Smart Defence: An Architecture for new Challenges to Cyber Security

Robert Koch, Mario Golling and Gabi Dreo Rodosek

Munich Network Management Team (MNM-Team)
Universität der Bundeswehr München
Werner-Heisenberg-Weg 39
D-85577 Neubiberg
{robert.koch, mario.golling, gabi.dreo}@unibw.de

Abstract: The last years have seen an unprecedented amount of attacks. Intrusions on IT-Systems are rising constantly - both from a quantitative as well as a qualitative point of view. Recent examples like the hack of the Sony Playstation Network or the compromise of RSA are just some examples of high-quality attack vectors. Since these Smart Attacks are specifically designed to permeate state of the art technologies, current systems like Intrusion Detection Systems (IDS) are failing to guarantee an adequate protection. In order to improve the protection, an analysis of these Smart Attacks in terms of underlying characteristics has to be performed to form a basis against those emerging threads.

Following these ideas, this paper starts by presenting individual facets of Smart Attacks in more detail. Inspired by the original definition of the term Advanced Persistent Threat of the Department of Defense, subsequently, the term Smart Attack is defined. Our architecture for Smart Defence focuses on three main elements: We propose the use of advanced geolocation for a geobased intrusion detection (e.g., inspecting new connections - originating from a location very close to where a recent attack was launched - more detailed than other connections). Furthermore, we will present our concepts on supervising Commercial Off-The Shelf (COTS) products (soft- and hardware), as both are nowadays used also in security environments. In addition, we will also show our concepts for similarity-based, multi-domain correlation as well as the corresponding proof-of-concept.

1 Introduction

The social and economic success of a society is increasingly dependent on a secure cyber space. Unfortunately, with the growing importance of the Internet for many areas within the last couple of years (e.g., for critical infrastructures such as energy or water supplies), securing the cyber space has not increased with the same speed. Quite the contrary, the number of attacks and their complexity is constantly increasing. The rising complexity, the variety of intelligence gathering techniques to access sensitive data, or the persistence and stealthy characteristics of sophisticated recent attack vectors are just some of the reasons why state of the art security systems like (Next Generation-) Firewalls, Antivirus Systems or IDSs are failing to guarantee an adequate protection. Given these *Smart Attacks*, defense mechanisms need to be adapted to fill the gap. Therefore, it is the goal of this publication to develop new solutions for IT-security to face these new challenges. In this context, our architecture, specifically designed to defend against Smart Attacks, is presented within this publication.

The paper is structured as follows: Section 2 identifies characteristics of Smart Attacks and defines the term. The subsequent Section 3 examines related work. The architecture itself is described in Section 4. In Section 5 a prototype of the proposed detection approach is presented. Finally, Section 6 concludes the paper.

2 Characteristics of Smart Attacks

Over the past years, the sophistication of attacks has increased dramatically. This evolution of attacks is reflected by well-known terms like “Targeted Attacks” or “Advanced Persistent Threat” (APT). An attack can be considered as targeted if it is *intended for a specific person* or organization, typically *created to evade traditional security defenses* and frequently makes use of advanced social engineering techniques [Sym]. Wrt. the original definition given by the DoD, APTs are regularly originated by *nation state actors* or at least state-driven. However, the term APT is often misused for different kinds of highly professional attacks, which are more likely Targeted Attacks. APTs are designed to stay below the radar, and remain undetected for as long as possible; a characteristic that makes them especially effective, moving quietly and slowly in order to evade detection [Sym].

With the ongoing sophistication of attacks, these terms are not able to describe the special characteristics in a proper way. For example, high security networks often demand what is called an “air gap”, meaning, that Intranet and Internet have to be physically separated. However, already several examples are known in which the air gap has been overcome, since there is often still a necessity of a controlled data flow between the secured and the insecure network (“swivel chair interface”). The most prominent example of this is likely to be Stuxnet. After the recent emerging of some documents leaked by Snowden, describing technologies implemented and used by the NSA to bridge the air gap, e.g., “HOWLERMONKEY”, “SOMBERKNAVE” and “COTTONMOUTH”, this endangerment is increasing dramatically. In addition of being *targeted*, *sophisticated* and *persistent*, Smart Attacks are *camouflaged* (using alternative and special covert communication channels), *multilayered* (attacking firstly the perimeter networks if no direct attack is possible and afterwards the isolated networks). Another characteristic of Smart Attacks is that they are *interdisciplinary*: For example, specialists for Human Intelligence (HUMINT) are responsible for social engineering and information gathering, programmers develop a special kind of sophisticated attack code while service technicians install a new malicious device in the network of a company (or as an "interdiction" process during the delivery).

Taking these new characteristics into account, *Smart Attacks* are defined as follows:

> An attack is considered *smart* if it is aiming at a single target or a limited target group, which is exploited in-depth. The attack is executed via the combined, interdisciplinary exploitation of multiple domains in a camouflaged way, where the means and levels of exploitation of the individual domain is below the particular detection respectively suspicion thresholds. Smart Attacks can contain, or be limited to, sleeper exploitations which are (pre-)installed in software or hardware and waiting for a specific time or an external trigger for activation.

In particular, Smart Attacks are utilizing the exploitation of multiple domains without attracting attention in the single domain and can even be pre-installed: Therefore, a detection with current security systems is highly unlikely.

3 Related Work

To encounter the extensive threats opened up by today's Smart Attacks, Smart Defence comprises numerous techniques: Effective classic protective measures as well as sophisticated new principles are required and have to be combined for an efficient protection. Because of the limited space of the publication, only a few examples of significant related work can be presented. Therefore, selected works of the most important areas, which are related with principles of our architecture, are discussed briefly as follows.

The basic principle of our architecture is the combination of established protection measures with our new developed components to achieve optimal detection results. In [EO10], the authors presented a collaborative intelligent intrusion detection system (CIIDS), which combines misuse- and anomaly-based systems. The resulting alerts of CIIDS are correlated and fuzzy logic is used to reduce the false alarm rate. While the idea of combining the advantages of the two detection principles sounds meaningful, no implementation or evaluation has been done by the authors. Even more, the DARPA 1999 IDS Evaluation dataset was proposed as being used for evaluation. While this dataset is well-known and has been used for numerous evaluations of IDSs, it also has been criticized widely because of the shortcomings wrt. data generation as well as the composition of the dataset (e.g., see [MC03]). An important possibility to identify outliers and therefore, abnormal and malign behavior, is the use of similarity measurements. Choi et al. give a comprehensive overview of binary similarity and distance measures [CCT10]. These methods can be used for intrusion detection. E.g., Yamada et al. proposed [Yam07] a system using similarity measures for the detection of malign access to servers within encrypted connections. In [FAH08], Foroushani et al. presented a system for intrusion detection in encrypted connections to servers secured by the SSH protocol. While these systems are able to search for intrusions in encrypted network traffic, comprehensive configurations and server profiles are required, prohibiting an application at large.

A fundamental problem for the implementation of a reliable IDS is located very deep - the untrustworthiness of the hardware. E.g., see the debate in 2012 about hardware backdoors implemented in the Actel/Microsemi ProASIC3 FPGA or the discussion about the security of network products from ZTE and Huawei (e.g., see [RR12]). The disclosure of more and more details about the NSA wiretapping scandal underlines this problem area of sophisticated hardware backdoors in COTS products. An intensive discussion about the security of COTS began back in 2000 [NAT00], but only focuses on the security of software products. Some work has been done to improve the reliability of the IC fabrication and for the detection of hardware backdoors and trojan circuits (e.g., see [WP12]). However, the available techniques are still limited, only detecting special kinds of circuits or not being applicable in practice.

Another opportunity for improving detection results is the use of geo-information within the intrusion detection process. [KGR13] gives a comprehensive overview of the possibilities and restrictions of different approaches for the geolocation of IP addresses.

4 Smart Defence Architecture

Because of their special characteristics, Smart Attacks can only be countered with Smart Defense. For this purpose, our architecture (which is depicted in Figure 1) makes use of the following three main principles: (1) perpetuation of successfully establish security concepts ("keep the tried and trusted") - gray color, (2) extension of local security mechanisms to include new elements (such as intrusion detection in encrypted environments, advanced geo-based intrusion detection, Layer 1 signal analysis) - orange color, and (3) use of external knowledge to improve the local detection results (from local to global knowledge) - blue color. Please note that a component for realizing the human intelligence knowledge, which is currently under development, is not depicted in the Figure.

Perpetuation of Well-Established Protective Measures
Although current protection systems are not able to ensure an adequate protection against new challenges, it must not be forgotten that in the course of IT security research over the last 30 years, a high number of security measures has been developed that are now an integral component of any IT security strategy. Referring to the technical measures, among others, this includes [Nat13]: Switch-based options like fine-grained Access Control Lists (ACLs), port security and community/isolated port-based Virtual Local Area Networks (VLANs), firewalls, patch and vulnerability management servers (such as Windows Server Update Services), virus scanners and Network Intrusion Detection Systems (NIDS), e.g., the

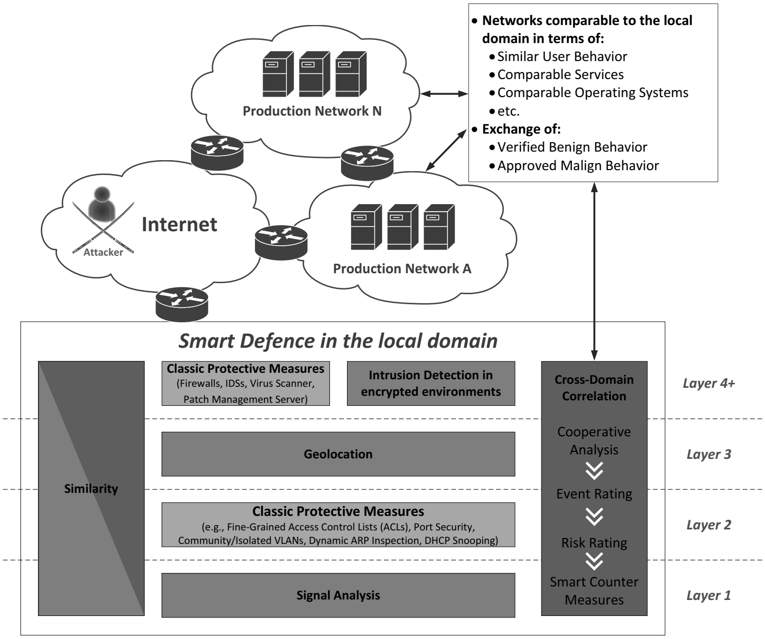
• Networks comparable to the local domain in terms of:
• Similar User Behavior
• Comparable Services
• Comparable Operating Systems
• etc.
• Exchange of:
• Verified Benign Behavior
• Approved Malign Behavior
Production Network N
Internet
Attacker
Production Network A
Smart Defence in the local domain
Similarity
Classic Protective Measures
(Firewalls, IDSs, Virus Scanner, Patch Management Server)
Intrusion Detection in encrypted environments
Geolocation
Classic Protective Measures
(e.g., Fine-Grained Access Control Lists (ACLs), Port Security, Community/Isolated VLANs, Dynamic ARP Inspection, DHCP Snooping)
Signal Analysis
Cross-Domain Correlation
Cooperative Analysis
Event Rating
Risk Rating
Smart Counter Measures
Layer 4+
Layer 3
Layer 2
Layer 1

Figure 1: Overview of the architecture

signature-based system SNORT. While these techniques are not able to prevent sophisticated attacks, they can oppress most of the daily threats which account for approx. 80 percent of all attacks. Therefore, they build the base of the defence architecture.

Integration of new Security Mechanisms

In order to address the new facets of Smart Attacks, well-established protective measures have to be enhanced with new security mechanisms (depicted in orange).

Similarity: Traditional forms of intrusion detection heavily rely on the recognition of preset patterns, which are an expression of previously acquired knowledge (usually about negative behavior). For example, IDSs like Snort or Antivirus Scanners are searching for known patterns (signatures). However, this involves in particular two disadvantages: (1) These systems are reactive by definition (i.e. the knowledge first needs to be collected) and (2) already slight variations can lead to the fact that a signature (representation of knowledge) is not applicable to a similar case. Even the use of heuristics cannot remedy this disadvantage sufficently, because only simple variations (e.g., new members of a family of a known virus) can be detected or, with a broader detection spectrum, false alarm rates are rising significantly. In contrast to this, our architecture applies similarity-based detection principles on different layers and within different components. A variety of events and data can be analyzed based on similarity evaluations, e.g., the behavior of multiple users or the degree of alikeness of two encrypted packet streams. The fundamental, anomaly-based technique used for the detection is as follows: Similarity search based on a distance function represents a way of quantifying the proximity of two objects [ZADB06]. With regard to detection, similarity based detection generally focuses on either how certain traffic is similar to (known) anomalies or how different it is from normal behavior. Even more, because often the benign behavior prevails malign behavior by far, similarity measurements can be used to identify attacks without any knowledge about the benign behavior or the service under consideration. For example, most connections to a web server of a company will be benign, while a minor portion will be malign (e.g., brute-force login attempts or SQL injections). By calculating the similarity between the different connections, the malicious ones can be identified and isolated with high probabilities [Koc11].

Signal Analysis: COTS products have revolutionized IT. While COTS products are perfectly suited for the consumer market, this may be different in security critical environments, for instance due to the loss of control over the production process. Here, using COTS may result in introducing harmful new attack vectors into these systems. On the other hand, special developments are nowadays not financially feasible in most cases. Therefore, COTS products are also used in security critical areas. To improve the security of these devices and to enable a detection of manipulation as well as attacks, we introduce a new security concept in our architecture. Focusing on the hardware layer, the idea is to perform a Layer 1 signal analysis and compare the output of multiple systems built up of hardware components of different vendors (comparative systems). On Layer 1, the signal analysis is done by evaluating the frequency spectrum as well as measuring signal similarities and using multiple techniques like cross covariance and cross correlation functions. By that, distances (*similarity*) of different flow and packet characteristics are evaluated for the detection of manipulated COTS hardware components. E.g., differing behavior or additional traffic (which will be typically highly suspicious if only one or a limited number of comparative systems exhibit this behavior) of specific hardware components can be detected and further investigations can be initiated. For a more comprehensive overview, take a look at [KGR14].

Geolocation: Similarity with regard to the origination of IP-addresses (*geolocation*) can be applied for intrusion detection as well. E.g., a recent study from the security company Mandiant - claiming to analyze Chinas Cyber Espionage Units - proclaimed, “a large share of hacking activity targeting the US could be traced to an office building in Shanghai”. Here, similar to greylisting in emails, geolocation allows to (i) correlate attacks detected with new connections and (ii) as a consequence to classify traffic a priori as more suspicious (thus particularly allowing to inspect this traffic in more detail). While suspicious IP addresses (e.g., proxies, TOR exit nodes or IPs marked by blacklists) can be blocked quite easily, this is not enough. For example, a hacker can use systems respectively IP addresses of an infiltrated network (stepping stone) to execute attacks onto the domain. Therefore, if an address is identified as suspicious, the IP as well as the addresses can be temporarily blocked to hamper the possibility of the attacker to use different IPs of the infiltrated network for the attack.

Similarity-Based Intrusion Detection: Todays IDSs either need large databases (signatures/

patterns) in case of signature and payload-based detection methods or often a learning phase to analyze the normal behavior of a network. In contrast to these techniques, our approach makes use of inherent knowledge of systems and services. For example, normally the overall majority of all connections to a website (e.g., a Web shop) will be benign, while only a little fraction will be malicious (attacking the Web shop, etc.). Due to the fact that the different actions - benign and malicious - are quite different concerning specific characteristics (like number of network packets, delay between the individual packets or number of individual connections), the majority of connections and their similarity can be used to separate the benign traffic form the malicious even without having to know - in advance - how the benign behavior looks like. Following the idea of a majority decision, a separation can be performed in real-time, without complex configurations and without the need to know specific details about the system/environment.

Cross-Domain Correlation
In particular, the last idea can also be extended to other domains. If the networks of the foreign domain have a high similarity (in terms of network and user behavior, services provided, operating systems installed, etc.), this knowledge can be used to extend the pool on which the comparative analysis is performed. In addition, the alerts of the different IDSs can be exchanged with each other in real time. Here, an analysis is first performed locally (*Event Rating*), to differentiate whether an alarm is present and if so, the corresponding severity. However, on the side of the individual IDS, diverse forms of rating models exist. In order to be applicable to a wide range of systems, the first step comprises the transition of the proprietary Risk Rating Models into the Common Vulnerability Scoring System (CVSS) by applying a transition formula. The advantage is that now the individual alerts are comparable and so a comparative analysis of the risk assessment is be made. This in turn is a prerequisite for smart countermeasures (as for instance an earlier notification of the operator or automated reactions).

5 Prototype

To verify the detection capabilities of our architecture, a proof-of-concept is currently under development. Not all modules are completely implemented yet, but most parts are ready and under extensive evaluation. At the moment, the modules are running in stand-alone mode; isolated evaluations are presented as follows. The complete architecture will focus on the detection of sophisticated Smart Attacks. Due to the usage of real-time measurements to identify malign behavior, an adaption of attackers to the detection scheme will be very challenging as long as the normal and the benign behavior has some distinctions (which should normally be the case). After the completion and integration of the final components, a comprehensive overall evaluation including detection capabilities and limitations will be done and real-time aspects and reactions will be discussed in-depth. Because of the limited space of the publication, only some results are presented as follows, namely for

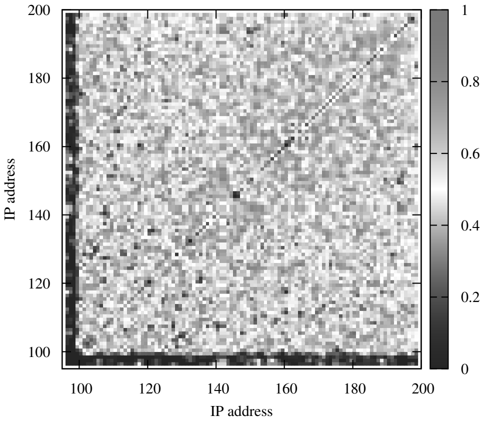

Figure 2: Similarity calculation of user behavior visualized by a colored map

the geolocation and the similarity module. The use of similarity is a basic functional principle throughout our architecture to, e.g., identify malicious behavior or the evaluation of user behavior. Based on the similarity measurements within (i) the local domain but also by (ii) the inclusion of knowledge from other domains, incidents can be checked and malicious behavior can be identified. See Fig. 2 for an example. The presented colored map was generated when encrypted connections of different users, accessing a server in the production network, were evaluated. While 99% of the users had been benign, 1% of the connections had been attacks executed against the server (brute force and SQL injection attacks). The streams of network packets of different users are correlated with each other to calculate the similarity of the data streams. While a correlation value of zero represents no similarity, increasing values show higher similarities. For a better illustration, IP addresses of attackers are below 100 (last octet), while benign users have addresses from 100 up to 200. Since this assignment of IP addresses is not used within our architecture, it does not influence the results, but simplifies the presentation. As shown in Fig. 2, there is a strong blue stripe on the left and at the bottom, representing low correlation values of the malicious connections. On the other hand, benign connections have higher correlation values, illustrated by red color: Because of the differences within the statistical features of benign respectively malicious data streams, benign sessions correlated with each other result in higher similarity values, while malicious sessions correlated with benign ones result in low values. Having 1% of malign connections, the Probability of Detection (PD) is 84.70% for a *single* evaluation. For the actual classification of a connection, multiple evaluations are used to increase the accurateness: Every connection is correlated with 11 other connections as this value was identified as the optimal number of correlation partners by empirical test runs. Based on that, the connections are classified: Having 1% of malicious connections, the final detection rate is about 99.3%. Therefore, our technique can be used to identify malign user behavior even without any knowledge about the used service, the transfered data (no need for decryption), a behavior model of the network or signatures. Note, that the system also doesn't need a configuration, in contrast to current systems using similarity for intrusion detection in encrypted communication (see Section 3).

Next, some results of the evaluation of the module for geolocation are given. Tab. 1 gives the detection accuracy of our IPv4 geolocation module for country and for city level. On country level, 99.78% of all addresses are identified correctly while on city level, our module is able to locate 90.49% of the addresses correctly. This outperforms the current other approaches, where the best one has an accuracy of 98% on country and 87.4% on city level (see [KGR13]).

Table 1: Accuracy of the geolocation module

LOCALISATION LEVEL	ACCURACY
Country	99,78%
City	90,49%

With an increasing usage of IPv6, the significance of an adequate IP geolocation will be even more important for the area of intrusion detection: While with IPv4, numerous private networks are used and masked by gateway and proxy IP addresses, the concept of IPv6 enables the possibility of identifying every device on the Internet because of the scheme for the address generation, using the hardware address of a device for building the IPv6 address. Even if Privacy Extensions (PE) are used for a randomized generation of the IPv6 address, at least the network part remains and can be used for identification. On the one hand, the IPv6 addressing scheme can be used to improve the geolocation-based portion of the proposed IDS architecture. On the other hand, because of the enormous number of possible IPv6 addresses, the actual geolocation process becomes much more challenging than in the case of IPv4. Therefore, the integration of an adequate IPv6 geolocation into our architecture will be part of our future work.

6 Conclusions and Outlook

Cyber attacks are increasing in complexity, persistence and stealthiness. APTs have more and more evolved towards so-called Smart Attacks. The paper defines the term Smart Attack and discusses characteristics, which are typically for this new very sophisticated kind of endangerment. Based on this, we propose a new architecture for Smart Defence, designed to cope with the special characteristics of this new generation of attacks. Geolocation and similarity-based intrusion detection are central components within our architecture. Based on similarity calculations, malign behavior (and therefore connections and users) can be identified without signatures or knowledge about the monitored services or characteristics of the traffic. The promising results of the evaluation of the prototypical implementations of the different modules on real data undermine the conceptual work. At the moment, the central management, control and coordination module has been implemented as well as the geolocation module. Our next steps will be to include the exchange of external domains in a better way as well as to concentrate more on Layer 1 signal analysis.

Acknowledgements

This work was partly funded by Flamingo, a Network of Excellence project (ICT-318488) supported by the European Commission under its Seventh Framework Programme.

References

[CCT10] Seung-Seok Choi, Sung-Hyuk Cha, and C Tappert. A survey of binary similarity and distance measures. *Journal of Systemics, Cybernetics and Informatics*, 8(1):43–48, 2010.

[EO10] H.T. Elshoush and I.M. Osman. Reducing false positives through fuzzy alert correlation in collaborative intelligent intrusion detection systems; A review. In *Fuzzy Systems (FUZZ), 2010 IEEE International Conference on*, pages 1–8, 2010.

[FAH08] Vahid Aghaei Foroushani, Fazlollah Adibnia, and Elham Hojati. Intrusion detection in encrypted accesses with SSH protocol to network public servers. 2008.

[KGR13] Robert Koch, Mario Golling, and Gabi Dreo Rodosek. Advanced Geolocation of IP Addresses. In *International Conference on Communication and Network Security (ICCNS)*, pages 1–10, 2013.

[KGR14] Robert Koch, Mario Golling, and Gabi Dreo Rodosek. Using Layer 1 Signal Analysis for the Supervision of COTS Products. In *9th International Conference on Cyber Warfare and Security ICCWS-2014*, 2014.

[Koc11] Robert Koch. *Systemarchitektur zur Ein-und Ausbruchserkennung in verschlüsselten Umgebungen*. BoD-Books on Demand, 2011.

[MC03] Matthew V Mahoney and Philip K Chan. An analysis of the 1999 DARPA/Lincoln Laboratory evaluation data for network anomaly detection. In *Recent Advances in Intrusion Detection*, pages 220–237. Springer, 2003.

[NAT00] NATO Research and Technology Organisation. Commercial Off-the-Shelf Products in Defence Applications (The Ruthless Pursuit of COTS). NATO, 2000.

[Nat13] National Institute of Standards and Technology. Security and Privacy Controls for Federal Information Systems and Organizations. *NIST Special Publication 800-53, Rev 4*, 2013.

[RR12] Mike Rogers and Dutch Rupperberger. Investigative Report on the U.S. National Security Issues Posed by Chinese Telecommunications Companies Huawei and ZTE. Technical report, U.S. HOR, 2012.

[Sym] Symantec Corporation. Industrial Espionage: Targeted Attacks and Advanced Persistent Threats (APTs).

[WP12] Sheng Wei and Miodrag Potkonjak. Scalable hardware Trojan diagnosis. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, 20(6):1049–1057, 2012.

[Yam07] Akira et al. Yamada. Intrusion detection for encrypted web accesses. In *Advanced Information Networking and Applications Workshops, 2007, AINAW'07. 21st International Conference on*. IEEE, 2007.

[ZADB06] Pavel Zezula, G Amato, V Dohnal, and M Batko. *Similarity search*. Springer, 2006.