User Interface Requirements for Authentication of Communication

Audun Jøsang[a] **Mary Anne Patton**[b]

[a]Distributed Systems Technology Centre*
QUT, Brisbane, Qld 4001, Australia
Email: ajosang@dstc.edu.au

[b]Department of Information Environments
University of Queensland Qld 4072, Australia
Email: mapatton@mac.com

Abstract

Authentication is a security service that consists of verifying that someone's identity is as claimed. There are a number of challenges to presenting information from the authentication process to the user in a way that is meaningful and ensures security. We show examples where authentication requirements are not met, due to user behaviour and properties of existing user interfaces, and suggest some solutions to these problems.

Keywords: Security, authentication, non-repudiation, usability, user interface

1 Introduction

Authentication is a basic security service without which most other communication security services become meaningless. For example, if you don't know with whom you are communicating, there is limited meaning in encrypting the communication. Authentication is also a complex concept rife with subtleties that only experts seem to notice.

Authentication is about verifying that an identity is as claimed, and there are various ways in which this can be done in theory (ISO 1998). In this paper we will focus on *authentication of communication* which is different from *user authentication*. The latter relates to when a system verifies the identity of users before granting them access to various parts of the system. This is typically obtained by using a password or some handheld token such as a smartcard. Authentication of communication on the other hand (a.k.a. message authentication), relates to verifying the identity of the origin of information that has been received through a communication channel. Typical examples are to verify the identity of the sender of en email message or to verify the identity of an organisation behind a Web site. When nothing else is specified we will simply use the term 'authentication' to denote 'authentication of communication'.

In practice the strength of authentication also depends on the quality of the implementation. In system-to-system communication authentication can be implemented to occur automatically, whereas when humans are involved the interface needs to present evidence of the authentication process to the user. One reason why authentication can not be completely transparent to the user is that the system does not usually know the entity with which the user intends to communicate, and the system is unable to comprehend the meaning of what is communicated. These and other issues will be discussed in the following sections. We will also suggest some requirements for usable authentication and present some novel approaches for making authentication as efficient and meaningful as possible to the user.

The study of how security information should be handled in the user interface forms part of *security usability*. Relatively little research has been carried out in this field to date and it has been largely overlooked by application and hardware developers. In (Whitten & Tygar 1999) it is argued that effective security requires a different usability standard, and that it can not be achieved through the user interface techniques appropriate to other types of consumer software.

2 Cryptography and Assumptions

2.1 Digital Signature

Cryptographic authentication can be obtained in several ways, and so-called strong authentication on the Internet is based on public-key cryptography. When a message is sent from Alice to Bob, then Alice prepares a digital signature using her private signature key, and appends the signature to the message. This process is illustrated in Fig.1.

Instead of a digital signature, Alice can also append a message authentication code (MAC) to the message. Without going into detail, it can be said that the processes of generating and verifying a MAC are done by using the same secret cryptographic key (symmetric key), whereas the processes of generating and verifying a digital signature are done by a private key and a public key respectively (asymmetric keys). A digital signature can not only be used to provide authentication, but also to provide the security service called *non-repudiation* which means that the sender can not falsely deny having sent the message. On the other hand, a MAC can only ever provide authentication. The difference between non-repudiation and authentication can be explained as follows. With authentication Bob is convinced that Alice is the sender of the message, but he is not necessarily able to convince anybody else about it. A third party could for example argue that Bob generated the message himself, and Bob would be unable to prove the opposite. With non-repudiation, not only does Bob believe that Alice sent the message, Bob is also able to convince others of this, such as for example a judge. In general it is only the set of external assumptions that determines whether non-repudiation or authentication is provided. The same cryptographic mechanism can thus provide non-repudiation in one setting, but is only able to provide authentication in another setting. In other words, if the assumptions needed for non-repudiation are not satisfied, a digital signature might only provide authentication.

On the recipient side, Bob verifies the authenticity of the message as illustrated in Fig.2. What the interface needs to communicate to Alice is not only that the cryptographic verification was successful, but also the identity

*The work reported in this paper has been funded in part by the Co-operative Research Centre for Enterprise Distributed Systems Technology (DSTC) through the Australian Federal Government's CRC Programme (Department of Industry, Science & Resources).

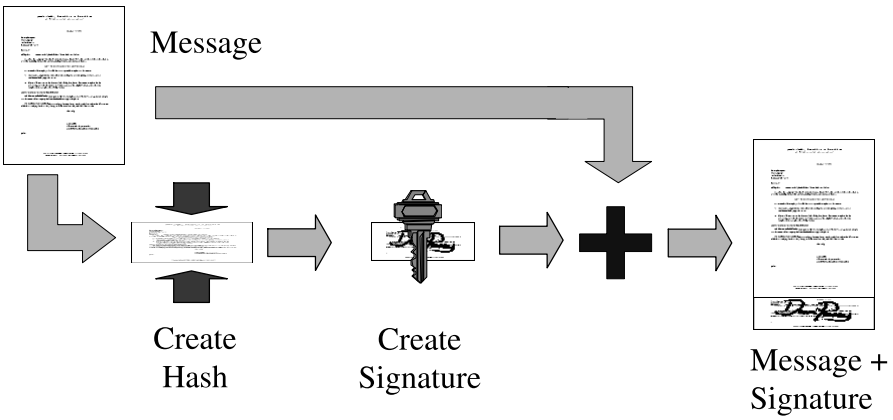

Figure 1: Generation of a digital signature on a message

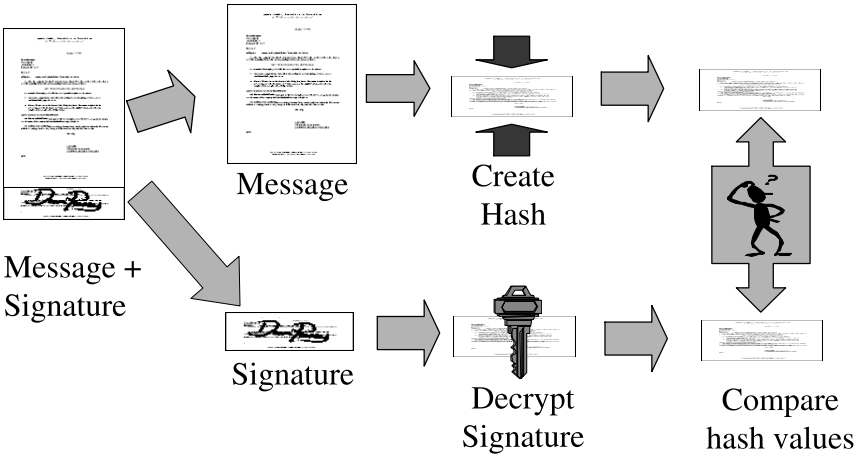

Figure 2: Verification of a message's digital signature

of the owner of the key used to generate the digital signature. If that identity is not communicated to the user, all he knows is that some key was used to generate the signature. He does not know to whom it belongs. Not only must Bob receive Alice's public key to verify the signature, but he must also be able to verify that the key really belongs to Alice. A public-key infrastructure (PKI) is typically used to solve that problem.

2.2 Public-Key Infrastructures

A PKI refers to an infrastructure for distributing public keys where the authenticity of public keys is certified by Certification Authorities (CA). A certificate basically consists of the CA's digital signature on the public key together with the key owner identity, thereby linking the two together in an unambiguous way. The structure of digital certificates is standardised by the ITU X.509 standard (ITU 1997). In order to verify a certificate the CA's public key is needed, thereby creating an identical authentication problem. The CA's public key can be certified by another CA etc., but in the end the signature verifier needs to receive the public key of some CA, usually called the root CA, out-of-band in a secure way, and this is can be seen as the Achilles heel of a PKI. The out-of-band channels can for example be physical delivery or encounter, physical mail, telephone or email. A common characteristic of out-of-band channels is that they are usually expensive to operate in comparison with the normal online channel. Because secure out-of-band channels are expensive there is often a pressure to use a cheaper but less secure online channel for distributing the root public key, but that would seriously affect the strength of the authentication and the PKI as a whole.

Most commercial PKIs are strict hierarchies, as illustrated in Fig. 3, and most only consist of one or two levels. The certification paths go strictly from the top root CA, eventually via intermediate CAs, and down to users/relying parties.

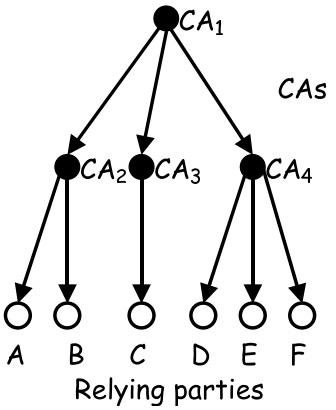

Figure 3: A strict certification hierarchy

In a strict hierarchy all users can be easily identified and found because of the hierarchic structure. A user must have out-of-band secure access the public key of the top root in order to resolve certificate chains and establish a certification chain to any other user in the hierarchy. The signature verification process thus has to resolve the certificate chain in a PKI in order to authenticate the owner of the key used to generate the signature on a message.

2.3 What's in a Name?

We humans are used to identifying people and organisations not only by name but also by additional elements of identity such as face, voice and company logo. Authentication is about verifying that an identity is as claimed, and we use a multitude of clues to assess the integrity of documents, buildings and other objects which carry these elements of identity.

Cryptographic authentication is almost exclusively based on verifying names because names lend themselves more easily to computerised processing than the additional elements mentioned above.

When Diffie and Hellman introduced public-key cryptography (Diffie & Hellman 1976) they proposed to use something similar to a telephone directory for looking up public keys. Instead of name, address and phone number it would list name, address and public key. If you wanted to communicate with John Smith you would look him up in the directory and send him a message for his eyes only using the public key found in the entry for John Smith. Attempts to create an international standard for global names include the X.500 Directory Standard (ITU 1993). X.500 was to be a global, distributed database of named entities: people, computers, printers, etc. The organisations owning some portion of the name space would operate and maintain that portion. However it is clear that this directory would be of limited value for inter-enterprise communication because access to it from outside the organisations would be restricted. Briefly said, the idea of having a unique global name space for people and organisations is not likely to become reality.

Telephone numbers, IP addresses and Internet domain names actually represent global identifiers but they are not suitable as stable and reliable identifiers of humans and human organisations. When a person or organisation moves to a new area or country their telephone numbers change. The IP address of a computer host often needs to be changed if the host is moved from one room to another, and humans have difficulty reading and remembering IP addresses anyway. Domain names are usually easy to read and remember and can in principle remain unchanged even if the hosts to which they refer are moved to another continent. Domain names were designed to provide a way of mapping physical IP addresses onto a logical name space. However domain names are linked to Internet hosts, and do **not** identify individuals or companies. An IP address or domain name, sometimes with additional elements, can be seen as the substitute of location or street address. When a name is prefixed to domain names with '@' in between it becomes an email address, and when a file or directory name is suffixed to a domain name with '/' in between it becomes a Web address. Unfortunately, the domain names in people's email addresses or organisations' URLs can change every time a person changes job or an organisation changes Internet service provider, which means that domain names are unsuitable as global identifier.

It seems that we have to stick to good old ordinary names for identifying humans and human organisations. The names we use for people and organisations on the Internet tend to be the same as the ones we use in the physical world, anything else would create unnecessary complications.

Despite the fact that names are ambiguous in a global context we rarely make mistakes because they are used in a local context or because we rely on additional elements such as face and voice for people, and company logo (trade mark) and physical location (street address) for companies. The reason logos are useful for trusted commerce in the physical world is not just familiarity, copyright and trademark protection, etc. An important part is that it is expensive and time-consuming to set up a physical shop that spoofs another with a fake logo and name, and that the risk and penalty of exposure is fairly high. This is precisely why appearance is usually enough in the physical world. That is much less true in virtual space, and public-key certificates were designed to solve this problem, but present standards and implementations fail to provide a good solution.

2.4 Syntax v. Semantics

Another problem relates to the interpretation of the message itself. The simplest approach is to look at a message as a stream of bits, and say that a digital signature applies to a particular stream of bits. However, the meaning of a message only emerges when the message is represented in some analogue form such as on a computer screen, or is interpreted by some automated system. These two views of a message, one as a stream of bits, and the other as semantic content, are radically different.

Digital documents can contain more than just text. Common examples of additional graphical features are pictures, drawings, table formatting and background colour. The correct and consistent rendering of such features is crucial for the meaning of digital documents. If graphical features are displayed differently in different applications, then the meaning is likely to change, making it meaningless to digitally sign such documents. The example below shows how inconsistent handling of table tags in HTML can make the same table look completely different.

In this scenario it is assumed that three building contractors have submitted a quote for building an office building for a company. The quotes are: "Alice Architects and Builders": $800,000, "Bob Building Contractors":$900,000, and Clark Constructions": $1,000,000. The company manager who evaluates the quotes is satisfied with the qualifications of all three contractors, and decides to list his preference as a function of price. The manager asks the company's web editor to create a table with the building contractors listed according to price. The manager digitally signs the HTML page seen in Fig.4 and submits it to the company's procurement department for further processing.

What the company manager does not know is that the web editor is a close friend of Clark, and therefore will try to make "Clark Constructions" win the contract. The web editor knows that the manager uses Opera 5, whereas the procurement department uses Netscape Navigator 4. The Web editor encodes the HTML table so that "Alice Architects and Builders" gets highest preference when viewed with Opera 5, and "Clark Constructions" gets highest preference when viewed with Netscape Navigator 4. Fig.4 shows what the tables look like when viewed in Opera 5, and Fig.5 shows what the tables look like when viewed in Netscape Navigator 4. Note that that the HTML code is identical in both cases.

This is possible because the HTML tag `tfoot` is handled inconsistently. Opera 5 always creates a row at the end of the table whereas Netscape Navigator 4 creates a row at the point where the tag appears in the HTML code. Thus by encoding "Clark Constructions" with the tag `tfoot` just after the table head, that row will appear to be the last table entry in Opera 4, and the first table entry in Netscape Navigator 4.

2.5 Assumptions

On the lowest level, a system that provides cryptographic authentication can only tell that a certain key has been used to sign a particular message. By including several external assumptions, it can be concluded that the meaning of the message was communicated by a specific named entity. In this regard, authentication can be interpreted in different ways depending on the assumptions, as described below.

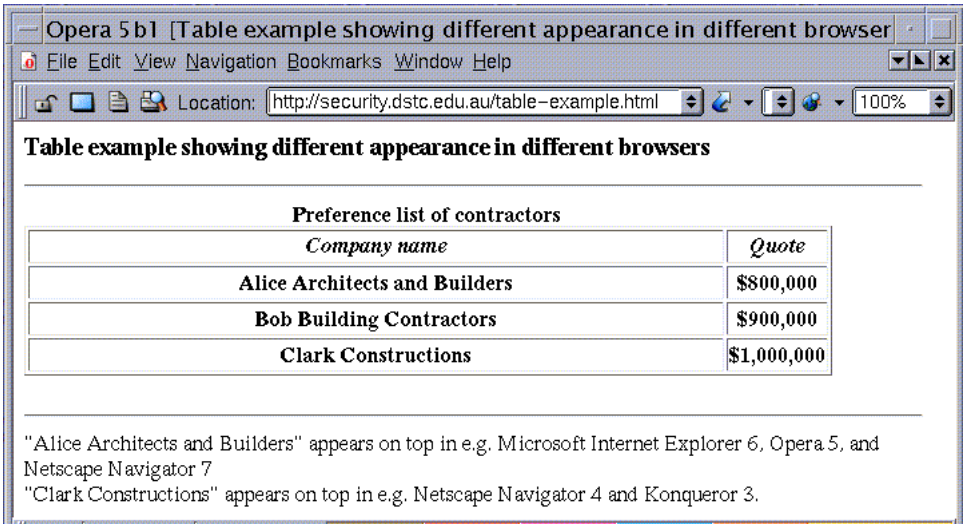

Figure 4: Viewing the table in Opera 5

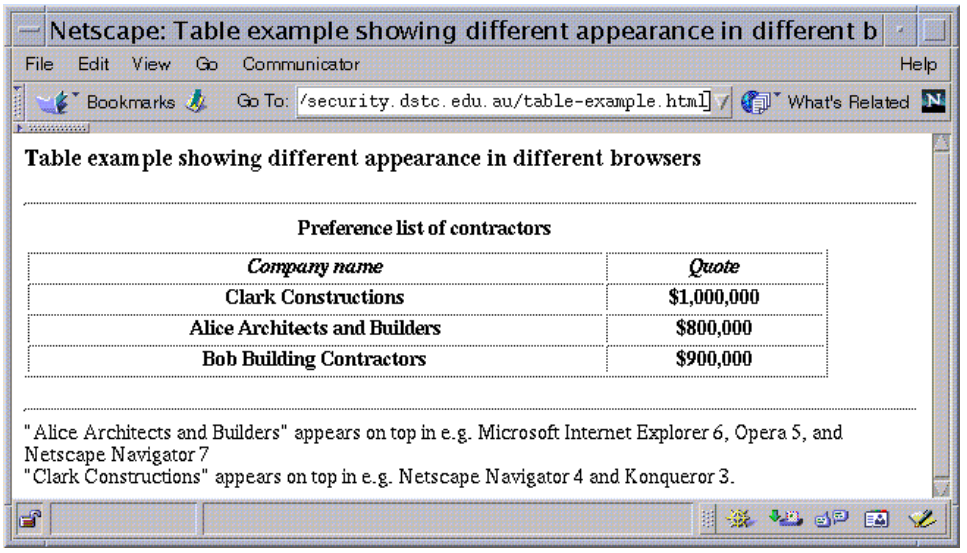

Figure 5: Viewing the table in Netscape Navigator 4

1. *A particular key was used to digitally sign the message.*

   This interpretation only requires the weak assumption that the cryptographic algorithm used to produce the signature can not be broken. Modern cryptographic algorithms can be designed with adequate strength to satisfy this assumption.

2. *Alice signed the message.*

   This interpretation requires several additional strong assumptions such as for example:

   (a) The CA has correctly verified that the entity that purchased the public-key certificate really is Alice.
   (b) Alice has protected her private key.
   (c) The name Alice is unambiguous.
   (d) Bob holds an authentic copy of the CA's public key.

3. *Alice communicated the meaning contained in the message.*

   This interpretation requires the additional assumption that the semantic rendering of a message is unambiguous.

The main issue when designing a user interface for authentication is to select the most suitable evidence and present it in a way which provides convincing proof of identity. The importance of the external assumptions must not be neglected, and the interface should as far as possible allow the user to get some information about the assumptions.

3 User Interface Requirements for Efficient Authentication

A common design philosophy is to make authentication as transparent as possible in order to reduce the mental burden on the user. Authentication can be implemented using various mechanisms, and if the mechanisms were totally hidden from the user he or she would not be able to tell whether they are working or not. This would allow successful attacks to remain undetected and thereby make the system insecure.

Authentication is a complex concept and evidence about the authentication mechanism needs to be presented in a concise and intelligible way to users. The average user has no security knowledge and user interfaces such as the Web browser have been designed to hide the authentication mechanisms and provide a minimal amount

of evidence. This is probably good for general usability, allowing users to get on with their primary task, but may be counter to meaningful and adequate authentication.

Systems do not know the name of the entity with which the user wants to communicate, and are therefor unable to judge the outcome of the authentication verification process depicted in Fig.2. All the system can do is to present some evidence from the verification process through the interface, and let the user decide if the sender identity is as expected.

The human attention span is limited and if too much evidence is presented to the user he or she will either be confused or simply tune out. This could allow authentication failures to pass undetected even if evidence about the failure is presented to the user. This poses a dilemma; too much evidence is just as bad as too little. Obviously it can not be more than the user can understand and handle but it must be sufficient for the required security level of the application. The challenge is to select the most appropriate evidence and present it to the user in an intelligible way.

An example that illustrates the subtleties of security is the padlock icon on Web browsers where an open padlock indicates insecure communication whereas a closed padlock indicates secure communication. This is seemingly a very neat and intuitive way of indicating that a Web server has been authenticated with SSL and that transmitted and received data is being encrypted. However a closed padlock only tells the user that some Web server has been authenticated but not which Web server in particular. As long as the user does not do the extra mouse clicks to view the server certificate he or she has in fact not authenticated anything at all. Despite its reassuring appearance, the padlock hides crucial aspects of security, which are required for meaningful authentication.

The Web browser does allow the user to view the server public-key certificate by clicking on the padlock icon, but users hardly ever do this, and even security aware users who view the certificate when accessing a secure Web site can have difficulty in judging whether the certificate really is what it claims to be. The browser usually checks that the domain name in the certificate is the same as the domain name pointed to by the browser, and aware users might notice when an intruder's domain name is different from the expected domain name of the bank. However, users do not usually inspect the URL for the domain name when browsing the Web, and many companies' secure Web sites have URLs with non-obvious domain names that do not correspond to the domain names of their non-secure Web sites. One example is the Norwegian bank Nordea with the URL: `http://www.nordea.no` and where its secure on-line banking has URL: `https://ibank.bbsas.no/iBank/Dispatcher`. Another vulnerability is the fact that distinct domain names can appear very similar, for example differing only by a single letter so that a false domain name may pass undetected. How easy is it for example to distinguish between the following URLs: `http://www.bellabs.com`, `http://www.belllabs.com`, and `http://www.bell-labs.com`?

In order to make authentication on the Web simpler some familiar elements from the physical world could be used. In (Jøsang, Patton & Ho 2001) it is proposed to display a digitally certified company logo in the Web browser to allow authentication at a glance and bridge the gap between the cryptographic mechanisms and humans. This idea is currently being discussed in the IETF and may become a standard feature in the future (see (Santesson 2001)). In addition a certificate can contain elements such as image and voice, which thereby can be presented to the user in order to allow meaningful authentication at a glance.

Fig.6 below illustrates the idea of using a certified company logo by showing the Web browser window when for example accessing the secure Web site of the Australian bank Westpac. We have added the certified company logo in the upper right corner outside the area displaying HTML content.

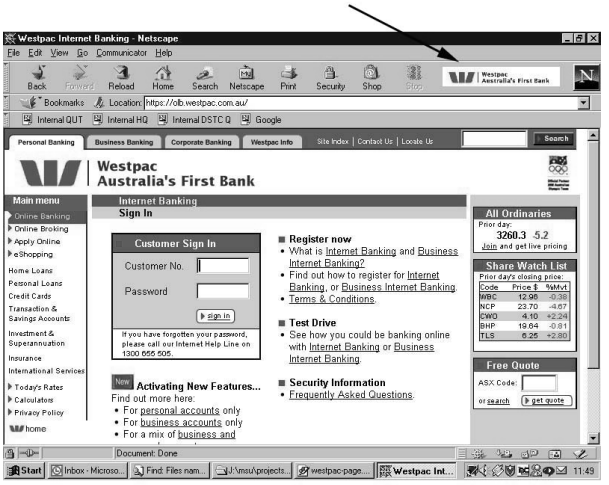

Figure 6: Example showing secure Web site with certified logo

Image and voice can only be used for strong authentication if the image and sound files are certified and included in digital certificates. This requires the CA to verify their authenticity before issuing certificates. A company logo must for example be sufficiently different from all other company logos and this requires the CA to perform a similarity check, but this is likely to create new problems. What are the criteria for a similarity check? If similar logos or names are used by companies in totally different businesses, is that OK? Hierarchies adequate to issue certificates are not by themselves adequate to ensure global uniqueness. (see e.g. (Stubblebine & Syverson 2000)). Suppose that a company obtains a certificate for a logo and then another company applies for a certificate for a much too similar logo, but it owns that logo as a registered trademark? More generally, what about revocation of a logo because of previously unrecognised problems? Does every little shop need to hire a graphic artist? What is the size of the space of meaningfully discernible logos? The authenticity of pictures of people can best be assured by taking the photos on the CA's premises. Similar requirements apply to sound files, i.e. they must be recorded in person on the CA's premises.

4 User Interface Requirements for Meaningful Authentication

When designing ways to presented authentication evidence to the user, not only must the suitability for efficiency be considered, but also the integrity. If the integrity of the evidence can not be protected from malicious manipulation, the evidence becomes meaningless.

Many weaknesses in the Web browser interface have been described, making it unsuitable for sensitive applications. It is for example quite simple for an applet to overwrite any part of the screen, and for example display a closed padlock in order to make the user believe that the communication is encrypted, or to overwrite the URL in order to create the appearance of being connected to a different Web site from what is actually the case (Lefranc & Naccache 2002).

For mobile devices the relatively small visual display will make it virtually impossible to inspect public-key certificates in the way in which they can be inspected with Web browsers. Cryptographic authentication by identity certificates such as X.509 will be unreliable because of

the difficulty of comparing an Internet site name with the identity stored in the digital certificate. Figure 7 shows a typical handheld WAP device with which server authentication will be meaningless.

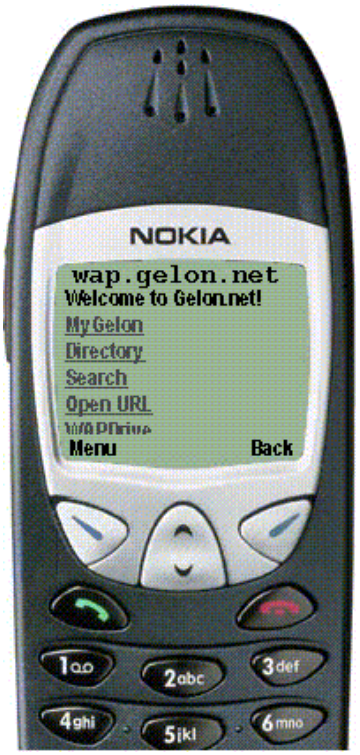

Figure 7: WAP interface that is unable to provide meaningful authentication of the WAP server

By typing the correct URL of a WAP or Web site, authentication is not really needed as long as the integrity of the network is preserved, i.e. you will access the right site as long as you type the URL correctly. WAP sites are more likely to be accessed through portals than by typing URLs, which makes cryptographic authentication the more important. However, cryptographic authentication mechanisms are only meaningful if the interface is able to provide authentication information in a secure way. For mobile devices with small display, certified company logos seem to provide the best solution.

The integrity of the authentication evidence presented to the user can be assured by having a reserved area for certified content on the interface which is never used for other types of content. Because of limited size of visual displays this might seem to be an expensive sacrifice. We therefore recommend using the normal display for displaying security information, but in a special security mode, and instead to reserve a small exclusive area to indicate that the display is in security mode. The exclusive security display area and the security display mode should not be accessible by content applications. This security mode should be easy to invoke and be distinguishable from the other display modes. The security mode of the interface then represents a separate interface channel that can be distinguished from the normal information content channel.

What represents the most suitable type of certified information to be displayed will depend on the application. A simple solution from an implementation point of view is to link the authentication directly to the logical network address used such as e.g. a telephone number or Internet domain name, and display the certified address in the separate control field. The user would then be required to know exactly which network address he or she wants to contact, and this would in fact mean that the network address becomes the global unique identifier directly associated with a person or an organisation, in the same way as telephone numbers.

5 Conclusions

The examples described in this paper outline some features of current user interfaces and user behaviour that pose a challenge to authentication of communication. A number of approaches to designing more efficient and meaningful interfaces have been outlined. Briefly said, the right type of authentication evidence must be selected for presentation, and there needs to be a separate secure channel so that it is easily distinguishable from the normal content channel.

1. The evidence that is presented must sufficient to perform a positive verification.
2. Familiar types of evidence should be used as far as possible.
3. The evidence must be protected from malicious manipulation.
4. The interface must provide a separate channel for presenting authentication evidence.
5. The meaning of the authenticated message must appear identical on any interface.

References

Diffie, W. & Hellman, M. E. (1976), 'New directions in cryptography', *IEEE Transactions on Information Theory* **22**(6), 644–654.

ISO (1998), *IS 9798. Information technology - Security techniques - Entity authentication mechanisms - Parts 1,2,3,4,5*, ISO/IEC JTC1.

ITU (1993), *Recommendation X.500, Data Communication Network Directory (also known as ISO/IEC 9594: Information Technology - Open Systems Interconnection - The Directory)*, International Telecommunications Union, Telecommunication Standardization Sector(ITU-T).

ITU (1997), *Recommendation X.509, The Directory: Authentication Framework* (also ISO/IEC 9594-8, 1995), International Telecommunications Union, Telecommunication Standardization Sector(ITU-T).

Jøsang, A., Patton, M. & Ho, A. (2001), Authentication for Humans, *in* B. Gavish, ed., 'Proceedings of the 9th International Conference on Telecommunication Systems (ICTS2001)', Cox School of Business, Southern Methodist University.

Lefranc, S. & Naccache, D. (2002), 'Cut and paste attacks with java', Cryptology ePrint Archive, Report 2002/010. http://eprint.iacr.org/.

Santesson, S. (2001), *Logotypes in X.509 certificates*, IETF PKIX Working Group INTERNET-DRAFT. URL: http://www.ietf.org/internet-drafts/draft-ietf-pkix-logotypes-00.txt.

Stubblebine, S. G. & Syverson, P. F. (2000), Authentic Attributes with Fine-Grained Anonymity Protection, *in* 'Proceedings of Financial Crypto'.

Whitten, A. & Tygar, J. (1999), Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0, *in* 'Proceedings of the 8th USENIX Security Symposium'.