting random numbers. The transmissions from the reader will then take about 1.5 milliseconds, provided RN(i) used is 64 bits in length. This will still allow over 600 tags to under go the mutual authentication process.
Backend Resource Requirements	Scheme still requires online resources such as access to secure backend databases containing tag profiles.
Overhead Costs	Tags still need to under go a verification phase prior to deployment to generate an adequate number of CRPs for each tag.
Power Consumption	Power consumption estimates will be higher than that for the previous schemes due to the integration of a LFSR but it is not expected to be greater than that required for writing to EEPROM memory.

Using a smaller challenge set as well as the incorporation of a LFSR has allowed the lightweight primitive to be implemented in a low cost RFID tag while meeting requirements of both performance and cost.

5 Practical Issues

5.1 Measurement Noise

As mentioned previously, the responses from a PUF are sensitive to environmental conditions such as temperature and power supply voltage. Thus a challenge that generates a reliable response may not generate a reliable response if environmental conditions change beyond a tolerance level. This issue is highlighted in Figure 11 where results of 500 repetitive measurements of 1000 challenges are shown. The design in Figure 2 presented in [15], is used to mitigate the effects of environmental noise based on the fact that noise would affect both signal propagations paths in an identical manner and thus the final results of the circuit are unchanged.

Because the circuit measures the relative delay difference, the PUF is robust against environmental variations. For realistic changes in temperature from 20 to 70 degree Celsius and regulated voltage changes of +/-2%, the output noise is 4.8% and 3.7%, respectively. Even when increasing the temperature by 100 degree Celsius and varying the voltage by 33%, the PUF output noise still remains below 9%. This variation is significantly less than the inter-chip variation of 23% (or 40%), allowing for the reliable identification and authentication of individual chips [24].

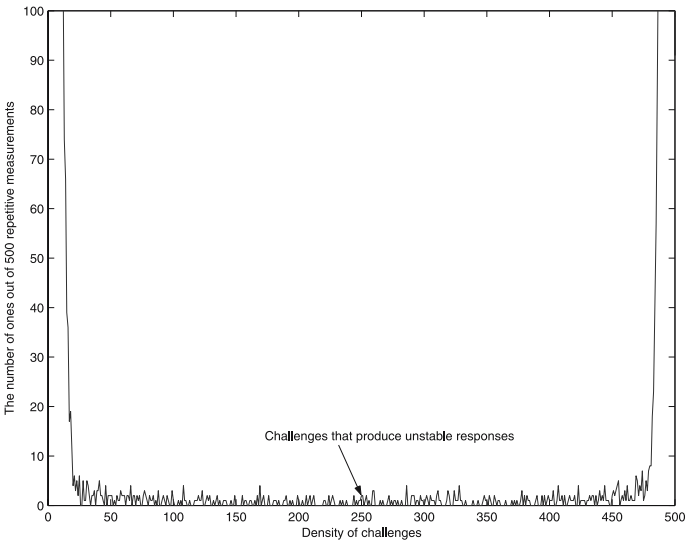


Fig. 11 The density function of the random variable k, where k is the number of 1's out of 500 repetitive measurements.

The generator is sensitive to the power supply voltage and the temperature of the surrounding environment [15]. However problems caused by operational voltage changes can be minimized by the fabrication of a voltage regulator on the PUF.

The problems caused by operational voltage changes can be minimized by the fabrication of a voltage regulator on the PUF and the authentication process can be refused if the PUF is inadequately powered. More recent work has also improved the reliability of PUF systems. Another simple solution to overcome measurement noise is to calculate and transmit redundant information extracted from a *RES* output along with the *CHAL* set to ensure that the response obtained can be evaluated and corrected for errors.

5.2 Addressing Reliability

The use of redundant information to improve the performance can be easily demonstrated by using an error correcting code to encode the measured PUF responses in the verification phase. This will allow error in the PUF responses received from tags as a result of measurement noise (discussed in above) to be corrected as far as that possible with the chosen error correcting code. Noise is a randomly distributed variable, thus it makes sense to use an error correcting code that is best suited for correcting random errors. BCH code named after its discoverers (Bose, Ray-Chaudhuri, Hocquenghem) is a logical choice for the purpose. BCH codes are a generalization of the Hamming codes.

Figure 12 illustrates the use of a BCH(n, k, d) code where n bits of the PUF output is used to calculate a syndrome *S(RES)* of e bits. The syndrome is evaluated during the verification phase and stored securely for use in one of the authentication protocols discussed previously.

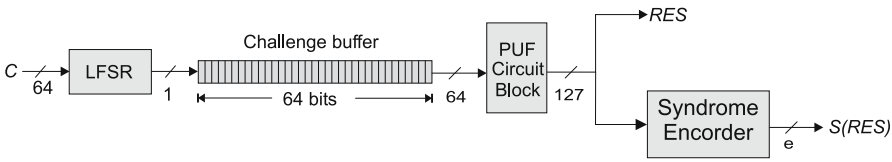


Fig. 12 Encoding of PUF responses using BCH codes at the verification phase.

The use of the *S(RES)* at the interrogator is illustrated in Figure 13. Upon receipt of the *RES* vector from a tag, a reader compares that to *RES* vector evaluated during the verification phase. A failure in the comparison stage results in the attempt to correct any measurement noise using the stored syndromes. The corrected *RES* vector is again evaluated to ensure the comparison failure most likely did not arise from measurement noise but from a fraudulent tag or a malicious user attempting to clone a tag.

Measurements of noise in PUF circuits have shown a bit error rate of around 5%. Thus we can expect about 7 random errors in a 127 bit PUF response vector. However the BCH code parameters used has an ability to correct up to 15 errors.

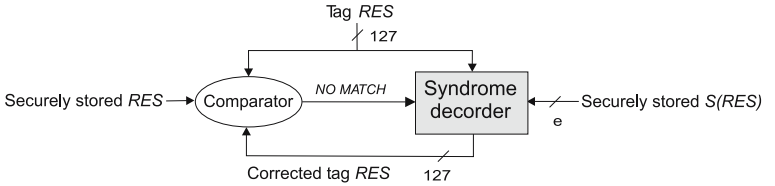


Fig. 13 Comparison of the received response from tag. If the *RES* received from the tag is not a match an attempt is made to correct to a challenge at the interrogator in the authentication protocols.

This is more than double the bit error rate expected from a PUF response and is adequate to ensure the reliability of the authentication process.

5.3 Verification Phase

Figure 14 gives an overview of the system implementation required. Clearly a practical implementation will require that tag IC manufacturers or product manufacturers in the supply chain initially obtain and securely store tag specific sets of challenge response pairs. This will be an expensive task but it can be automated similarly to the test and verification of ICs. While the mechanism for authentication appears simple, issues regarding the secure storage of challenge sets, and the corresponding responses along with a method for distributing the challenge sets and responses to parties in a the supply chain need to be addressed. There are various existing solutions to such issues and they are not considered in this chapter.

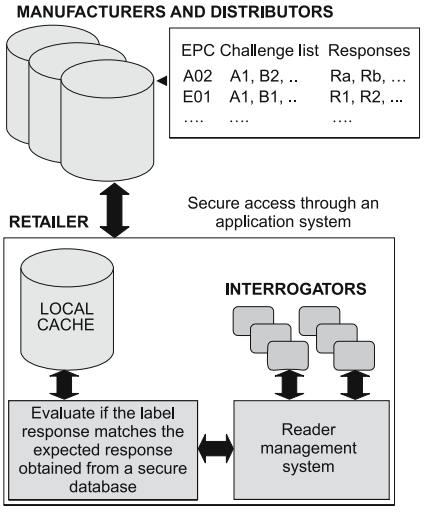


Fig. 14 An overview of an implementation of a PUF based RFID system.

5.4 Air Interface

What has not been addressed thus far is the compatibility of the mechanism with the recently ratified EPCglobal air interface protocol [19] for Class I labels (C1G2). Given the wide scale adoption of C1G2 for consumer goods supply chain applications, the largest market for low cost RFID technology, it is appropriate to consider the ease with which the authentication protocols discussed above can be incorporated into the C1G2 air interface protocol. The schemes described above require the transmission of a challenge from a tag to a reader and the transmission of a tag response. There is no command in the current specification of C1G2 to transfer such a challenge.

The current version of C1G2 protocol does not have commands that could be used to implement the mechanism outlined above. However, the specification does allow proprietary commands to be defined. Implementing the tag authentication mechanism and the mutual authentication scheme would require the definition of at least one proprietary command after initiating an authentication: `WriteChl` (instructs the tag of the number of `WriteChl` commands to expect and transmits a challenge to the tag) and instructs the tag to scroll out the response to the challenges. In order to prevent the challenge lists from being discarded in the event of sudden power loss, it is important for the tag to transfer the results from the PUF output buffer to the tag memory and only to transmit back the response once the entire challenge sequence is evaluated. The reader is required to keep the tag powered until the tag responds with a challenge vector or signals a failure of the authentication process by way of a NAK.

In addition to these commands, the mutual authentication mechanism in Section 2.4 will require a further command; `SendEnc(`*<flag>*`,`*<data>*`)`. The reader command `SendEnc` will be used to send encrypted *RN(i)* to the tag, where the *flag* value will be used to denote the authentication step. Implementing these additional commands will increase IC cost, albeit being only a small cost increase.

6 Possible Attacks

The security of the above system relies on a PUF to securely store a unique secret key in the form of fabrication variations. The PUF based security systems are susceptible to reliability issues as discussed in Section 2.7 of this chapter; however recent work has addressed these issues. The most probable attacks on a PUF based challenge response system are outlined in [15].

The security of the systems based on PUFs will depend on the difficulty of replicating a PUF circuit and on the difficulty of modeling the PUF circuit successfully. This is not a simple process and is therefore an adequate deterrent depending on the value of the article being authenticated by the reader.

7 Conclusion

The PUF provides a cost effective and a reliable solution to meeting the security objective of authentication in low cost RFID systems while meeting end-user performance requirements. This security engine can be easily constructed using standard digital gates and layout tools and fabricated using standard CMOS technology. A 64-stage PUF circuit costs less than 1000 gates. Additionally, various kinds of low power techniques such as sub-threshold logic design and multi-thresholds CMOS design can be utilized to reduce the power consumption to make it suitable for use in devices sensitive to low power consumption.

The effects of environmental conditions on the measurements obtained from a PUF are documented in [20], and the symmetrical nature of the circuit counter acts to reduce much of the variation provided otherwise.

The effects of power supply voltage are being investigated to discover practical performance boundaries such that the PUF can operate reliably. Nevertheless it is possible to fabricate a voltage regulator onboard the PUF to prevent effects from higher voltage variations, but it will not be able to counteract conditions induced by voltages below a calibrated power supply voltage.

Future work should involve the investigations into the effects of voltage on the performance of the PUF. It is also left to investigate whether the generator throughput can be improved to reduce the time taken for the execution of a challenge response protocol.

However, we have shown a technique of using a PUF as a lightweight primitive to evaluate a given challenge reliably in the presence of acceptable levels of measurement noise by utilizing error correcting codes.

References

1 Ranasinghe, D.C., Leong, K.S., Ng, M.L., Engels, D.W., Cole, P.H.: A distributed architecture for a ubiquitous item identification network. In: Seventh International Conference on Ubiquitous computing, Tokyo, Japan, (2005)

2 Bono, S., Green, M., Stubblefield, A., Juels, A., Rubin, A., Szydlo, M.: Security analysis of a cryptographically-enabled RFID Device. In: Proceedings of 14th USENIX Security Symposium, (2005) 1–16

3 Westhues, J.: Hacking the prox card. In: RFID: Applications, Security and Privacy, Addison-Wesley, (2005) 291–300

4 Verichip corporation home page. Available from: http://www.4verichip.com/ (06/2006)

5 Albrecht, K.: Chipping workers poses huge security risks. In: Freemarketnews. Available from: http://www.freemarketnews.com/Analysis/139/3812/ 2006-02-15.asp?wid = 139&nid = 3812 (06/2006)

6 Ranasinghe, D.C., Cole, P.H.: Security and privacy issues. In: Ranasinghe, D.C., Cole, P.H. (eds.): Networked RFID Systems and Lightweight Cryptography, Springer-Verlag, Berlin Heidelberg New York (2007)

7 Schnorr, C.P.: Efficient signature generation by smart cards. In: Journal of Cryptology, Vol. 4 (1991) 161–174

8 Okamoto, T.: Provably secure and practical identification schemes and corresponding signature schemes. In: Lecture Notes in Computer Science, Vol. 196, (1993)
9 Menezes, P. van O., Vanstone, S.: Handbook of Applied Cryptography. CRC Press, (1996)
10 Kommerling, O., Kuhn, M.G.: Design principles for tamper-resistance smartcard processors. In: proceedings of USENIX Workshop Smartcard Technology (1999) 9–20
11 Ravikanth, P.S.: Physical one-way functions. In: PhD dissertation, Department of Media and Art Science, Massachusetts Institute of Technology, Cambridge, (2001)
12 Pappu, R., Recht, B., Taylor, J., Gershen-Feld, N.: Physical one-way functions. In: Science, Vol. 297 (2002) 2026–2030
13 Gassend, B.: Physical random functions. In: M.S. thesis, Department of Electrical Engineering Computer Science, Massachusetts Institute of Technology, Cambridge (2003)
14 Gassend, B., Clarke, D., van Dijk, M., Devadas, S.: Silicon physical random functions. In: Proceedings of Computer Communications Security Conf. (2002). 148–160
15 Lim, D.: Extracting secret keys from integrated circuits. In: Master thesis, Massachusetts Institute of Technology (2004)
16 Lim, D., Lee, J.W., Gassend, B., Suh G.E., van Dijk, M., Devadas, S.: Extracting Secret Keys from Integrated Circuits. In: IEEE Transactions on VLSI Systems. Vol. 13(10) (2005)
17 Lee, J.W., Lim, D., Gassend, B., Suh, van Dijk, M., Devadas, S.: A technique to build a secret key in integrated circuits for identification and authentication applications in 2004 Symposium on VLSI circuits (2004) 176–179
18 Ranasinghe, D.C., Engels, D.W., Cole, P.H.: Low cost RFID systems: confronting security and privacy. In: Auto-ID Labs White Paper Journal. Vol 1 (2005)
19 EPCglobal UHF Class I Generation II Air Interface Protocol v1.0.0 (2006): Available from: http://www.epcglobalinc.org/standards_technology/ (06/2006)
20 Ranasinghe, D.C., Engels, D.W., Cole, P.H.: Security and privacy solutions for low cost RFID Systems. In: Proc. of the 2004 Intelligent Sensors, Sensor Networks & Information Processing Conference, Melbourne, Australia (2004) 337–342
21 Aigner, M., Feldhofer, M.: Secure Symmetric Authentication for RFID Tags. In: Telecommunications and Mobile Computing TCMC2005 (2005)
22 Feldhofer M., Dominikus S., Wolkerstorfer J.: Strong authentication for RFID in Systems using the AES algorithm. In: Lecture Notes in Computer Science, Vol. 3156 (2004) 357–370
23 Wolkerstorfer, J.: Is Elliptic-Curve Cryptography suitable to secure RFID tags. In: Workshop on RFID and Light-Weight Cryptography, Graz Austria (2005)
24 Suh, G.E., O'Donnell, C.W., Sachdev, I., Devadas, S.: Design and implementation of the AEGIS single-chip secure processor using Physical Random Functions. In: Proceedings of the 32nd International Symposium on Computer Architecture, Madison, Wisconsin (2005)
25 Weigart, S.H.: Physical security devices for computer subsystems: a survey of attacks and defences. In: Workshop on Cryptographic Hardware and Embedded Systems. LNCS, Vol. 1965 (2005) 302–317
26 Kommerling, O., Kuhn, M.G.: Design principles for tamper-resistance smartcard processors. In: Proceedings of USENIX Workshop Smartcard Technology (1999) 9–20
27 Ranasinghe, D.C.: New directions in advanced RFID systems. In: PhD dissertation submitted to the School of Electrical and Electronic Engineering, The University of Adelaide, Australia (2007)

Chapter 18
Lightweight Cryptography for Low Cost RFID

Damith C. Ranasinghe[1]

[1] The School of Electrical an Electronic Engineering, The University of Adelaide, Adelaide SA 5005, Australia. damith@eleceng.adelaide.edu.au

Abstract. Security and privacy concerns as well as the need for security services to enable the development of novel applications using low cost RFID have been illustrated in previous Chapters, and in particular various proposals made to address issues regarding information security and end-user privacy have been discussed in Chapter 6. However, some of these ideas are not practicable for secure low-cost RFID on account of their demand for circuit size and operational power while others fail to meet various security and privacy objectives adequately. The solutions presented have not considered: aspects unique to low cost RFID, system performance requirements and consequences and practicability of implementation in a system wide context. This chapter aims to propose a number of practicable solutions based on lightweight cryptography that address the security objectives and privacy goals outlined in Chapter 6 and are based on the low cost RFID framework outlined therein. The proposed solutions are then evaluated for their merits using the evaluation framework developed in Chapter 8.

Keywords: Authentication, Confidentiality, Message content security, Anonymity, Untraceability, PUF, Stream cipher, LFSR, Shrinking generator, Knapsack generator

1 Introduction

It is clear from the discussions in Chapter 6 that it is not possible to incorporate any encryption engine of significance in a Class I or a Class II label. It is in this view that the majority of proposals below will aim at removing complexity from the label to other proxy systems and limit any security related computation on the chip to simple operations.

The proposals below will also be considered on their merits based on meeting the performance metrics and the framework outlined in Chapter 8 while implementations of the mechanisms will be considered in view of the C1G2 air interface protocol.

Prior to proceeding with the proposals a note on notational aspects and an introduction to the hardware and operations used in the mechanisms outlined in this chapter is given in the following sections. The rest of the chapter will outline and evaluate various mechanisms for achieving the security and privacy objectives discussed in Chapter 6.

1.2 Notation

It is appropriate, before proceeding any further, to discuss a number of notational aspects to improve the clarity of the discussions below.

An encryption function performed using a key *K* will be indicated by the expression $e_K(<plaintext>)$ while a decryption function using the same key will be given by $d_K(<ciphertext>)$. If a pair of keys used is public and private they will be distinguished as $K_{private}$ and K_{public}. However a hash function operation on a string of *plaintext* using key *K* will, in particular, be expressed as $hash_K(<plaintext>)$. When a keystream is used, as is the case with a stream cipher, a unique sequence of a keystream will be indicated using italic roman character such as *Ks* and different segments of the same keystream will be noted by appending a number to *Ks* such as *Ks1*, *Ks2* and *Ksn* for the nth segment. Two different keystreams will be distinguished by using a subscript, such as Ks_1 and Ks_2.

A signature scheme used to sign a message using a signing algorithm *sig*, and key *K*, will be denoted as $sig_K(<plaintext>)$ while the corresponding verification algorithm *ver* using key *K* will be denoted as $ver_K(sig_K(<plaintext>),<plaintext>)$.

The exclusive-or operator will be notated in diagrams and equations using the '$\oplus$' symbol throughout this chapter, while its usage in a sentence will be termed as XOR.

A random number or a nonce will be denoted by *RN* where there is a series of random numbers used it will be denoted as *RN1*, *RN2*, *RN3*, … , *RNn*, while the notation *RN*(*i*) will note the *i*th random number chosen or used.

The CRC (cyclic redundancy check) of a number will be noted by preceding the number with the string *CRC_*. For instance the CRC of a random number *RN* will be denoted as *CRC_RN*.

In order to distinguish commands specified in the C1G2 protocol specification, or commands proposed for usage in a protocol, the usage of these commands will be in bold, italic and roman characters. For instance the command issued by a reader to write data to a tag's memory will be expressed as ***Write***.

2 Related Work

The following sections provide a familiarisation with various lightweight hardware and concepts discussed in the security proposals, and also highlight the significance of these in the context of low cost RFID.

2.1 XOR Operation

The 'exclusive-or' operation (XOR) is a function that requires minimal hardware to implement. The XOR operator is both commutative and associative, and it satisfies the properties outlined in Table 1 for a Boolean variable identified by the symbol 'A'. The XOR operation will be used extensively in the following security proposals as a lightweight operation for encrypting data.

Table 1 XOR properties.

$A \oplus A = 0$
$A \oplus \overline{A} = 1$
$A \oplus 0 = A$
$A \oplus 1 = \overline{A}$

2.2 CRC Generation

RFID tags falling into the Class I and II category include CRC generation hardware utilised for error detection. A number of proposals outlined below also use the CRC generator on a tag as part of a security engine. A CRC is a type of a hash function used on fixed length bit strings to generate a message digest. It is probably important here to stress the properties of CRCs.

Mathematically, an n-bit CRC is the remainder of a division operation generated by performing a division (modulo 2) of a message bit string of length m bits by a predefined bit stream of length n ($m > n$), where the predefined bit stream is defined by a generator polynomial of degree n. In order to guarantee the n consecutive bit error detection property of a CRC, the polynomial chosen must be primitive. While any primitive polynomial of the same degree will guarantee the error detecting ability, the polynomial chosen may not provide a good hash function (a discussion of choosing a primitive polynomial is discussed in detail in [1], from pages 130–132).

CRCs by themselves can not be used for data integrity checks because of the linearity of the division process which permits changes to the data in a message string, with relative ease, without altering its CRC. This can be easily illustrated by a property of a CRC given as its Hamming Distance (HD), which is the minimum possible number of bit errors that is undetected by a message's CRC. Hence if a CRC has a HD of 4 then there are no possible combinations of 1, 2, or 3 bit errors that will pass undetected by the CRC, however there is at least one possible combination of 4 bits, such that the bit errors will not be detected by the CRC check.

CRCs are also not collision free. The collision rate of an n bit CRC is the probability that two messages will have the same CRC. The theoretical collision rate of

an n bit CRC can be given as $1/2^n$. While collisions may be a problem in error detection, they can be seen as an advantage in a security scheme as an attacker may be faced with the reconstruction of the message given only its n bit CRC.

An n bit CRC generator consists of an n-bit shift register with XOR gates that form a linear feedback function to the first stage of the register. While it is possible to use additional gates to reconfigure the CRC to produce a LFSR for generating pseudorandom numbers as part of a security mechanism, this provides no added security due to the weaknesses of LFSRs as discussed later in this Chapter. However, it may be possible to use LFSRs or NLFSRs in a context where the initialisation key and the connection polynomial are kept secret. The security of a LFSR scheme is also increased if the encryption scheme can not be subjected to a known plain text attack. However, generally it is almost always assumed that the encryption scheme can be subjected to a known plaintext attack. Nevertheless, low cost RFID provides a unique environment in which such an assumption is not always valid.

2.3 Stream Ciphers

As suggested in [2] there are four essential approaches to stream cipher design listed below.

1. Information-theoretic approach
2. System-theoretic approach
3. Complexity-theoretic approach
4. Randomised approach

Keystream generators have proliferated as a result of the information-theoretic security possible with one time pads. However, one-time pads are impractical in most implementations due to the problems associated with large purely random key sizes, key generation and distribution. Keystream generators provide a scalable means of generating a keystream that is of the same size as the plaintext. However, contrary to one time pads, the keys produced are not purely random as the sequences are pseudorandom, and hence keystream generators do not provide the same perfect secrecy as one-time pads. A detailed analysis of LFSRs and desirable properties of stream ciphers based on LFSRs can be found in [3, 4]. Prior to a discussion of LFSRs it is important to define the term linear complexity and the irreducibility of a polynomial which are important concepts in the study of stream ciphers [5].

Definition 1. Assume all sequences are binary sequences where s denotes an infinite sequence with terms $s_0, s_1,...$; s^n refers to a finite sequence of length n. Then the linear complexity of a finite binary sequence s^n denoted $LC(s^n)$ is the length of the shortest possible LFSR that can generate the sequence s^n as the first n terms [3].

Definition 2. An nth degree polynomial $f(x)$ is irreducible if no polynomial of degree k, where $0<k<n$, divides $f(x)$. There are a total of $(2^n-1)/n$ primitive polynomials for a LFSR of length n.

Since no mathematical proof of security can be found for feedback shift register based keystream generators, system theoretical designs based on established guidelines and testable security properties have been developed. Developments over time have led to a number of desirable criteria for keystream generators outlined in [3, 6–8] listed below.

- Large period before the keystream repeats itself.
- Large linear complexity.
- Good statistical properties so that the pseudorandom sequence satisfies statistical tests for randomness.
- Confusion (so that the output keystream bits are some complex transform of all or most of the bits of the key stream).
- Diffusion – use of redundant information in the output sequence.
- Meeting nonlinearity criteria to provide correlation immunity.

While the above design criteria are not proven to provide a secure keystream generator, the criteria above have been proven to be sufficient or necessary for security. There are various stream ciphers, designed around these criteria, that are based on non linear feedback shift registers, linear combination generators (the use of several LFSRs to build a single stream cipher), nonlinear filter generators and clock controlled generators which eliminate the linearity properties of the LFSRs. The shrinking generator is such a stream cipher where the generator can be implemented using simple shift registers and yet is secure provided that it is implemented prudently (discussed in more detail in Section 2.3.4).

More secure stream ciphers are built using a complex-theoretic approach (or number theoretic approach) where breaking the generator is based around an NP-hard problem. Generally these generators are more complex, slow and result in greater silicon cost for implementation [10]. However, the knapsack generator is an exception where the generator can be implemented using simple shift registers. The latter generator is discussed in more detail in Section 2.3.3.

The final approach to designing stream ciphers is based on the randomised approach where the aim is to assure security by ensuring that breaking the cipher requires an adversary to perform an impractical amount of work. The security of such a cipher can be evaluated as the average number of bits an adversary has to examine before improving his probability of ascertaining the key by way of random guesses. An example of such a cipher can be found in [2]. Generally these schemes require very large random bit sequences of the order of 10^{20} and thus require massive communication overheads, as well as computations [10].

An examination of the stream ciphers available in literature summarised in [3, 10] reveals that there is a large body of stream ciphers based on feedback shift registers despite the availability of a variety of different theoretical techniques and methods for constructing stream ciphers.

2.3.1 Linear Feedback Shift Registers

Feedback shift registers form the basic building block of a number of keystream generators. Figure 1 gives a schematic of a LFSR with L registers and a connection

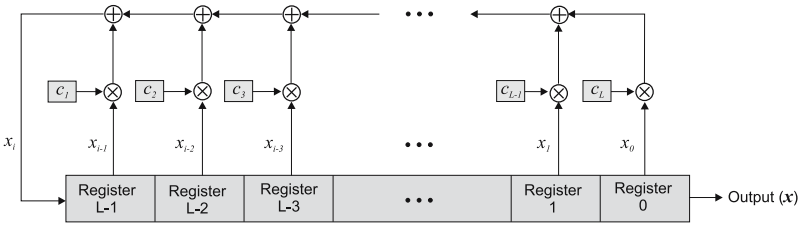


Fig. 1 Illustration of LFSR of length L where the weights of the connection polynomial are given by $c_L c_1$.

polynomial defined by the coefficients in $\boldsymbol{c} = \{c_1, c_2, \ldots, c_L\}$. LFSRs consist of a set of shift registers whose input is computed based on a linear recursion of the current state of the shift registers as shown in (1) where x is the current state of the shift register i at time t and c_i, $x_i \in$ GF(2). The unit of time t is advanced by clocking the registers after each recursion.

$$x_{L-1}(t+1) = \sum_{i=0}^{L-1} c_{L-i} x_i(t) \tag{1}$$

LFSRs form the backbone of many stream ciphers or keystream generators. They provide a number of important advantages outlined below.

- Maximum length LFSRs can produce pseudo-random sequences (sequences with statistical properties satisfying various randomness tests such a Golomb's randomness postulate) [3].
- LFSR can be easily implemented in hardware.
- Maximal length LFSRs are capable of generating sequences of length 2^{L-1} where L is the length of the LFSR.
- Since the generators are linear they can be easily analysed.

However, the output bit string of LFSRs are not secure even if the feedback scheme is not known [3]. It only takes a substring t of length $2L$ consecutive bits of an output sequence s from a LFSR to calculate the feedback scheme of the generator (commonly defined as a connection polynomial) using the Berlekamp-Massy algorithm of time order complexity $O(n^2)$ [3]. Once the connection polynomial is determined the LFSR obtained can be initialised with a substring of t with length L and used to generate the remaining bits of the sequence s.

Then the remaining question is regarding procurement of the substring t. This requires an adversary attacking the generator using a known plaintext or a chosen plaintext attack. Thus if the adversary knows the plaintext sequence m and the corresponding ciphertext sequence c, the corresponding keystream k can be obtained as m XOR c. Hence using LFSR in RFID will demand in addition to the LFSR being maximum length, that an adversary is not given the opportunity to perform a chosen plaintext attack or a known plaintext attack. The latter may be a difficult task even in a well designed system; however it is possible to prevent a known plaintext attack by a passive adversary or an active adversary who is not capable of a physical attack, in an RFID context. The mechanisms presented in this chapter illustrate the latter remark.

Based on a system-theoretic approach the most common practice of making LFSRs secure is to use a nonlinear Boolean function to generate nonlinearity in the output or the irregular clocking of LFSRs. Two generators based on the previous ideas and suitable for RFID applications are considered below. Unfortunately nothing can be proven regarding the level of security they provide. However these generators have survived much public scrutiny and they can be concluded to be computationally secure (refer to [3]). Nevertheless implementation of LFSRs must be done with care. There are a number of practical guidelines that should be followed to avoid stream ciphers based on LFSRs falling prey to adversaries. These guidelines are discussed below.

2.3.2 Implementation Considerations

In the choice of a connection polynomial it is important to choose a primitive polynomial (irreducible polynomial) that will lead to a large number of feedback connections or in other words a primitive connection polynomial of the same degree as the length of the LFSR where only a number of the coefficients are zero. This will make attacking the stream cipher more difficult [3] while ensuring that the resulting LFSR is maximum length.

The secret key is used to initialise the LFSR to provide the initial state of the registers. However it is possible to have secret connections that can be initialised using a secret key that consists of the connection polynomial coefficients and the initial state of the LFSR registers. Keeping the connection polynomial secret provides greater security [3] but with the added cost of extra hardware for implementation.

An extensive coverage of LFSR based generators can be found in [2]. The nonlinear filter generator and the clock controlled generator are two such generators based on LFSRs discussed below.

2.3.3 Nonlinear Filter Generators

Figure 2 shows a general structure of a nonlinear filter generator where the function f is a nonlinear Boolean function called the filtering function that operates on the vector $\boldsymbol{x}$ consisting of the outputs from the registers at various stages [3].

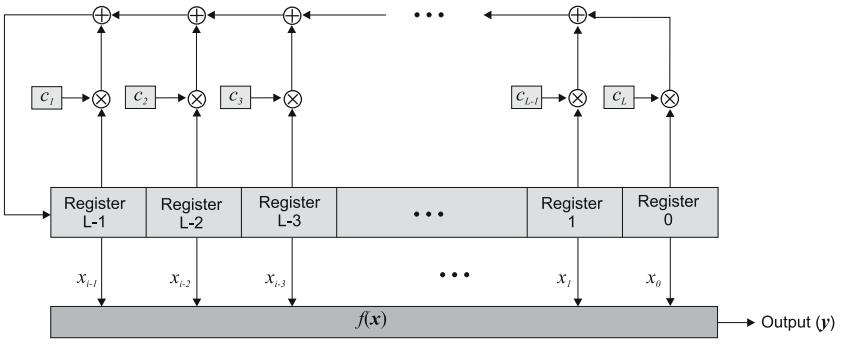


Fig. 2 Nonlinear filter generator based stream cipher.

The stream cipher generates a keystream ***y***, based on some nonlinear Boolean combination of the output of various stages of the LFSR. A stream cipher suitable for RFID applications can be found in a nonlinear filter generator called the knapsack generator. While its classification may be questioned as the security of the generator is based on an NP-hard problem, it is nevertheless based around a LFSR and a function f which acts on $\boldsymbol{x} = \{x_1, x_2, \ldots, x_L\}$ which are the outputs of each stage of the shift register and the key stream consist of selected bits from f. The knapsack generator is discussed in more detail below.

Knapsack Generator

A keystream generator based on the summation of a set of weights selected based on the register values of a LFSR to generate an integer sum S is called the knapsack generator in respect of the subset sum problem which involves the determination of a subset of weights which when added together equals a given integer S. Provided that such a subset exists the problem is proved to be NP-hard [3].

Integer addition over GF(2) is non linear as shown by Rueppel in [8]. The knapsack generator is based on mapping the state of a LFSR at time t to a knapsack sum S_t is also non-linear. The key stream is generated by calculating the knapsack sum S_t as given in (2) at time t by stepping the LFSR forward by one step, where $[x_L, \ldots, x_2, x_1]$ is the state of the LFSR registers at time t and $[a_1, \ldots, a_L]$ are the integer weights each of bit length L. Here $Q = 2^L$ where L is the length of the LFSR. The generator is illustrated in Figure 3.

$$S_t = \sum_{i=1}^{L} x_i a_i \bmod Q \tag{2}$$

While the knapsack generator has been extensively analysed by Rueppel [8] there are no published weaknesses of the knapsack generator available in literature. Unfortunately there are also no concrete suggestions on selecting the length of the LFSR or the knapsack weights.

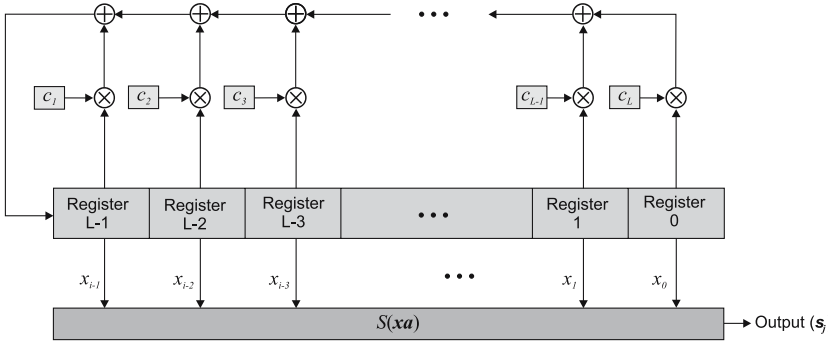


Fig. 3 Schematic of a knapsack generator.

2.3.4 Clock Controlled Generator

In general the registers in an LFSR are all clocked using the same clock signal, thus all the registers are clocked at regular times, and the contents of the registers are updated at each rise or fall of a clock signal. However, in clock controlled generators the idea is to use a combination of LFSRs so that the output of one LFSR controls the clocking of a second LFSR. This introduces nonlinearity as the second LFSR is clocked irregularly and hence the stream cipher designer hopes to defeat attacks based on the regular clocking of LFSRs. The shrinking generator proposed in [9], described below, is an example of a clock controlled generator that is suitable for implementation on an RFID label.

Shrinking Generator

This generator is a more recent proposal outlined in [9] that utilizes two LFSRs, *R1* of length *L1* and *R2* of length *L2* clocked in parallel. At any given transition of the clock the output of the generator is that of *R1* given that the output of *R2* is a logical one. If *R2* output is a logical zero then nothing is output from the generator. Hence the output from *R1* is shrunk to produce an irregularly decimated subsequence [3] as shown in Figure 4.

An outline of the properties of the shrinking generator can be found in [3]. The security of the generator has survived many known attacks on LFSR based systems, especially due to the very long period of the generator.

Attempts at breaking the shrinking generator have been based on an exhaustive search using a divide and conquer attack [12], a correlation attack [13, 14], a distinguishing attack appropriate for when *R2* has a sparse polynomial [15] and more recently using linear hybrid cellular automata to characterise the shrinking generator provided the connection polynomials, the lengths *L1* and *L2*, and as many keystream bits as the linear complexity of the shrinking generator (that is $2^{L2}.L1$ bits)[16] are known. The order of complexity of known attacks has exponential time complexity as they are a function of the length of both LFSRs. The complexity of known attacks are summarised in [9].

Despite all of the above attacks, shrinking generators are still considered resistant against efficient cryptanalysis attacks due to the difficulty of the attack scenarios and the time order complexity of the algorithms. It should however be stated here that for maximum security the following implementation consideration should be satisfied.

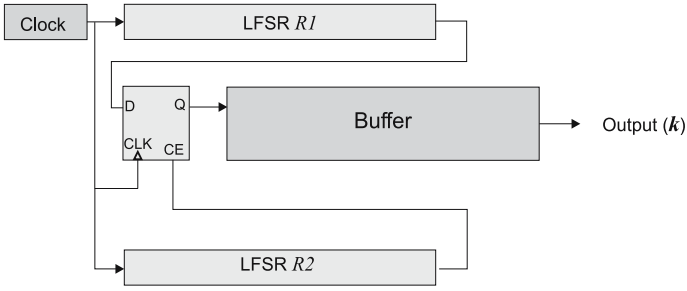


Fig. 4 Configuration of a shrinking generator.

- Use secret connection polynomials that are not sparse
- Use maximum length LFSRs for *R1* and *R2*
- The lengths of LFSR should be such that $\gcd(L_1, L_2) = 1$

One drawback of this generator is its irregular output but this may be solved by buffering the key stream prior to its use. In [17] an estimate of buffer size calculated using a Markov chain to model the buffer has shown that the probability of not having a byte of keystream data available is very small even when a buffer size of only several bits is used. Generally running *R1* and *R2* at twice the throughput required, using a small buffer size of 16 bits yields a probability of 5×10^{-3} of not having a byte of keystream data available [9]. The probability of not having a byte of keystream data decreases exponentially with buffer size and the rate at which *R1* and *R2* are running with respect to the required throughput [17].

2.3.5 Power Consumption

Power consumption of CMOS digital circuits was discussed in Chapter 6 (Addressing Insecurities and Violations of Privacy) and estimating power dissipation in modern CMOS digitial circuits was presented in Chapter 8 (An Evaluation Framework). It has been shown in Chapter 8 that the power dissipated can be estimated by considering the average capacitance switched per clock cycle. The relevant equation for estimating power consumption is reiterated below.

Equation (3) models the power dissipation of a node that may consist of a number of logic gates and it is probably most practicable when the logic circuit complexity is a minimum.

$$P = p_{0\to1} C_L V_{dd}^2 f_{clk} \quad W \tag{3}$$

In (3), C_L is the output load capacitance along the critical path and $p_{0\to1}$ is the fraction of time the node makes a power consumption transition (that is a logic $0\to1$ and $1\to0$) in a single clock cycle. The combination of $p_{0\to1}$ C_L can also be stated as the average capacitance switched during each clock cycl. In (3), f represents the clock frequency and V_{dd} is the maximum signal swing, more generally taken as the operating supply voltage.

Hence the power consumption of a LFSR based stream cipher is directly dependent on the frequency of operation, the supply voltage and the physical parameters of the gates (equivalent output capacitance). The L shift operations and XOR operations performed during each clock cycle results in a significant high switching activity and hence a considerable power dissipation. The estimation of switching activity is investigated in [18, 19] while [20] gives power estimation of a number of stream ciphers including the shrinking generator. A power consumption of 65 microW was estimated for a shrinking generator of length 65 bits with programmable connection polynomials [20]. This is comparable in magnitude to the power requirements for writing to the EEPROM memory of a passive tag.

One technique of reducing power dissipation is to reduce the equivalent output gate capacitance but this is a fabrication process based parameter that is difficult for a digital circuit designer to control. However there are low power CMOS digit-

al circuit design techniques that can be implemented to reduce the power consumption of LFSRs. There are two major principles used in low-power digital circuit design: supply voltage scaling and minimizing average switched capacitance per clock cycle [21, 22]. These techniques are briefly discussed below while a detailed coverage of the topic can be found in [21, 23]. The most effective method of reducing power dissipation is by reducing the supply voltage as the resultant reduction is proportional to the square of the supply voltage while the relationship between average switched capacitance is linear (refer to Chapter 8 – An Evaluation Framework).

2.3.5.1 Supply Voltage Scaling

It is clear from (3) that the power dissipated scales with the square of the supply voltage. Therefore a simple method of reducing power dissipation is to reduce the supply voltage. An unfortunate consequence of reducing the supply voltage is the resultant increase in the propagation delay, T_d of a gate given in (4) as a simple first order derivation which ignores the nonlinear characteristics of a CMOS gate [21].

$$T_d = \frac{kV_{dd}}{\left(V_{dd} - V_t\right)^2} \tag{4}$$

In (4), k is a parameter dependent on the size of the CMOS transistor and the fabrication process, V_t is the threshold voltage and V_{dd} is the supply voltage. Hence the supply voltage can only be reduced until the delay along the critical path (the slowest path) of the circuit is the same as the clock period required to achieve a given throughput from a stream cipher. Further power reductions may be obtained by minimising the delay of the critical path by using pipelining techniques albeit at the cost of extra hardware. The critical path of LFSRs are generally limited by the XOR summation operations [20], however for stream ciphers based on LFSRs the critical path may also be the clock control circuit as in the threshold generator or the output combining function as in the case of a non linear filter generator.

2.3.5.2 Reducing the Switched Capacitance

Minimizing the average switched capacitance per clock cycle achieved by lowering the switching activity by reducing the number of logic transitions per clock cycle also reduces the power dissipated. Therefore circuits should be clocked at the slowest rate possible and as seldom as possible to achieve the required functionality.

The average switched capacitance of LFSR circuits may be reduced by the choice of LFSR architecture, logic design style, layout style and the use of gated clocks [21, 22]. The LFSR architecture must be designed to minimise the possibility of unnecessary transitions. For instance, a XOR tree may be used in a LFSR instead of a chain of XOR gates to reduce the number of state transitions. In Figure 5 the output of the final XOR in the chained OXR structure can change three times before a valid output is available where as in the tree architecture of the XOR

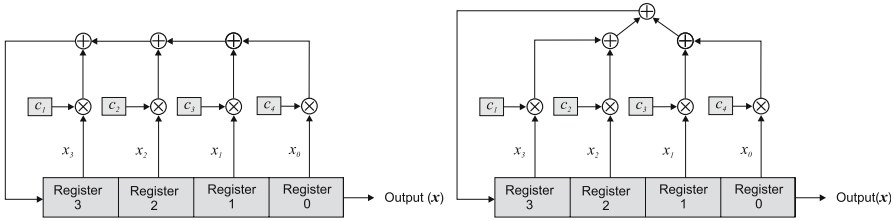


Fig. 5 Comparing a chained XOR architecture and a tree based XOR architecture.

gates, the final output will change only once while the XOR gate outputs propagate up the tree. Similarly portions of the circuit that are not in use may be powered down to attain the same goal of minimising unnecessary transitions.

As discussed previously reducing the clock rate also reduces the average switched capacitance. This can be achieved by using parallelised LFSR hardware but this is accomplished at an additional cost and therefore is not a suitable method for reducing the power consumption of stream ciphers for low cost RFID.

Finally gated clocks to prevent the clocking of portions of logic that are not immediately being used can reduce the average switched capacitance of the stream cipher. This can be done with the knapsack generator where the LFSR can be halted from being clocked while the adder circuit computes the integer sum of the knapsack weights.

2.4 Physical Unclonable Functions

The underlying concepts and the use of Physical Unclonable Functions (PUF) has already been illustrated in Chapter 17 (A Low Cost Solution to Cloning and Authentication Based on a Lightweight Primitive) of this book. The use of PUF will be further extended in this chapter to provide a confidentiality service to low cost RFID.

3 Confidentiality and Authentication

While authentication was discussed in the previous Chapter, the following sections will consider mechanism for providing a service to deliver confidentiality to low cost RFID systems. This is a difficult task as the tags and readers communicate over an insecure communication channel, where tags can not be trusted to store secret information and strong cryptographic mechanisms can not be used. Achieving confidentiality involves creating a secure communication channel in an untrusting environment over an insecure channel. This is not a novel problem (refer to [3, 4, 5]) but the solution space has little offerings for resource intensive environments.

3.1 Secure Forward Link

Achieving a secure communication channel between a tag and an interrogator using a PUF for the secure storage of a secret key and a stream cipher as a fast and efficient source of a key stream is discussed below.

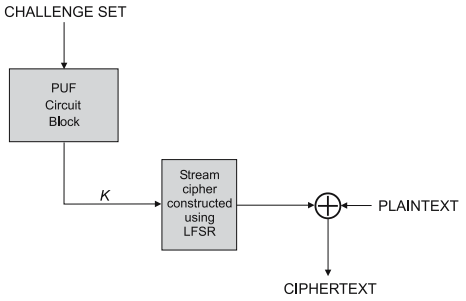


Fig. 6 Tag implementation of a stream cipher performing an encryption operation.

The method assumes that an interrogator has uniquely identified the tag using its EPC and obtained a *CHAL* list from a secure database corresponding to the tag EPC. The interrogator then initiates the establishment of a secure channel by signalling to the tag and once the tag acknowledges the interrogator's request, the interrogator transmits the *CHAL* list. The tag then uses the *CHAL* list to generate a key *K* which is used to initialise the stream cipher and to obtain a keystream *Ks* which can be used to encrypt the forward link as illustrated in Figure 6. The interrogator is able to generate the same keystream for decryption as the reader can obtain the secret key *K* corresponding to the *CHAL* set as the tag's PUF was characterised using the *CHAL* list prior to its deployment. An outline of the communication protocol is given in Figure 7.

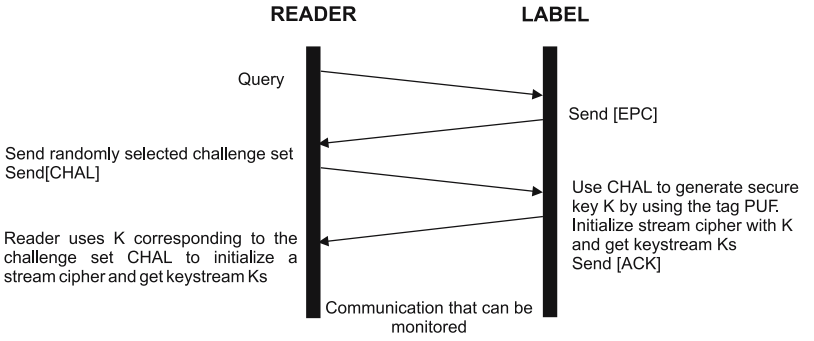


Fig. 7 Communication protocol for achieving a secure communication channel.

3.2 Tag and Reader Authentication (Mutual Authentication)

Section 3.1 above discussed a mechanism for achieving confidentiality by allowing the establishment of a secure communication channel. However, a simple extension to the communication protocol using the existing hardware required for achieving a secure channel can provide mutual authentication of a tag and a reader.

The mutual authentication algorithm is based on using the stream cipher as a means of generating a one time pad. In the method outlined in Figure 8 an interrogator initiates the authentication after the establishment of a secure communication channel. Once the tag acknowledges the request, the interrogator sends a randomly generated number *RN1* encrypted using the keystream *Ks* generated previously (indicated as Ks_1 in Figure 8). A legitimate tag is able to obtain *RN1* and then use it to reinitialize the stream cipher to generate the new keystream Ks_2. In order to make the scheme more efficient the tag uses the previously generated keystream bits Ks_1 and encrypts them with an identical length sequence of bits $Ks_2 1$ from the keystream Ks_2 generated from the reinitialised stream cipher. The interrogator is then able to authenticate the tag since only a legitimate tag would have been able to produce the keystream Ks_2.

If the tag is legitimate, the reader transmits a self encrypted keystream sequence to the tag. In the event the tag is not authentic the reader will pretend to complete the authentication by transmitting a nonce to the tag. Thus an adversary who does not have access to the reader output has no indication of whether the authentication attempt was successful.

The reader uses the previously generated keystream Ks_2 to obtain the bit sequence $Ks_2 2$ and encrypts it with the keystream $Ks_2 4$ generated from the reinitialised stream cipher to enable the tag to authenticate the reader based on the shared secret key generated using the stream cipher. $Ks_2 3$ subsequence is not used in the protocol because it is possible to use a byte of the *RN1* transmitted to skip a desired number of bits in the Ks_2 bit stream before a secret key $Ks_2 4$ is selected. This will strengthen the security of the protocol by using the idea of confusion discussed in Section 2.3.

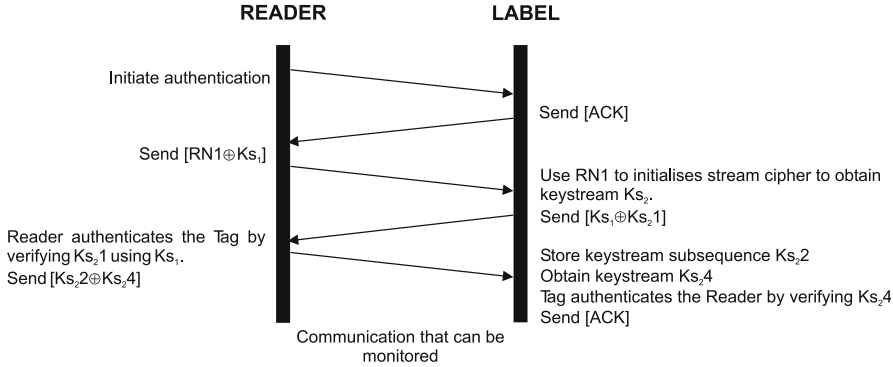


Fig. 8 Protocol for mutual authentication after establishing a secure communication channel.

3.3 Evaluation

The two schemes outlined above can be evaluated together, and the results are detailed in Table 2. Using the knapsack generator as a stream cipher provides the most cost effective solution with a total gate cost of 2416. However, performance may still be a problem and increasing performance by storing the challenges in

Table 2 Security mechanism evaluation.

Achieved Security Objectives	Confidentiality Tag authentication Reader authentication
Tag Cost	PUF block: 856 gates *Stream cipher* Shrinking generator with 64 bit LFSRs and a 16 bit buffer: 1730 gates. Knapsack generator with 64 bit LFSRs and a 64 bit adder (1 HA + 63 FA) = 1560 gates. Memory cost of storing 64, 64 bit challenges: 6144 gates.
Performance	*Considering that the challenges are stored in memory* Initialising the PUF will take an amount of time dictated by the tag memory access time for 64, 64 bit CHALs. Time to complete the transmission process using 128 bit random numbers: 2.2 ms, hence 455 tags per second can be authenticated on average. *Considering that the challenges are not stored in memory* Initializing the PUF:32 ms. Time to complete the transmission process using 128 bit random numbers: 2.2 ms. Hence 29 tags per second can be authenticated on average. *Using the stream cipher* There will be a small delay in initialising the stream ciphers but this will be in the time order of 10s of clock cycles and can be ignored. *Using a knapsack generator* On average 32, 64 bit additions need to occur prior to producing a 64 bit keystream and provided the adders are clocked with the worst case timing constraints in mind, 32 additions will take 62 clock cycles where the clock cycle length will be based on the worst case carry propagation path. However such a delay is only experienced at the initialising stage as the 64 bit keystream will buffer the time taken for the next output from the knapsack generator. Even if a 4 MHz tag oscillator is assumed the additions take negligible time and the bottle neck is still the data transmission times.
Backend Resource Requirements	Requires online access to secure database with CHAL lists or offline authentication may be performed using a local cache.
Overhead Costs	Requires creating profiles of chips prior to deployment.
Power Consumption	Power consumption is not expected to be greater than that required for writing to EEPROM memory. Refer to Section 2.3.5 for a detailed discussion on LFSR power consumption.

memory will cost an additional 6144 gates. Thus the total security engine will cost at least 8560 gates. This is still far less than that required for a heavyweight encryption scheme, delivered at a much greater performance benchmark (312 mutual authentications per second).

3.4 Practical Issues

The first instance use of LFSR is in the form of a synchronous stream cipher as the key stream is generated by initialising the LFSR with the response from the PUF and thus has the advantage that a single bit error will only cause a single bit of the plaintext to be corrupted after decryption. However, subsequent use of RNs transmitted from the tag to initialize the stream cipher may cause a failure in the authentication process due to undetected errors in the transmission from the reader.

There is the added overhead of requesting and obtaining the *CHAL* list from a secure database and transmitting the challenges to the RFID IC. This process could be avoided by storing the *CHAL* list on the tag, as the discovering the *CHAL* list by way of a physical attack can not reveal the response of a PUF circuit on an IC but as shown in Table 2 this is an expensive option.

The mechanism outlined above will require a number of proprietary commands as permitted by the C1G2 protocol. ***WriteChl*** is one such command described in Chapter 17 for sending a challenge to a PUF on a tag. However, the mutual authentication scheme will require further two commands: ***SendEnc(***<*flag*>***,***<*data*>***)*** and ***ResEnc(***<*data*>***)***. The reader command ***SendEnc*** will be used to send encrypted text to the tag, where the *flag* value will be used to denote the authentication step. The ***ResEnc*** command will be used by the tag to respond with the encrypted ciphertext to the reader. If either party detects a fraudulent tag or an unauthorised reader, the authentication process will be continued by the legitimate party, but the encrypted data will then be replaced by irrelevant bit sequences.

3.5 Possible Attacks

Similarly to the methods outlined in Chapter 17, the vulnerabilities and weaknesses of the PUF circuit also applies to the schemes described here. In addition the vulnerabilities of the stream ciphers discussed in Section 2.3.3 and Section 2.3.4 also apply to the generators. However, it should be noted that as a result of the protocol used, the stream ciphers can not be subjected to a known plaintext attack, and attacks based on a known plaintext attack can be disregarded. As discussed previously, the existing vulnerabilities of these generators can be prevented with careful implementation.

3.6 Conclusions

The combination of a PUF circuit block with a stream cipher has created a practicable and a powerful solution capable of delivering both an authentication service and an encrypted communication channel. The mechanism is suitable for both Class I and Class II tags, especially Class II tags requiring an encrypted communication channel.

The performance of the knapsack generator may be increased by using a carry save adder instead of the ripple carry adder, but this will only be achieved at almost double the cost (in terms of the gate equivalent cost) of the ripple carry adder.

4 Anonymity and Untraceability

An RFID label implementing the C1G2 protocol will scroll out its EPC when queried after being singulated by any transceiver implementing the C1G2 air interface protocol. This unique identity carried by the RFID label poses various security threats and privacy violations illuminated in Chapter 6. Thus, it is important to control access to a label's EPC, or to allow an RFID label to respond with a non-identifying response as a way of concealing its unique identity to unauthorised readers. It is possible to avoid sending a tag's EPC which allows the tag, the associated object and perhaps even the person in possession with the tag to be identified. The following sections will consider two methods to achieve anonymity and untraceability.

4.1 Pseudonyms

The schemes presented in this section are based on the idea of a tag using randomly changing pseudonyms during the identification process as a way of addressing the privacy concerns outlined in Section 9.3.7. All pseudonym schemes generally rely on two means for changing the identifier on a tag. The mechanisms are based on where the pseudonym is generated. There can be two types of pseudonym updates; reader or backend database generated pseudonym updates or tag generated pseudonym updates. The security mechanisms presented in the sections below are based on re-encryption and randomly varying object identifiers and use simple lightweight protocols for pseudonym updates generated by backend databases or readers.

4.2 Re-encryption

The idea of re-encryption was discussed in Chapter 6. Despite the resource limitations of an RFID label, it is possible to allow a low cost RFID tag to become

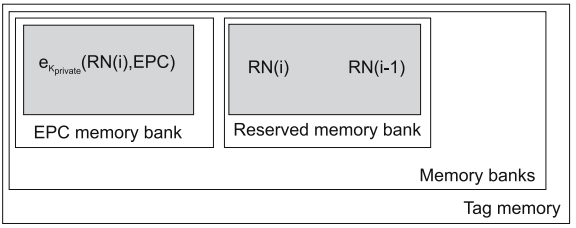


Fig. 9 Tag memory contents in a typical implementation of a Class I tag.

a party to a computationally intensive security mechanism provided that the designer is able to transfer the computationally demanding aspects to a backend system, such as the reader itself, or a backend network of computers which may act as a proxy to the security mechanism. In essence it is a transfer of complexity out of the chip onto a shared resource with greater capability to execute a computationally intensive task in a timely manner. While this will essentially reduce the ability to perform offline operations it does allow tag cost to be kept to a minimum. Clearly pushing complexity further up the EPC Network reduces the ability to execute security services offline; an unfortunate compromise that must be made in view of tag cost constraints.

The following security mechanism is based in the idea of transferring complexity out of the tag silicon, and the ideas behind re-encryption and a method used for establishing a secure communication channel outlined in Section 4. Re-encryption offers a novel perspective on achieving the privacy objectives of anonymity and location privacy. The scheme is described below.

Instead of storing the EPC in a write once format, it is possible, for instance, for a retailer in the supply chain to store an encrypted version of a tag EPC XORed with a nonce or a random number *RN(i)* on the tag where the EPC is now [EPC ⊕ *RN(i)*]. The XORing of the EPC ensures that the encrypted EPC appears to be different each time the encrypted version of the EPC is updated on the tag.

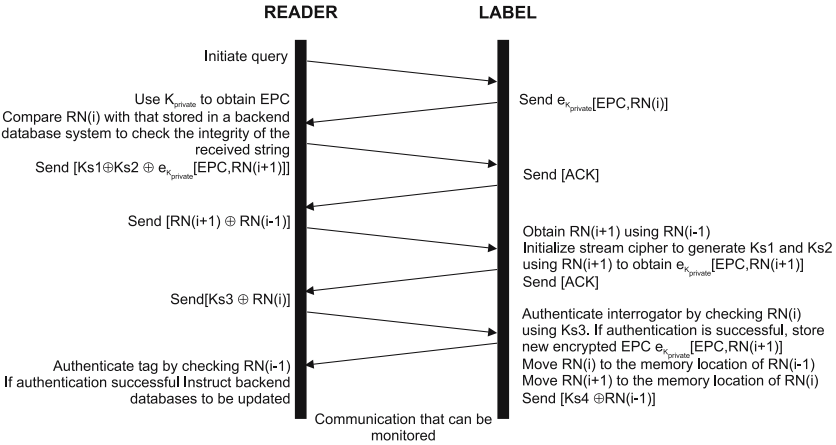


Fig. 10 Communication protocol for the re-encryption scheme.

However for simplicity of the following discussion the [EPC ⊕ *RN(i)*] operation is assumed to be implicit in the mention of an EPC.

The scheme also requires the storage and the transmission of a random number *RN(i)* so that the actual EPC can be obtained from [EPC ⊕ *RN(i)*] and an initial random number *RN(i-1)* on the tag as illustrated in Figure 9. The primary storage of the *RN*s needs to be performed in a secure environment such as an electromagnetically shielded room or a Faraday cage. Then it is possible for the tag owners to use their own private keys to encrypt the tag identifiers prior to their use within a facility such as at a supermarket store.

The encryption may be performed using a secret key that is known solely to the retailer or the person having ownership of the tags. Thus, when a label is requested to scroll out its EPC, it will scroll out an encrypted version of the EPC, which to a third party will appear as a stream of random bits. The protocol is illustrated in Figure 10.

The encrypted EPC alone will prevent profiling as the information obtained by eavesdropping does not reveal any information regarding the object to which the tag is attached as an adversary is unable to obtain the EPC without the private key $K_{private}$. However, the label still emits a predictable response and thus untraceability is not achieved.

On identification of the label, the reader has the option of storing back an encrypted version of the EPC padded with a new random number *RN(i+1)* used to perform the [EPC ⊕ *RN(i+1)*] operation. The protocol is outlined in Figure 10. Hence authorised readers may randomly, or on every occasion the tag is read, update the encrypted version of the EPC by storing a new encrypted version of the EPC.

The above mechanism has the added advantage that the *RN(i)* padded to the EPC can be used as a message authentication code to ensure the integrity of the received tag response to a query by comparing the *RN(i)* obtained after decrypting the transmission $e_{Kprivate}[EPC, RN(i)]$ and also to authenticate the tag as being legitimate. It should be noted here that it is possible for the encrypted data transmitted from the tag to contain product specific information such as an EPAC that will extend the tag authentication scheme to a product authentication scheme. The use of EPAC is discussed in Section 5.

Alternatively, it is possible to use the protocol in Section 3.2 after the tag has transmitted its encrypted tag identifier. Then the $e_{Kprivate}[EPC, RN(i+1)]$ can be transmitted from the reader to the tag using the encrypted communication channel established after a mutual authentication process. The difference between the protocols in Figure 7 and in Figure 8 and that given in Figure 10 is the removal of the overhead of using a PUF to initialize the stream cipher.

An attractive feature of the above approach is that it can be used throughout the supply chain at various stages by setting up a tag data security infrastructure based around a PKI (public key infrastructure) to allow the benefits of tag or product authentications, anonymity and untraceability to be available, at any time and anywhere along the supply chain all the way into the living quarters of consumers. Implementing such an infrastructure will require careful standards and agreements between various parties in the supply chain. The idea is illustrated in Figure 11. Considering a simple supply chain model, the manufacturer would use the public key of a distributor to encrypt the EPC prior to the transfer of goods to the

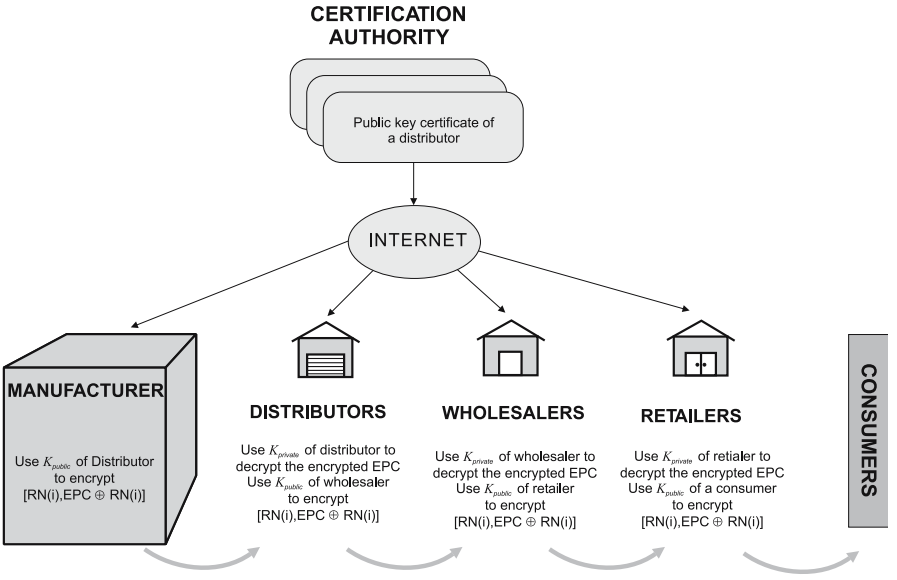


Fig. 11 Tag data security infrastructure based on a PKI.

supplier. The public key can be obtained from a certificate authority, with which the distributor's public key is published. The distributor is then able to decrypt the encrypted EPC using the distributor's private key and prior to transportation of the goods to a retailer, the EPCs of the products going to a particular retailer can be encrypted using that particular retailer's public key. It is still possible for a party along the supply chain or even a consumer to then use their own private key to encrypt the EPC for both greater efficiency and greater security while the tagged item is in their ownership.

4.2.1 Evaluation

The scheme outlined above is evaluated and discussed in detail in Table 3 Re-encryption, while at the cost of online requirements and tag initialisation requirements is capable of delivering adequate performance, using strong cryptographic primitives with minimal cost consequences to a tag. Unlike the mechanisms discussed previously, the present mechanism is able to satisfy a majority of the security objectives and all of the privacy objectives considered necessary in Chapter 8.

The total implementation cost of the mechanism (using a knapsack generator) on an IC is estimated to be 2328 gates, however this cost might be higher in practice due to additional registers and the changes required to the finite state machine to implement the new commands necessary to execute the protocol and recover from a sudden power loss.

Table 3 Security mechanism evaluation.

Achieved Security Objectives	Confidentiality Tag authentication Reader authentication Message content security
Achieved Privacy Objectives	Anonymity Untraceability
Tag Cost	*Gate equivalent cost estimate based on the memory storage required for the keys* C1G2 tags already have the necessary hardware for XOR operations, string comparisons, CRC generation, registers for temporary storage of operands, and memory for the storage of the EPC. Using a 96 bit EPC, a 128 bit RN value (adequate for initialising two 64 bit LFSRs, as in the cases of a shrinking generator and also adequate for initialising both the initial state and the connection polynomials of a knapsack generator of length 64 bits) with a 128 bit private key where the encrypted message will be the same size as the key size (as is the case with the AES block cipher with a 128 bit key). Memory cost for the encrypted EPC: 384 gates Memory cost for *RN(i-1)* and *RN(i)*: 384 gates Cost of a shrinking generator: 1730 gates Cost of a knapsack generator: 1560 gates
Performance	Neglecting network delays and computation time for the new encrypted version of a tag's EPC, the greatest delays will result from the time required for transmitting data between tags and readers. There will also be a small delay in initialising the stream ciphers but this will be in the time order of 10s of clock cycles and can be ignored. Estimated time to compete the protocol: approximately 5 ms Hence the number of tags that can be read, authenticated and pseudonyms updated: 200 tags This is a best case scenario and in reality the string comparisons and the calculation of the encrypted EPC will reduce the estimated performance.
Backend Resource Requirements	Real time authentication requires access to secure backend databases with *RN* values. However if real time authentication is not required and all that is required is a pseudonym change, where backend databases can be consulted at a later time for both the update procedure and the authentication process, the protocol can be executed without the need for online resources.
Overhead Costs	Tags must be subjected to an initialisation phase prior to deployment in an electromagnetically secure environment for the initial storage of the *RN* values. However this will be a one-time cost and can be carried out at the RFID chip verification phase.
Power Consumption	The most power consuming operation is the operation required to write two strings to the EEPROM and thus the mechanism will not violate power constraints outlined in Chapter 8 (An Evaluation Framework). Refer to Section 2.3.5 for a detailed discussion on LFSR power consumption.

4.2.2 Practical Issues

The implementation of the scheme outlined in Figure 11 requires the creation of a PKI whereby parties can publish their public keys signed by a trusted third party. A simple system overview of a manufacturer in possession of a list of public keys associated with a product's destination party and a retailer using his own symmetric key encryption scheme is illustrated in Figure 12.

During the journey of a product through the supply chain where a party along the supply chain is in ownership of that product, that party then has the option of using a symmetric key primitive to encrypt the EPCs to both improve performance by speeding up the encryption process and to utilise the increased security provided by symmetric keys in relation to public keys of similar size.

It has also not been discussed whether the present mechanisms can be accommodated by the recently ratified C1G2 protocol. Tags will initially need to be placed in a locked state where the query command will initiate the execution of the protocol outlined, or a modified version of the existing query command is required to signal the finite state machine to execute the protocol in Figure 5. The mechanism outlined above will also require two proprietary commands as permitted by the C1G2 protocol: ***SendEnc(***<*flag*>***,***<*data*>***)*** and ***ResEnc(***<*data*>***)***. The reader command ***SendEnc*** will be used to send encrypted text to the tag, where the *flag* value will be used to denote the protocol step. The ***ResEnc*** command will be used by the tag to respond with the encrypted ciphertext to the reader. If either party detects a fraudulent tag or an unauthorised reader, the protocol will be continued

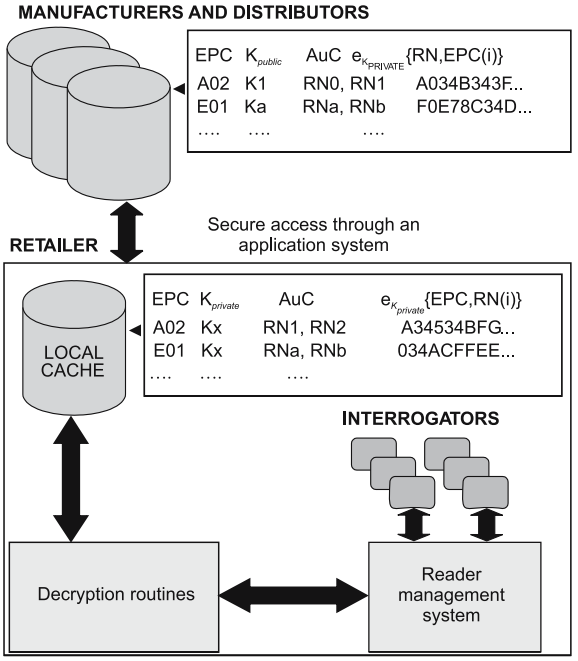


Fig. 12 An overview of an implementation of an RFID system based on re-encryption.

by the legitimate party, but the encrypted data will then be replaced by irrelevant bit sequences. The commands mentioned above were also used and described in Section 3.4.

4.2.3 Possible Attacks

The security of the above mechanism relies on the difficulty of predicting the output of the stream cipher given only the ciphertext. Both the knapsack generator and the shrinking generator have been found to be secure against known ciphertext only attacks, even if the connection polynomials are known. The stream ciphers used have been discussed in detail in Section 2.3.

Figure 13 provides a protocol verification schematic outlining the execution of the protocol over three consecutive interrogations of the same tag. At each stage the tag memory contents and also the possible information that can be obtained by a passive eavesdropper that can eavesdrop on both the forward and the backward communication channel is indicated to show that the information collected is not sufficient to defeat the security mechanism.

The information that is available to an eavesdropper provides an insight into looking at the weaknesses of the protocol to a passive or a malicious adversary who may be mobile or stationary. Clearly the protocol may be interfered with by a disruptive adversary conducting a man-in-the middle attack but this is a difficult proposition given the difficulty of capturing and altering the messages between

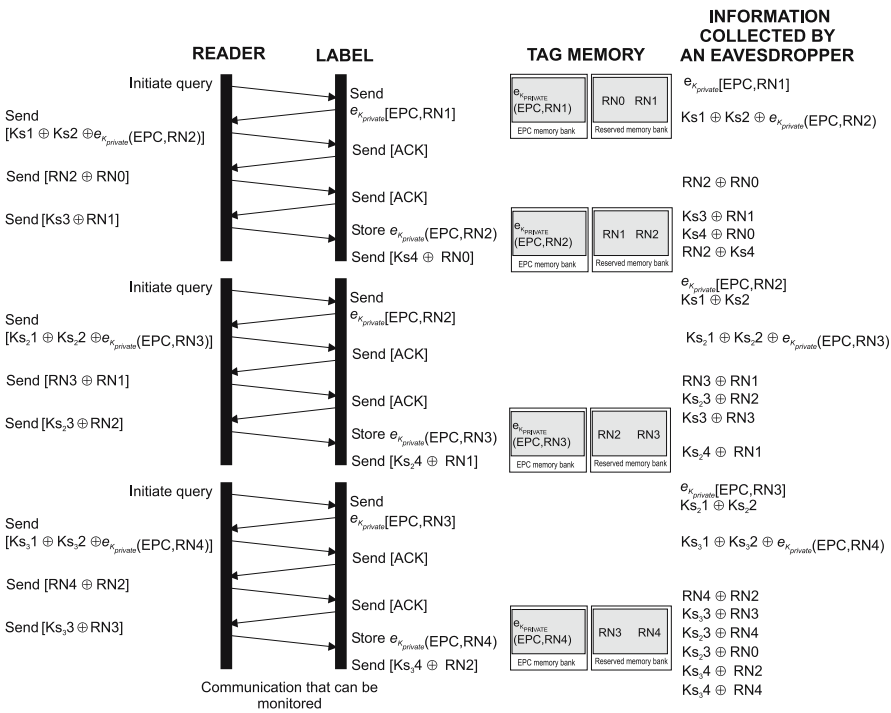


Fig. 13 Verification of the re-encryption protocol.

a tag and reader during the rapid transition of messages and the adversary would also have to ensure that either the tag or the reader never receives the unaltered message.

An adversary with access to specialised equipment may subject a tag to a physical attack to obtain the *RN* values. This will only allow a single tag to be replaced with another tag, with identical memory contents but such an attack provides no useful information to defeating the security mechanism of other tags. However, a proximity mobile adversary will then be able to decipher all future communications between a tag and a reader. Preventing a physical attack requires the secure storage of keys on the tag as discussed in the protocol presented in Section 3.

The use of forward secrecy, that is the use of ephemeral keys, *Ks*, which are generated during each communication session, to be used only once, invalidates any information that the adversary is able to obtain about previous keys or any other information. Even if an adversary is able to recover a previous keystream, that is a Ks_i, that information is not useful in the next interrogation session. The security of the scheme relies on the ability of the label owners to keep the encryption key secret and thus the mechanism is vulnerable to an adversary who is a temporary or a permanent insider.

4.3 Randomly Varying Object Identifiers

An alternative to the re-encryption scheme outlined above for providing privacy protection and authentication, is a scheme based on the concept of using a pool of completely random EPCs. Here the EPC number no longer has an information bearing structure and it is a random number which only acts as a temporary pointer to the actual EPC. The relationship created between the random EPC and the true EPC is securely stored in a backend database. Thus, it is possible for RFID labeled items of sensitive nature (such that required for use in supply chain logistic operations of the defence department) to be labelled with a randomly generated EPC.

The scheme requires the backend systems to manage a large pool of random numbers and to be able to search through such a large pool in a short time. This database will also need to be able to perform concurrent updates and searches while possibly being distributed in nature.

It is possible to use an encrypted version of the random EPCs as done in the re-encryption scheme outlined in Section 4.2 to add another layer of security. Although using a random EPC will prevent an adversary from obtaining any useful information in the event the encryption scheme is compromised, encrypting the EPC data will be at the cost of performing the encrypting and decrypting operations, on the reader or by way of a trusted third party (proxy). It should also be mentioned here that it is possible for an encrypted version of the random EPC to contain product specific information such as an EPAC (Electronic Product Authentication Code) that will extend the tag authentication scheme to a product authentication scheme. The use of EPACs are discussed in Section 5.

However, the use of random EPCs eliminates the need to encrypt the EPC to hide the information bearing nature of the EPC while preventing the tag from em-

anating a predictable response to a query by an unauthorised reader. The protocol is detailed in Figure 14 and is similar to that given in Figure 10.

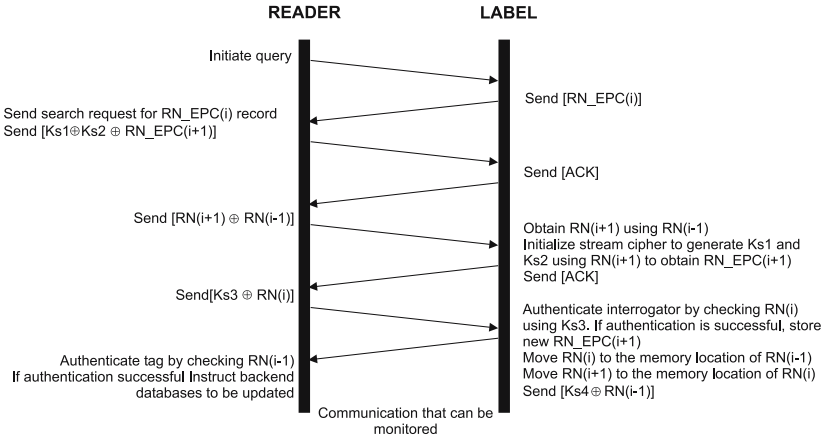


Fig. 14 Protocol based on using randomly varying object identifiers.

4.3.1 Evaluation

The randomly varying object identifier scheme outlined above is evaluated for its suitability for low cost RFID in Table 3. Similarly to the mechanism discussed in Section 4.2 the present mechanism is able to satisfy a majority of the security objectives and all of the privacy objectives considered necessary in Chapter 6 and outlined in the framework provided in Chapter 8.

An a consequence of transferring tag complexity to backend systems, with a complex data structure capable of concurrent updates and efficient search algorithms, the randomly varying object identification techniques has reduced the security related tag costs. In comparison with the re-encryption protocol in Section 4.2 requiring from 2328 gates for the security engine the current scheme requires only 1944 gates. This reduction is achieved by removing complexity and the associated overhead resulting from encrypting the EPC related information.

4.3.2 Practical Issues

The scheme outlined above requires, as indicated in Figure 15, rapid access to databases over network infrastructure and secure databases maintaining records for each individual EPC. The management of such a database is not a significant hurdle. A data structure such as a self-balancing binary search tree implementation in the form of a red-black binary tree [25] with the ability to perform concurrent updates and search operations will perform efficient search, insert and delete operations. Infrastructural issues are important but can generally be addressed with greater investment to build networks of adequate bandwidth to mange the network delays. Currently Internet traffic network delays are in the rage of 10s of milliseconds [26] but these delays can be improved using appropriate infrastructure.

Table 4 Evaluation of the randomly varying object identification scheme.

Achieved Security Objectives	Confidentiality Tag authentication Reader authentication Message content security
Achieved Privacy Objectives	Anonymity Untraceability
Tag Cost	*Gate equivalent cost estimate based on the memory storage required for the keys* C1G2 tags already have the necessary hardware for XOR operations, string comparisons, CRC generation, registers for temporary storage of variables, and memory for the storage of the EPC. *Assume:* Using a 96 bit RN_EPC and a 128 bit RN value (adequate for initialising two 64 bit LFSRs, as in the cases of a shrinking generator and also adequate for initialising both the initial state and the connection polynomials of a knapsack generator of length 64 bits). Memory cost for RN(i-1) and RN(i): 384 gates Cost of a shrinking generator: 1730 gates Cost of a knapsack generator: 1560 gates
Performance	Neglecting network delays and computation time for new encrypted versions of a tag's EPC, the greatest delays will result from the time required for transmitting data between tags and readers. The small delay in initialising the stream ciphers in the time order of 10s of clock cycles can be ignored. Estimated time to compete the protocol: approximately 3.5 ms Hence the number of tags that can be read, authenticated and pseudonyms updated: 285 This is a best case scenario and in reality the string comparisons and the calculation of the encrypted EPC will reduce the estimated performance.
Backend Resource Requirements	Real time authentication requires access to secure backend databases with *RN_EPC* values. If real time authentication is not required and all that is required is a pseudonym change, where back end databases can be consulted at a later time for both the update and the authentication process, then local databases caching a list of available random EPCs for future use, will greatly speed up the protocol and provide a method of offline authentication. Refer to Section 4.3.2 for database cost considerations.
Overhead Costs	Tags must be subjected to an initialisation phase prior to deployment in an electromagnetically secure environment for the initial storage of the *RN* values.
Power Consumption	The most power consuming operation is the operation required to write two strings to the EEPROM and thus the mechanism will not violate power constraints outlined in Chapter 8 (An Evaluation Framework). Refer to Section 2.3.5 for a detailed discussion on LFSR power consumption.

TPC (Transactions Processing Performance Council [27]) data provide a guide to assessing performance and cost criteria to evaluate the feasibility of a system

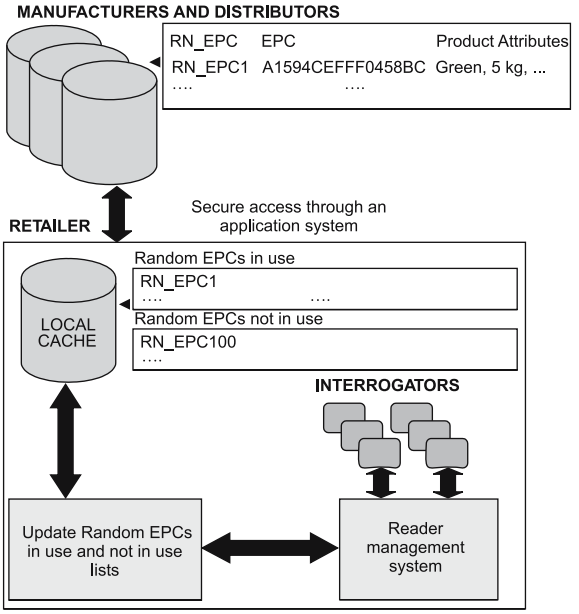


Fig. 15 An overview an RFID system based on randomly varying object identifiers.

implementation. TPC is a non-profit organisation founded to define transaction processing and database benchmarks. The term transactions has a broad meaning, however the TPC benchmarks define a transaction as it is referred to commonly in the business world. Thus a typical transaction defined by the TPC will encompass updating a database system for the purpose of inventory control, banking or the purchase of goods. TPC benchmarks generally measure transaction processing and database performance in terms of the number of transactions a given system and database can perform as transactions per second (tpc) or transactions per minute (tpm) TPC pricing and performance metrics are widely used by the industry to estimate IT infrastructure cost required to provide adequate system performance [28].

The TPC-C performance benchmark [29] for OLTP (on-line transaction processing) is a suitable benchmark to establish the cost of installing a clustered or a non clustered server capable of delivering the database processing times required to make randomly varying object identification practicable. The price performance number obtained for a non clustered system configuration capable of over 4 million transactions per minute, or 15 microseconds per transaction, has a cost estimate of is approximately 15 million Australian dollars. The latter cost estimate comprises of the cost of establishing and maintaining a system, including the cost of software, backup storage and three years of maintenance.

While complexity is pushed further back up the EPC Network to reduce the cost of tags and readers, a significant investment will be required to establish and run the required backend systems. However this is not an ongoing cost and a significant portion of the cost will be one-time items (such as network infrastructure costs). In addition, usage over time (that is an accumulating number of trans-

actions over the life time of the system), will reduce the cost per transaction to a level much less than that estimated in the TPC-C benchmark figures.

Reading a tag is clearly no more complicated than current implementations and requires no special commands. However a special query command or a modified version of the query command is required to signal the finite state machine to execute the protocol in Figure 14. The execution of the protocol following a query can be performed using the two additional commands outlined in Section 4.2.2.

4.3.3 Possible Attacks

The use of random tag identifiers provides anonymity by altering the tag response to a query command and thus never transmitting a predictable response. Figure 16 provides an outline of the execution of the protocol to query a tag on three consecutive occasions. The tag memory contents at the start and the end of the protocol are also indicated. The vulnerabilities of the system are identical to those discussed previously in Section 4.2 under the re-encryption based mechanism. Clearly given a mobile adversary who may be passive or malicious and who is able to collect all the information indicated in Figure 16, still has the task of breaking the stream cipher given only the ciphertext.

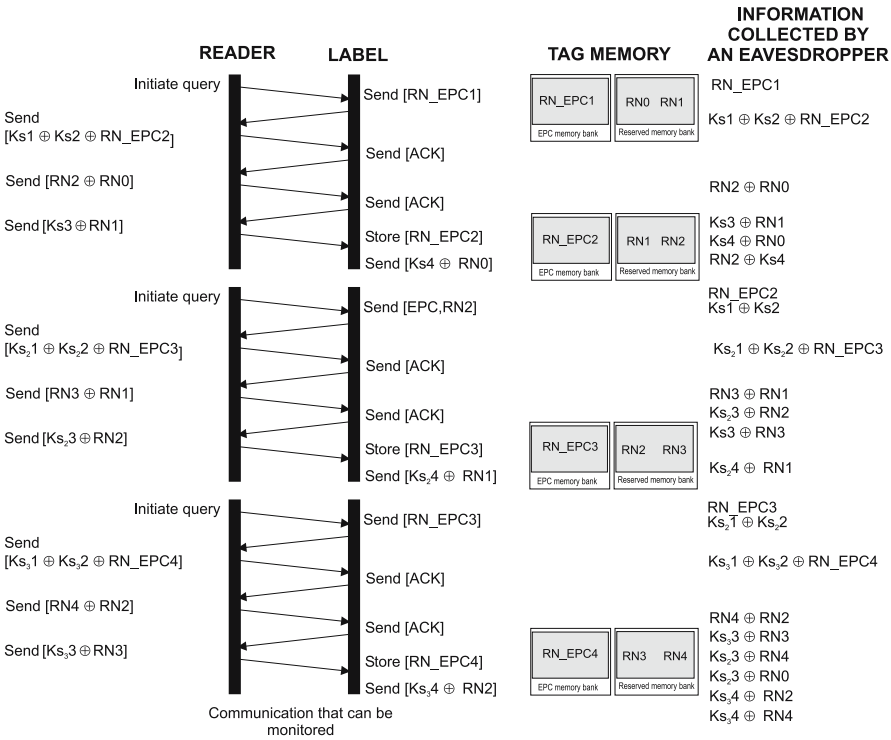


Fig. 16 Verification of the randomly varying object identifier protocol.

5 Anonymity, Untraceability, Product Authentication and Preventing Counterfeiting

Section 4 outlined a scheme for providing anonymity and untraceability, while Section 2 outlined methods of authentication. The following scheme is aimed at combining the previous solutions to provide, in addition, a product authentication service that will provide a method for detecting counterfeit goods.

Authentication implies the establishment of a tag's legitimacy. However, in a supply chain logistics environment, use of authentication services to establish the authenticity of a tag and hence the legitimacy of the article to which it is attached is not sufficient to guarantee the genuineness of the article; though authentication of a tag does eliminate cloning of tags. The absence of a method to ascertain the genuineness of goods may be a special concern in the secondary market for goods, and in the processing of returned items.

5.1 Product Authentication and Anti-Counterfeiting

The absence of a method to physically or electronically bind the tag identity to a product identity in existing RFID deployments implies that an authentication of a tag does not necessarily guarantee the authenticity of the object to which the tag is attached. The genuine article may be replaced with a counterfeited article or counterfeited goods fitted with stolen tags.

There are a variety of existing techniques for product authentication, based on optical technologies (watermarks, holograms, micro printing and biochemical technology [30]). These techniques are not without their list of associated problems. All solutions devised from the technologies above are based on static markers that are typically applied uniformly to a single class of products. Biochemical marker tests provide the ability to detect markers but they do not generally quantify the marker, thus leaving open avenues of counterfeiting by dilution, while most optical technologies no longer present an adequate deterrent due to the reduction in the cost of producing watermarks and holograms.

It is into this environment that the following proposal introduces an electronic maker. Each tag attached to a product will contain an Electronic Product Authentication Code (EPAC) illustrated in Figure 17 and the various data fields are explained below.

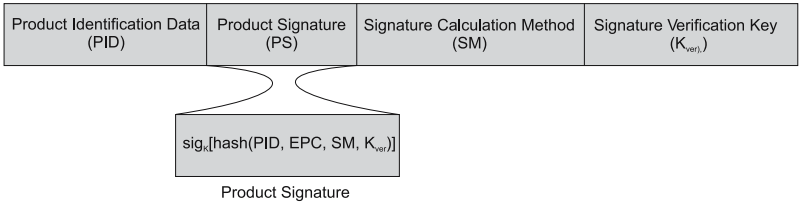


Fig. 17 Electronic Product Authentication Code (EPAC).

Product Identification Data (PID)
It is unique product identification data that characterises the product using verifiable, measurable or observable product specific information determined by the product manufacturer or the retailer. The product identification data may consist of, for instance, weight of the product, measurable physical characteristics such as the dielectric constant, or conductance, size and shape of the product or an electronic copy of a signature printed or embossed on the product.

Product Signature (PS)
The product signature allows the confirmation of the integrity of the EPAC and also allows a third party to authenticate the identity of the signatory (such as a manufacture or a retailer). As illustrated in Figure 17 the signature value is calculated by hashing the PID, EPC, SM, and K_{ver}. This will allow the creation of a message digest to reduce the size of the bit string that needs to be transmitted as the product signature. Hashing the data will produce savings in transmission time as well as memory storage costs. Various signature methods such as RSA-PSS, DSA, ECDSA and ElGamal signature scheme can be used and unkeyed hash functions such as SHA-1 (produces a 160 bit hash value) or MD5 (produces a 128 bit hash value) can be used [3, 24].

Signature Calculation Method (SM)
The binary sequence in this field will indicate the digital signature method and the type of hash function used to calculate the message digest.

Signature Verification Key (K_{ver})
This is the public key of the party that produced the product signature. The signature verification key can be used to verify the product signature.

Assuming that the product signing takes place at the manufacturer, any strong cryptographic signing algorithm with a reasonable digital signature size may be used for the process where the signing method used can be indicated as part of the EPAC along with the verification key, K_{ver}. The protocol for product authentication is illustrated in Figure 18.

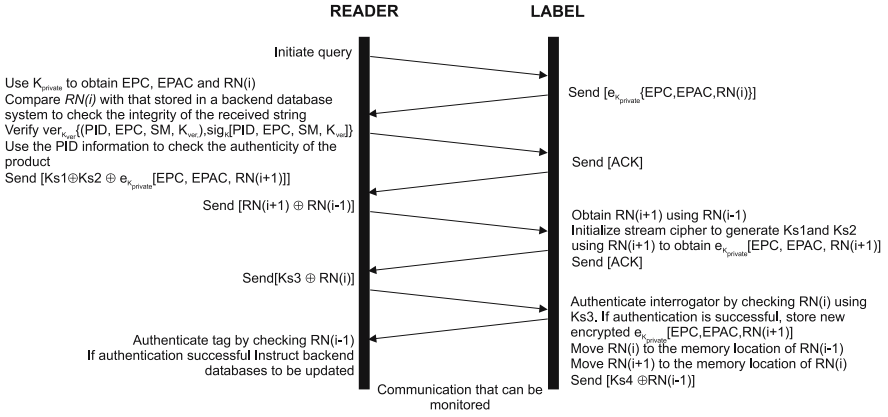


Fig. 18 Protocol for product authentication.

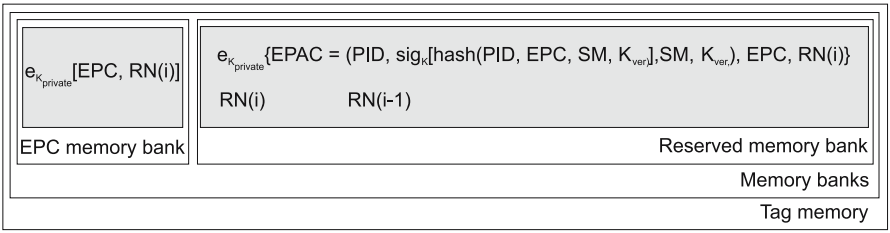


Fig. 19 Tag memory contents with EPAC information.

Similarly to the re-encryption scheme in Section 4.2 the EPC stored in the tag memory is a result of the [EPC $\oplus$ *RN(i)*], and *RN(i)* is transmitted along with the EPC. This ensures a random variation in the tag identifier. However for simplicity of the following discussion the [EPC $\oplus$ *RN(i)*] operation is assumed to be implicit in the mention of an EPC. Once a tag transmits its tag identifier as indicated in Figure 18, a reader can decrypt the received information to obtain the tag EPC, EPAC and the *RN(i)* used in the XORing operation of the tags actual EPC. Although not mentioned previously, the EPC memory bank can also contain an encrypted version of the EPC without the EPAC data, as shown in Figure 19. This will allow a reader who does not wish to authenticate the product to execute the protocol based on the re-encryption scheme outlined in Figure 10 and if the EPAC is required, the reader can request the tag to transmit the EPAC data over the encrypted channel.

In order to provide a complete solution to product authentication and anti-counterfeiting of goods, it then essential for the client application to check the electronic pedigree of the item to ensure that a verifiable and a valid path for the product's life through the supply chain exists from the manufacture to the point of sale of that item. The idea behind an electronic pedigree is discussed in Chapter 4 and Chapter 11 and it is not considered in this chapter. The electronic pedigree and the EPAC information can then be used both to authenticate the product and to thereby detect counterfeit goods. Thus for a product to be authentic the EPAC information must be verifiable and the product must have a verifiable and a valid path through the supply chain. A failure of either test implies the detection of a counterfeited item.

5.2 Evaluation

Table 5 provides an evaluation of the product authentication mechanism discussed above. Examination of Table 5 reveals that there is a significant penalty in hardware costs (storing a product authentication code costs 1536 gates) and time required for transmitting the 962 bit long string from the tag to the reader and from the reader to the tag. The total hardware cost is less than 4000 gate equivalents. However, there is a serious performance limitation. This performance limitation may not be a hindrance in practice as product authentication may require a lengthy measurement or visual examination process.

Table 5 Evaluation of the product authentication protocol.

Achieved Security Objectives	Confidentiality Tag authentication Reader authentication Product authentication and counterfeit item detection Message content security
Achieved Privacy Objectives	Anonymity Untraceability
Tag Cost	*Gate equivalent cost estimate assumptions:* A 96 bit EPC, a 128 bit RN(i) value (adequate for initialising two 64 bit LFSRs, as in the cases of a shrinking generator and also adequate for initialising both the initial state and the connection polynomials of a knapsack generator of length 64 bits) A 128 bit private key for encrypting the EPC and EPAC. A 256 bit PID, and assuming that ECDSA is used to create the digital signature with a key size of 160 bits as recommended in FIPS 186-2 [11] which generates a digital signature of size 320 bits using the SHA-1 hash algorithm, which produces a 160 bit message digest. An ACM of 2 bits Verification key of size 160 bits and RN(i) of 128 bits Memory cost for encrypted EPC: 384 gates Memory cost for encrypted EPC and EPAC: 1536 gates Memory cost for RN(i-1) and RN(i): 384 gates Cost of a shrinking generator: 1730 gates Cost of a knapsack generator: 1560 gates Total cost (using only the encrypted EPC and EPAC along with a knapsack generator): 3288 gates
Performance	Neglecting network delays and computation time for new encrypted versions of a tag's EPC and EPAC, the greatest delays will result from the time required for transmitting the 962 bit long string consisting of the EPAC between tags and readers. There will also be a small delay in initialising the stream cipher but this will be in the time order of 10s of clock cycles and can be ignored. Estimated time to complete the protocol: approximately 11 ms Hence the number of tags that can be read, authenticated and pseudonyms updated: 88 This is a best case scenario and in reality the string comparisons and the calculation of the encrypted EPC on the reader side will reduce the estimated performance.
Backend Resource Requirements	Real time authentication requires access to secure backend databases with RN values. However if real time authentication is not required and all that is required is a pseudonym change, where back end databases can be consulted at a later time for both the update procedure and the authentication process, the protocol can be executed without the need for online resources.
Overhead Costs	Tags are required to undergo an initialisation phase prior to deployment in an electromagnetically secure environment where the initial RN(i) values can be imprinted in memory.
Power Consumption	The most power consuming operation is the operation required to write two strings to the EEPROM and thus the mechanism will not violate power constraints outlined in Chapter 8 (An Evaluation Framework). Refer to Section 2.3.5 for a detailed discussion on LFSR power consumption.

5.3 Practical Issues

Storing the EPAC on the tag is an expensive option both in terms of memory storage costs and transmission costs, but it does allow the offline authentication of the product. Alternatively it is possible to store the EPAC on a backend database, along with other product related data pointed to by the EPC. This will reduce tag complexity and reduce bottlenecks produced by long transmission times accumulated during the product authentication protocol.

It is also possible to store a portion of the EPAC such as the PID on the tag, and the rest of the data on a secure database pointed to by the tag EPC. This will greatly reduce product authentication times by reducing the data load transmitted during the protocol. If the PID data needs to be verified, it can be achieved by opting to retrieve the remaining EPAC data from the secure database.

As discussed in Section 4.2.2 tags will initially need to be placed in a locked state where the query command will initiate the execution of the protocol outlined or a modified version of the existing query command is required to signal the tag's finite state machine to execute the protocol in Figure 18. A modified query command can also signal if the current interrogation round requires the tag to participate in a product authentication round or if the query will only result in the update of the tag identifier based on the encrypted version of the EPC stored in the EPC memory bank shown in Figure 19. The mechanism outlined above will also require two proprietary commands as discussed in Section 4.2.3.

5.4 Possible Attacks

The use of re-encryption provides anonymity by altering the tag identifier using a randomly generated number. Thus tags never transmit a predictable response. Figure 16 provides an outline of the execution of the re-encryption protocol, without the added product identification service, to query a tag on three consecutive occasions. The vulnerabilities of the system are identical to those discussed previously in Section 4.2 under the re-encryption based mechanism.

6 Conclusions

Strong cryptographic solutions are too area or power hungry to satisfy the limitations of RFID systems and much of the encryption hardware available is for smart card technology. Even though the solutions can be applied directly to RFID, the main obstacle is that smart card processors are much more powerful than a typical RFID label. Thus, the solutions are not portable to an RFID platform if we expect the cost of the secure labels to remain below the 5 cents target value. The chapter has used lightweight hardware and lightweight protocols to both enforce privacy and to provide security services to address various vulnerabilities identified in

Chapter 6 and has presented the application of low cost RFID technology to security related applications such as anti-counterfeiting.

The solutions presented have recognised that the resource limitation of low cost labels require the consideration of simplicity at the tag silicon level provided by small one time pads, which involve one or more small shared secrets between a label and an interrogator. Such methods required the use of shielded electromagnetic communications between the label and the reader system to store secret information at an initialisation phase.

The solutions presented have concentrated on the simple concepts of removing label IC complexity, and using the abundant resources available to the reader and application systems of an RFID system to counterbalance the resource limited nature of RFID labels.

Security mechanisms discussed overcome privacy concerns by addressing profiling and, tracking and surveillance. However, it should be noted here that issues concerning privacy are also public policy issues and require a combination of security mechanisms and properly formulated public policy.

The security mechanisms presented has been evaluated using the criteria outlined in Chapter 8 to appraise their suitability for low cost RFID applications.

It is evident that RFID privacy and security are challenging areas of research that have led to a blossoming new cryptographic paradigm called lightweight cryptography. There are three specific areas of research (lightweight hardware, lightweight primitives and lightweight protocols) which will greatly benefit low cost RFID security and privacy and the outcome of this research will be the widespread adoption of this technology.

It is important to note that the level of security and privacy will depend on the application. It is evident that there is no universal solution but a collection of solutions suited to different applications based on compromises and on security services required.

An important consideration that is often overlooked is the ability for a cryptographic system to use a piece of hardware repeatedly to result in a more secure encryption engine. Most modern UHF RFID chips use on board oscillators with frequencies over 1 MHz. Thus within the operational timing constraints imposed as a result of US regulations, it is conceivable to allow a tag to expend around 400,000 clock cycles during a 400 millisecond period. Thus, it may be possible to redesign hardware for existing cryptographic primitives to exploit this unique scenario. However, this will be at the compromise of tag reading speeds. In addition a security mechanism that is capable of leveraging existing hardware on the tag will also reduce the cost of implementation; such a possibility has been found by using hardware used to calculate the CRC (cyclic redundancy checks) on the tags.

References

1 Tanenbaum, A.S: Computer Networks, Prentice Hall (1981)
2 Rueppel, R.A: Stream Ciphers, Contemporary Cryptology. In: Simmons, G.J. (eds.): The Science of Information Integrity. IEEE Press, (1992) 65−134
3 Menezes, A., Van Oorchot, P., Vanstone, S.: Handbook of Applied Cryptography, CRC Press (1996)
4 Mollin, R.A.: Introduction to Cryptography, Chapman & Hall/CRC, London, 2001
5 Massey, J.L: Cryptography and System Theory. In: Proceedings of the 24th Allertong conference on communication, control and computing (1986)
6 Shannon, C.E.: Communication theory of secret systems. In: Bell System Technical Journal, Vol. 28 (4) (1940) 656−715
7 Beker, H., Piper, F.: Cipher systems: the protection of communications, London, Northwood Books (1982)
8 Rueppel, R.A.: Analysis and Design of Stream Ciphers, Berlin, Springer-Verlag (1986)
9 Coppersmith, H., Krawczyk, H., Mansour, Y.: The shrinking generator. In: D.R. Stinson, (eds.): Advances in Cryptology – Crypto ’93, Springer-Verlag, New York (1994) 22–39
10 Schnier, B.: Applied Cryptography Protocols: Algorithms, and Source Code in C, John Wiley & Sons, Inc, New York (1994)
11 NIST FIPS 186-2 standard, (2004). Available from: http://csrc.nist.gov/publications/fips/fips186-2/fips186-2-change1.pdf (01/2005)
12 Simpson, L., Golic, J.D., Dawson, E.: A probabilistic correlation attack on the shrinking generator. In: Proc. ACISP ’98, LNCS, Vol. 1438, Springer Verlag (1998) 147−158
13 Golic, J.D., O’Connor, L.: A cryptanalysis of clock-controlled shift registers with multiple steps. In: Cryptography: Policy and Algorithms (1995) 174−184
14 Johansson, T.: Reduces complexity correlation attacks on two clock-controlled generators. In: Proc. Asiacrypt’98, LNCS, Vol. 1541 (1998) 342−356
15 Ekdahl, P., Meier, W., Johansson, T.: Predicting the shrinking generator with fixed connections. In: Proc Eurocrypt’03, LNCS, Vol. 2656, Springer Verlag (2004) 345−359
16 Caballero-Gil, P., Fuster-Sabater, A.: Using linear hybrid cellular automata to attack the shrinking generator. In: IEICE Trans. Fundamentals, Vol. E90-A (5) May (2006) 1166−1172
17 Kessler, I., Krawczyk, H.: Buffer length and clock rate for the shrinking generator. In: IBM Research Report, RC 19938 (88322) (1995)
18 Murashko, I., Yarmolik, V., Puczko, M.: The power consumption reducing technique of the pseudo-random test pattern generator and the signature analyser for the built-in self-test. In: CDSM, Liv-Slasko, Ukraine, Feb. (2003)
19 Puczko, M., Yarmolik, V.N.: Designing cryptographic key generators with low power consumption. In: Proceedings of the Third IEEE International Workshop on Electronic Design, Test and Applications (2006)
20 Goodman, J., Chandrakasan, A.P: Low power scalable encryption for wireless systems. In: Wireless Networks, Vol. 4 (1) (1998) 55−70
21 Chandrakasan, A.P., Sheng, S., Brodersen, R.W.: Low-Power CMOS Digital Design. In: IEEE Journal of Solid-State Circuits, Vol. 27(4) April (1992)
22 Jr. Thomas, B.: BodyLAN: A low power communications system. In: Master thesis, Massachusetts Institute of Technology, January (1996)

23 Rabaey, J.M., Chandrakasan A., Nikolic, B.: Digital integrated circuits − A design perspective, 2nd Edition, Prentice Hall, New Jersey (2003)
24 Stinson, D.R.: Cryptography Theory and Practice, CRC Press (1995)
25 Cormen, T.H., Leiserson, C.E., Rivest, R.L.: Introduction to Algorithms, first edition, MIT Press and McGraw-Hill (1990)
26 AT&T Global IP Network website, active measurements data website. Avalialable from:
http://ipnetwork.bgtmo.ip.att.net/pws/network_delay.html, date accessed (15/08/2006)
27 Transaction Processing Performance Council web page. Available from:
http://www.tpc.org (08/2005)
28 Gray, J., Reuters A.: Transaction Processing: Concepts and Techniques, Morgan Kaufmann (1993)
29 Transaction Processing Council, TPC-C Results web page. Available from:
http://www.tpc.org/tpcc/results/tpcc_perf_results.asp (10/2006)
30 Avoine, G.: Radio frequency identification: adversary model and attacks on existing protocols. In: Technical Report LASEC-REPORT-2005-001, September (2005)

Index