Systems Security research at Politecnico di Milano (PoliMi)

Federico Maggi
Dipartimento di Elettronica e Informazione
Politecnico di Milano
Milano, Italy
fmaggi@elet.polimi.it

Stefano Zanero
Dipartimento di Elettronica e Informazione
Politecnico di Milano
Milano, Italy
zanero@elet.polimi.it

***Abstract*—This paper summarizes the past, present and future lines of research in the systems security area pursued by the Performance Evaluation Lab of Politecnico di Milano. We describe our past research in the area of learning algorithms applied to intrusion detection, our current work in the area of malware analysis, and our future research outlook, oriented to the cloud, to mobile device security, and to cyber-physical systems.**

***Keywords*-intrusion detection; malware analysis; computer virology; cloud security; systems security**

I. Introduction

This position paper describes the main research lines we are currently pursuing in our small research group at Politecnico di Milano [1], the largest school of engineering in Italy, with over 35,000 students and over 1,400 faculty members, it has a long tradition of research and teaching in all the domains of technology.

We work within the *Performance Evaluation Lab* (VPLab) [2], part of the Systems Architecture research area[3], within the department of computer science, *Dipartimento di Elettronica e Informazione*[4] (DEI), which comprises all of the ICT related research areas, with 185 faculty members and slightly short of 230 PhD students and post-doc researchers.

Our group was founded in 2005 and now includes one Assistant Professor (Stefano Zanero) and one post-doctorate research fellow (Federico Maggi), has a steady-size of 3–4 research assistants, and an yearly average of 5 BSc and MSc students. Our research group participates actively in research projects, including the FP7 STREP project WOMBAT[5], the NoE SysSec[6], and the NATO SfP project SCADA-NG.

Our research has started from the application of unsupervised learning techniques to security issues (§II), particularly in the field of anomaly-based network intrusion detection. Our current research interests is malware analysis (§III), smart devices, the cloud and cyber-physical systems (§IV).

II. The past: Learning in Intrusion Detection

Our original interests lied in applying unsupervised learning algorithms to intrusion detection tasks [1]. Over the years, this evolved into several different projects:

- ULISSE, a network based unsupervised learning IDS based on self-organizing maps and outlier detection [1], [2]. Interestingly, ULISSE was one of the first NIDS that proposed to apply learning to the payloads of network packets, and also one of the first IDS to apply a double tier of learning. The architecture of ULISSE is shown in Fig. 1
- S^2A^2DE, a host based IDS based on the analysis of the sequence and the arguments of system calls on Linux [3], [4], an evolution of SyscallAnomaly [5]. The architecture of S^2A^2DE is shown in Fig. 2
- Masibty, a web application IPS based on the analysis of the sequence, the parameters and the contents of HTTP messages, correlated with SQL queries and results to detect anomalies [6], [7].

Over the years, our research explored two interesting concepts: the possibility of using multiple layers of algorithms to analyze complex interactions, as we did, for instance in [1], [3], [6], and the conversely important issue of aggregating multiple models and facets of a phenomenon into one coherent outcome, needed to take decisions. The latter problem was explored both in the area of aposteriori aggregation of data coming from different sources [8], and in the area of combination of multiple models inside a single intrusion detector. Specifically, we were the first to propose to exploit cooperative negotiation among agents to combine the outputs of different intrusion detection models into one outcome needed to decide whether or not an attack has occurred [7], [9].

III. The present: malware analysis and computer virology

Our original interest in this area was triggered by virus propagation models and their mathematical expression [10], [11]. However, our focus quickly shifted to the issue of propagation of bluetooth and wireless malware, with a seminal experiment known as the Bluebag which received worldwide media attention [12]. More recently, our BlueBat bluetooth

[1] http://www.polimi.it
[2] http://www.vplab.elet.polimi.it
[3] http://sagroup.ws.dei.polimi.it
[4] http://www.dei.polimi.it
[5] http://wombat-project.eu
[6] http://syssec-project.eu

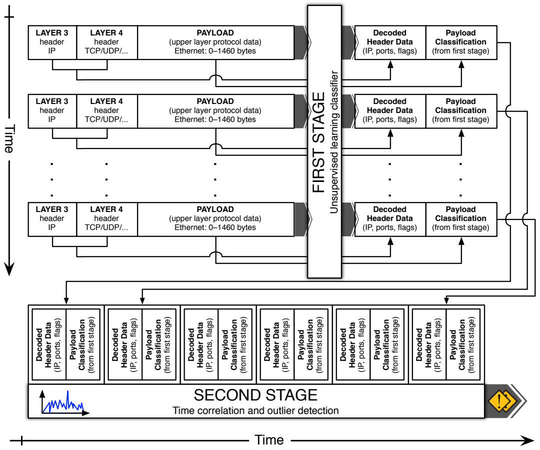
LAYER 3 header IP
LAYER 4 header TCP/UDP/...
PAYLOAD (upper layer protocol data) Ethernet: 0–1460 bytes
FIRST STAGE
Unsupervised learning classifier
Decoded Header Data (IP, ports, flags)
Payload Classification (from first stage)
Time
SECOND STAGE
Time correlation and outlier detection
Time

Figure 1. The architecture of the NIDS ULISSE

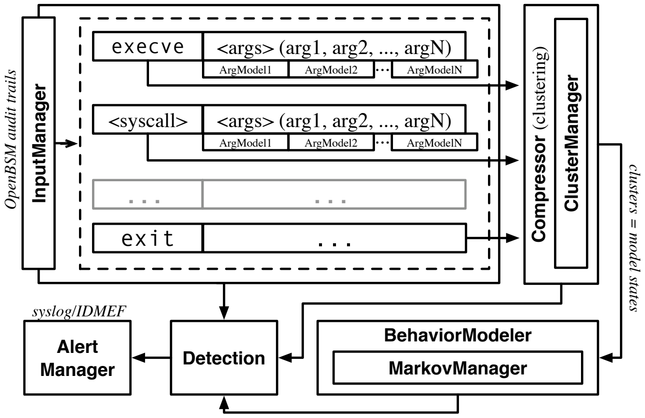

Figure 2. The architecture of the HIDS S^2A^2DE.

honeypots [13] led us to express serious doubts about the actual dangerousness of wireless-enabled malware [14]. Our ongoing research is now trying to say a final word on the subject of viability of wireless-spread infections.

Another issue we explored, jointly with colleagues from Technical University of Vienna and University of California, Santa Barbara, is automation of malware analysis. Today, each newly discovered malware binary must be analyzed mostly by hand, to understand its capabilities, its level of threat and its potential impact. We developed an hybrid approach mixing dynamic and static analysis, to overcome their symmetric limitations. Dynamic analysis is unlikely to explore all of the malware capabilities, i.e., to execute all of the reachable code, as most modern malware includes triggers that execute certain functions only if some conditions are verified. On the other hand, static analysis is very difficult to automatize. We proposed a system called Reanimator [15] that exploits similarities in the code base among different malware samples, specifically by identifying interesting behaviors, mapping them back to the code implementing them, and creating a resilient set of fingerprints based on the Control Flow Graph (CFG) of said genotypes. A limitation of Reanimator is that the analyst needs to define manually the behaviors of interest. For this reason, we are working to exploit clustering on both the structural features of a malware collection and the dynamic features. Our objective is threefold:

1) as dynamic clustering necessarily works on an incomplete set of features (because of the inherent incompleteness of dynamic analysis), comparison between the dynamic and static analysis may reveal the presence of dormant behaviors in malware samples, thus helping to improve the clustering in malware families.
2) we can use dynamic clustering to automatically derive "interesting behavior" sets from the malware dataset itself, thus further automating the approach laid out by Reanimator and removing one manual step.
3) as arguably a behavior can be reimplemented from scratch, thus "blinding" the techniques used by Reanimator (but not the dynamic behavioral clustering), we can use the matching between statically and dinamically derived clusters to expand the set of signatures for a behavior, in order to improve the ability to detect it in other dormant samples.

Another research problem we have been devoting attention to is the analysis malware naming inconsistencies. In particular, we recently came up with a demonstration that major inconsistencies plague the naming convention and malware taxonomy employed by different vendors. This creates an obvious issue for researchers focusing on integrating and systematizing threats, for instance to create ground truths for automated analysis approaches.

In addition to topics strictly related to the malware domain, we have been focusing on two aspects of the current threat scenario. First, we have been conducting the largest and most realistic data collection experiment on the World Wide Web that features more than 5,300 users. The goal of this experiment is to determine accurately the extent to which short URLs, one of the most revolutionary technologies in Web 2.0, masquerade significant threats, by acting as "amplifiers" of the attack surface (i.e., web clients) with respect to attack vectors such as phishing, drive-by download and spam URLs. The first phase of this large experiment is concluded and its results, which will be soon submitted for review, represent a significant improvement over the state-of-the-art work [16]. Second, we have designed and released the beta version of **BURN** [17], an interactive visualization tool for displaying autonomous systems exhibiting rogue activity that helps at finding misbehaving networks through visual and interactive exploration. Last, we are investigating underground economy and the usage of captured credentials is an issue we are currently devoting some research efforts. In this context, we have explored the phenomena of vishing [18], [19] and have built an architecture, described in Fig. 5, to capture relevant data for analyzing this growing phenomenon.

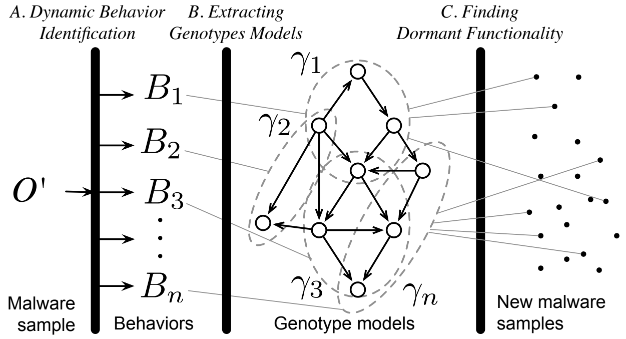

Figure 3. The architecture of Reanimator.

IV. The future: smart devices, the cloud, and cyber-physical systems

Arguably, the future of ICT is characterized by a pervasive access to the Internet through smart devices, by the extension of the cloud computing paradigm, and by the increasing interaction between the digital and the physical world. We have already mentioned our work on bluetooth-enabled smartphones [12], [13]. Building upon this expertise, we are currently working on the vulnerabilities of smartphone user interfaces and input systems.

We have some past experience on analyzing the grid computing paradigm [20], arguably one of the ancestors of the upcoming cloud computing revolution. In this area, we are working on the basis of the observation that there is a strong parallel between the emerging paradigm of cloud computing and the traditional time-sharing era [21]. Clouds are the modern reincarnation of mainframes, available on a pay-per-use basis, and equipped with virtual, elastic, disks-as-a-service that replace the old physical disks with quotas. This comparison, beyond being fascinating in its own self, prepares the ground for a constructive critique regarding the security of such a computing paradigm and, especially, of one of its key components: web services. Along this line, we concentrate on a few, critical hypotheses that demand particular attention. Although in this emerging landscape only a minority of threats qualify as *novel*, they could be difficult to recognize with the current countermeasures, given the change that the new computing paradigm has induced in the use of the network stack (see Fig. 4), and thus can expose web services to new attacks. Our current research works by analyzing the traditional countermeasures such as intrusion detection systems, developed to mitigate well-known web security threats, and by trying to explore the affinities and differences when trying to use them within the cloud computing paradigm.

The final new trend that we wish to study is the emerging class of security issues that arise in the interstitial layer between safety-critical, physical systems and digital, pervasive systems (a typical example is a SCADA-controlled industrial process, but in the near future we can foresee more and more such interactions).

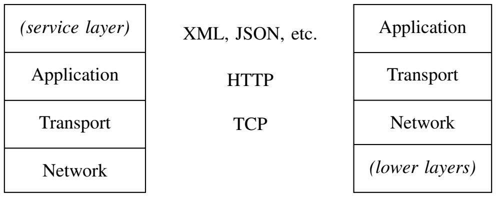

Figure 4. The change in the networking stack is noticeable from the traditional application layer (left) that, in the case of HTTP, is playing the role of a transport protocol (right) to encapsulate upper layer protocols (e.g., SOAP, JSON, XML), typical of modern web services.

The increasing interconnection between such systems creates new attack surfaces that are neither physical nor digital, and which cannot be identified if such systems are studied and secured separately as is customarily done nowadays. As such systems are prevalent in critical infrastructures such as power grids or water-distribution plants, they are primary target for cyber-terrorism and cyber-warfare attacks.

We are beginning to investigate this emerging class of vulnerabilities with an empirical, bottom-up methodology. We will start from devising real-world attacks against both the digital and the physical side of carefully selected target systems, and strive to unveil recurring vulnerability patterns that can be generalized to a (possibly novel) class of vulnerabilities. To ensure their real-world applicability, we will validate, step by step, assumptions, results and countermeasures on replicas, models or simulated systems in controlled environments, with the help of industrial partners.

As a first result, we aim to generate actionable assessments of the security of representative, high-profile systems, structured in a series of novel attack papers, and correspondingly produce targeted countermeasures. However, our more ambitious long term goal is to formalize the general problem and its root causes, by attempting to develop a theory, a taxonomy and a methodology for describing, identifying and assessing this novel class of vulnerabilities. Leveraging this general theory, and systematizing our observations, we will also try to produce preemptive secure design patterns and general solutions for this class of problems.

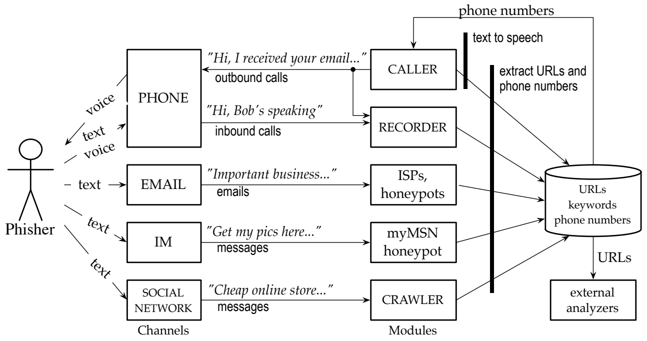

Figure 5. Our social-engineering centric data collection architecture.

ACKNOWLEDGMENT

The research described in this position paper has received funding under the 7th Framework Programme (FP7/2007-2013) under grant agreements nr. 257007 (SysSec) and 216026 (WOMBAT). The opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the European Commission. Further funding for our research has been received from NATO under the SfP programme, contract nr. 983805 "SCADA-NG".

REFERENCES

[1] S. Zanero and S. M. Savaresi, "Unsupervised learning techniques for an intrusion detection system," in *Proc. of the 2004 ACM Symposium on Applied Computing (SAC), Nicosia, Cyprus, March 14-17, 2004*, H. Haddad, A. Omicini, R. L. Wainwright, and L. M. Liebrock, Eds. ACM, 2004, pp. 412–419.

[2] S. Zanero, "Ulisse, a network intrusion detection system," in *Proc. of the 4th annual workshop on Cyber security and information intelligence research: developing strategies to meet the cyber security and information intelligence challenges ahead*, ser. CSIIRW '08. New York, NY, USA: ACM, 2008, pp. 20:1–20:4.

[3] F. Maggi, M. Matteucci, and S. Zanero, "Detecting intrusions through system call sequence and argument analysis," *Dependable and Secure Computing, IEEE Transactions on*, vol. 7, no. 4, pp. 381 –395, Oct-Dec 2010.

[4] A. Frossi, F. Maggi, G. L. Rizzo, and S. Zanero, "Selecting and improving system call models for anomaly detection," in *Detection of Intrusions and Malware, and Vulnerability Assessment, 6th International Conference, DIMVA 2009, Como, Italy, July 9-10, 2009. Proc.*, ser. Lecture Notes in Computer Science, U. Flegel and D. Bruschi, Eds., vol. 5587. Springer, 2009, pp. 206–223.

[5] D. Mutz, F. Valeur, G. Vigna, and C. Kruegel, "Anomalous system call detection," *ACM Trans. Inf. Syst. Secur.*, vol. 9, pp. 61–93, February 2006.

[6] C. Criscione, G. Salvaneschi, F. Maggi, and S. Zanero, "Integrated detection of attacks against browsers, web applications and databases," in *Proc. of the 2009 European Conference on Computer Network Defense*, ser. EC2ND '09. Washington, DC, USA: IEEE Computer Society, 2009, pp. 37–45.

[7] A. Volpatto, F. Maggi, and S. Zanero, "Effective multimodel anomaly detection using cooperative negotiation," in *Decision and Game Theory for Security - First International Conference, GameSec 2010, Berlin, Germany, November 22-23, 2010. Proc.*, ser. Lecture Notes in Computer Science, T. Alpcan, L. Buttyán, and J. S. Baras, Eds., vol. 6442. Springer, 2010, pp. 180–191.

[8] F. Maggi, M. Matteucci, and S. Zanero, "Reducing false positives in anomaly detectors through fuzzy alert aggregation," *Information Fusion*, vol. 10, no. 4, pp. 300–311, 2009.

[9] F. Amigoni, F. Basilico, N. Basilico, and S. Zanero, "Integrating partial models of network normality via cooperative negotiation: An approach to development of multiagent intrusion detection systems," in *Proc. of the 2008 IEEE/WIC/ACM International Conference on Intelligent Agent Technology, Sydney, NSW, Australia, December 9-12, 2008*. IEEE, 2008, pp. 531–537.

[10] G. Serazzi and S. Zanero, "Computer virus propagation models," in *Performance Tools and Applications to Networked Systems*, ser. Lecture Notes in Computer Science, M. Calzarossa and E. Gelenbe, Eds. Springer Berlin Heidelberg, 2004, vol. 2965, pp. 26–50.

[11] E. Filiol, M. Helenius, and S. Zanero, "Open problems in computer virology," *Journal in Computer Virology*, vol. 1, no. 3-4, pp. 55–66, 2006.

[12] L. Carettoni, C. Merloni, and S. Zanero, "Studying bluetooth malware propagation: The bluebag project," *IEEE Security & Privacy*, vol. 5, no. 2, pp. 17–25, 2007.

[13] A. Galante, A. Kokos, and S. Zanero, "Bluebat: Towards practical bluetooth honeypots," in *Proc. of IEEE International Conference on Communications, ICC 2009, Dresden, Germany, 14-18 June 2009*. IEEE, 2009, pp. 1–6.

[14] S. Zanero, "Wireless malware propagation: A reality check," *IEEE Security & Privacy*, vol. 7, no. 5, pp. 70–74, 2009.

[15] P. M. Comparetti, G. Salvaneschi, E. Kirda, C. Kolbitsch, C. Kruegel, and S. Zanero, "Identifying dormant functionality in malware programs," in *31st IEEE Symposium on Security and Privacy, S&P 2010, 16-19 May 2010, Berleley/Oakland, California, USA*. IEEE Computer Society, 2010, pp. 61–76.

[16] D. Antoniades, I. Polakis, G. Kontaxis, E. Athanasopoulos, S. Ioannidis, E. Markatos, and T. Karagiannis, "we.b: The web of short URLs," in *WWW 2011*, 2011.

[17] F. Roveta, L. di Mario, F. Maggi, G. Caviglia, S. Zanero, and P. Ciuccarelli, "Burn: Baring unknown rogue networks," 2011.

[18] F. Maggi, A. Sisto, and S. Zanero, "A social-engineering-centric data collection initiative to study phishing," in *Proc. of the First Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS)*. ACM Digital Library, Apr 2011. [Online]. Available: http://home.dei.polimi.it/fmaggi/downloads/publications/2011_maggi_sisto_zanero_vishing.pdf

[19] F. Maggi, "Are the con artists back? a preliminary analysis of modern phone frauds," in *Proc. of the International Conference on Computer and Information Technology (CIT)*. IEEE Computer Society, 2010, pp. 824–831. [Online]. Available: http://home.dei.polimi.it/fmaggi/downloads/publications/2010_maggi_vishing.pdf

[20] S. Zanero and G. Casale, "Givs: integrity validation for grid security," *IJCIS*, vol. 4, no. 3, pp. 319–333, 2008.

[21] F. Maggi and S. Zanero, "Is the future web more insecure? distractions and solutions of new-old security issues and measures," 2011.