Computer Security and Machine Learning: Worst Enemies or Best Friends?

Konrad Rieck
Technische Universität Berlin
Germany

***Abstract*—Computer systems linked to the Internet are confronted with a plethora of security threats, ranging from classic computer worms to involved drive-by downloads and bot networks. In the last years these threats have reached a new quality of automatization and sophistication, rendering most defenses ineffective. Conventional security measures that rely on the manual analysis of security incidents and attack development inherently fail to provide a timely protection from these threats. As a consequence, computer systems often remain unprotected over longer periods of time.**

The field of machine learning has been considered an ideal match for this problem, as learning methods provide the ability to automatically analyze data and support early detection of threats. However, only few research has produced practical results so far and there is notable skepticism in the community about learning-based defenses. In this paper, we reconsider the problems, challenges and advantages of combining machine learning and computer security. We identify factors that are critical for the efficacy and acceptance of learning methods in security. We present directions and perspectives for successfully linking both fields and aim at fostering research on intelligent security methods.

I. Introduction

The amount and diversity of security threats in the Internet has drastically increased. While only few years ago most attacks have been developed for fun rather than profit, we are now faced with a plethora of professional security threats, ranging from stealthy drive-by downloads to massive bot networks. These threats are employed by an underground economy for illegal activities, such as theft of credit card data, distribution of spam messages and denial-of-service attacks [see 1, 2]. As part of this development, attacks and malicious software have been systematically advanced in automatization and sophistication. Today's attack tools comprise a wide range of functionality, including various techniques for propagation, infection and evasion.

This change in the threat landscape confronts computer security with new challenges. Basically, computer security can be viewed as a cyclic process, which starts with the discovery of novel threats, continues with their analysis and finally leads to the development of prevention measures (Figure 1). This process builds on manual processing of data, that is, security practitioners take care of updating detection patterns, analyzing threats and crafting appropriate defenses. With the growing automatization of attacks, however, this cycle increasingly gets out of balance. The amount and complexity of threats renders manual inspection time-consuming and often futile. Thus, only a minor fraction of novel security threats is sufficiently analyzed for protecting computer systems in the future (arrows in Figure 1).

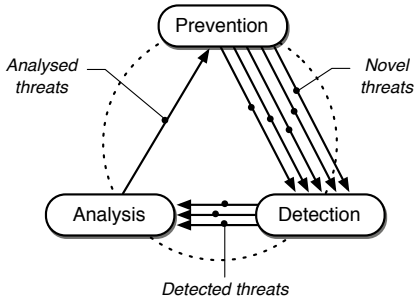

Figure 1: Computer security as a cyclic process.

Clearly, there is a need for techniques that help to analyze and fend off novel threats more quickly. If the attackers are systematically automatizing their instruments, why not try the same in the context of defense? The field of machine learning has been considered an ideal match for this problem, as learning methods are able to automatically analyze data and provide timely decisions, for example when identifying attacks against services [3, 4] or web browsers [5, 6]. Unfortunately, many researchers have exploited security solely as a playground for benchmarking learning methods, rather than striving for practical solutions. Despite a large body of work, only few research has produced practical results and there is notable skepticism in the security community about machine learning [7, 8].

In view of the possible advantages of learning-based defenses and the demand for alternative security measures, its worth reconsidering the combination of computer security and machine learning. In this paper, we study the problems, challenges and perspectives of linking the two fields. We identify key factors that contribute to the efficacy and acceptance of learning methods in security. While previous work has largely focused on making learning effective, we also emphasize the need for transparent and controllable

methods that can assist a human expert during analysis. Based on these observations, we present directions for future work on combining learning and security, where we point out new perspectives in detecting, analyzing and preventing security threats.

II. Problems and Challenges

Computer security fundamentally differs from other application domains of machine learning. The sound application of a learning method requires carefully addressing various constraints that are crucial for operating a security system in practice. While the performance of machine learning in other areas is often determined by a single quality, such as the classification accuracy, security involves several factors that require attention. Sommer and Paxson [8] have studied some of these factors for network intrusion detection. We extend this work to the generic application of machine learning and identify five key factors that impact the efficacy of learning-based security systems.

(a) *Effectivity:* First, any learning method applied in the context of security needs to be effective—either in detecting, analyzing or preventing threats. In contrast to other areas, this effectivity is highly problem-specific and may involve several quality metrics. For example, an intrusion detection system must accurately identify attacks as well as attain a reasonable low false alarm rate, as otherwise it is inapplicable in practice.

(b) *Efficiency:* A second important factor is efficiency. The main motivation for using learning methods in security is their ability to automatically provide results. Thus, learning needs to be fast to achieve a benefit over conventional security techniques. A good example is the work of Bayer et al. [9] which systematically improves the run-time performance of a clustering method for malicious software [10].

The majority of previous research has focused on these two factors when considering learning in security applications. Operating a system in practice, however, also requires addressing demands of practitioners. A main reason for the lack of machine learning in practical security is that effectivity and efficiency alone are not sufficient for designing successful security systems.

(c) *Transparency:* One central aspect is transparency. No practitioner is willing to operate a black-box system, which fails to provide explainable decisions. Fortunately, machine learning is not per se opaque and there exist several approaches for explaining the decisions of learning methods. One example is the visualization developed by Rieck and Laskov [11] which enables explaining the decisions of several learning-based intrusion detection systems.

(d) *Controllability:* Many security experts are deterred by the idea of handing over control to a learning method. This concern reflects a relevant problem of machine learning in security: learning-based systems must retain control of the operator, such that false decisions can be immediately corrected and the system adapted to dynamics in the environment. A key to this problem is changing the role of machine learning from operating totally autonomously to being actively supervised by a human expert.

(e) *Robustness:* Finally, any extension to a security system will become a target of attacks itself. Hence, machine learning must also deal with the problem of being attacked, for example, if an adversary tampers with the learning process or tries to evade detection and analysis [12, 13]. If considered during the design however, learning methods can be constructed in a robust manner and withstand different attack types, for example by randomization [14] and diversification [15] of the learning process.

We need to note here that none of these factors is new in the field of computer security and actually any practical security system should address these key factors—whether it applies machine learning or not. It thus comes as no surprise that even many conventional security instruments fail to satisfy all factors equally well. For example, many tools for attack detection suffer from false alarms and analysis systems for malicious software are often vulnerable to evasion. Nevertheless, it is a pity that a substantial body of previous work on learning for security has ignored these factors and there is a clear demand for research that brings the promising capabilities of machine learning to practical security solutions.

III. Perspectives and Applications

Based on these observation, we are ready to explore perspectives for machine learning in computer security. In view of the presented constraints and problems, this research is quite challenging and demanding. Sommer and Paxson [8] thus suggest to apply machine learning solely as a tool for preprocessing data. However, the ability of machine learning to provide protection from novel threats, only comes into effect if learning methods are deployed in the first front of defenses. Consequently, we herein argue that machine learning resembles a tool for directly strengthening the full cycle of computer security (Figure 1)—provided practical constraints and factors are carefully addressed.

In the following, we give a brief description of some promising applications of machine learning, including the detection of unknown network attacks, the automatic analysis of malicious software and the assisted search for vulnerabilities in software.

Proactive Detection of Attacks: One of the main advantages of machine learning over regular security techniques is its ability to detect anomalous events and identify novel attacks. Starting with the seminal work of Denning [16], learning methods for anomaly detection have been applied in different contexts of security. In particular, for network intrusion detection several methods have been developed which attain remarkable effectivity in practice [3, 4, 15]. However, all of these methods operate as black-box systems and do not provide interfaces for controlling and amending the detection process.

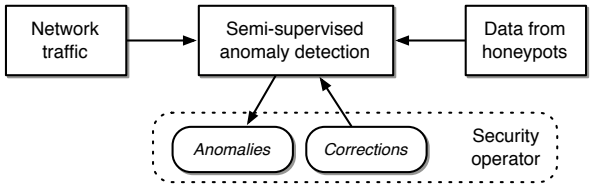

Figure 2: Schematic depiction of proactive threat detection.

A first step towards improving the practicability is thus the development of transparent anomaly detection methods which enable understanding and adapting their detection models during operation. One direction for addressing this problem is linking learned models back to their underlying features, for example, by automatically transforming statistical models into equivalent string patterns and rules. In contrast to numbers and vectors, strings and rules can be easily adapted and thereby allow an operator to carefully refine learning models in practice.

A further addition is the combination of anomaly detection and proactive techniques, such as server-based and client-based honeypots. Honeypots allow to automatically monitor malicious activity and provide a valuable source for training and calibrating learning methods. If combined with techniques for semi-supervised anomaly detection, these information can be directly fed into the learning process. For example, malicious web sites detected using honeypots and sandboxes [5, 17] can be transferred to a learning-based web proxy [e.g., 6] to create a dynamic defense against the threat of drive-by-download attacks. A corresponding detection system is illustrated in Figure 2.

Automatic Analysis of Threats: Another promising area for the application of learning in security is the analysis of threats. Security researchers are swamped by the amount of malicious activity in the Internet. Whether analyzing malicious programs, faked profiles in social networks or web pages of spam campaigns, in many settings there are thousands of data instances per day that need to be analyzed and fit into a global picture of threats. Machine learning can greatly assist in this process and provide a valuable instrument for accelerating threat analysis.

In particular, the automatic analysis of malware has proved to be a fruitful ground for learning. In the last years techniques for automatic classification and clustering of malware have been developed [9, 10, 18] which allow to identify malware variants as well as discover new families of malicious software. However, grouping malware into classes is only a one step in defending against malicious code. What is needed are analysis techniques that extract relevant information from these groups and propose patterns for signature generation to the analyst. Hence, novel learning systems need to be developed that automatically extract discriminative patterns from malicious code and guide the construction of anti-malware signatures.

While clustering has been studied for analysis of malicious programs, malicious web pages and malicious network flows, none of these approaches provides the ability to selectively correct the learned grouping. Often however, a security expert can clearly indicate some instances of a sample that need to be grouped into the same cluster and identify pairs that should to be placed in different groups. Currently, this information is lost. A possible direction of research hence lies in semi-supervised clustering methods that group data instances into clusters while at the same time satisfying the constraints given by a human expert. A corresponding system is illustrated in Figure 3.

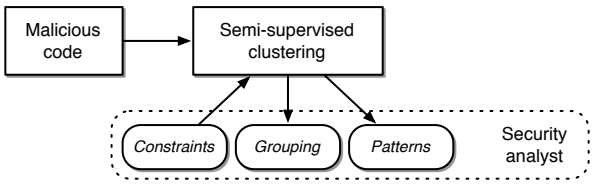

Figure 3: Schematic depiction of automatic threat analysis.

Assisted Discovery of Vulnerabilities: A third area for application of machine learning that has received almost no attention so far is the discovery of vulnerabilities. Security ultimately aims at eliminating threats and thus finding vulnerabilities is a crucial step for protecting computer systems. The search for security flaws is usually carried out in one of two extremes: on the hand vulnerabilities are often discovered in a brute-force manner using fuzzing techniques, whereas on the other hand security researchers devote considerable time into manually tracking down software vulnerabilities in program code.

Machine learning can help in establishing a link between these contrasting workflows. Instead of blindly scanning for possible vulnerabilities, the search may be actively guided by learning methods that incorporate knowledge about problematic programming constructs and known vulnerabilities of similar software. For example, known flaws in a web browser may be used to discover similar though not identical

vulnerabilities in the code of another browser. Similarly, an expert may actively control the search of a learning-based auditing tool, once positive or negative results are reported. A schematic depiction of this concept is illustrated in Figure 4.

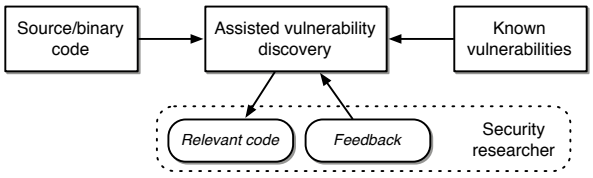

Figure 4: Schematic depiction of vulnerability search.

IV. Conclusions

In conclusion, we can note that computer security and machine learning are far from being "worst enemies". Instead, there is good hope to make them "best friends" in the near future. To this end, the proposed directions and perspectives for linking the two fields are currently explored by a group of security and learning researchers at Technische Universität Berlin and soon at the University of Göttingen.

While this paper can not generally rule out the difficulties of applying machine learning in the field of security, it pinpoints the relevant challenges and advantages of linking the two and aims at fostering interesting security research to keep abreast of future attack developments.

Acknowledgments

The author acknowledges funding from the *Bundesministerium für Bildung und Forschung* under the project PROSEC (FKZ 01BY1145).

References

[1] J. Franklin, V. Paxson, A. Perrig, and S. Savage, "An Inquiry Into the Nature and Causes of the Wealth of Internet Miscreants," in *Proc. of Conference on Computer and Communications Security (CCS)*, 2007, pp. 375–388.

[2] "Symantec Global Internet Security Threat Report: Trends for 2009," Vol. XIV, Symantec, Inc., 2010.

[3] Y. Song, A. Keromytis, and S. Stolfo, "Spectrogram: A Mixture-of-Markov-Chains Model for Anomaly Detection in Web Traffic," in *Proc. of Network and Distributed System Security Symposium (NDSS)*, 2009.

[4] W. Robertson, F. Maggi, C. Kruegel, and G. Vigna, "Effective anomaly detection with scarce training data," in *Proc. of Network and Distributed System Security Symposium (NDSS)*, 2010.

[5] M. Cova, C. Kruegel, and G. Vigna, "Detection and analysis of drive-by-download attacks and malicious JavaScript code," in *Proc. of the International World Wide Web Conference (WWW)*, 2010.

[6] K. Rieck, T. Krueger, and A. Dewald, "Cujo: Efficient detection and prevention of drive-by-download attacks," in *Proc. of 26th Annual Computer Security Applications Conference (ACSAC)*, 2010.

[7] C. Gates and C. Taylor, "Challenging the anomaly detection paradigm: A provocative discussion," in *Proc. of New Security Paradigms Workshop (NSPW)*, 2006, pp. 21–29.

[8] R. Sommer and V. Paxson, "Outside the closed world: On using machine learning for network intrusion detection," in *Proc. of IEEE Symposium on Security and Privacy*, 2010, pp. 305–316.

[9] U. Bayer, P. Comparetti, C. Hlauschek, C. Kruegel, and E. Kirda, "Scalable, behavior-based malware clustering," in *Proc. of Network and Distributed System Security Symposium (NDSS)*, 2009.

[10] M. Bailey, J. Oberheide, J. Andersen, Z. M. Mao, F. Jahanian, and J. Nazario, "Automated classification and analysis of internet malware," in *Recent Adances in Intrusion Detection (RAID)*, 2007, pp. 178–197.

[11] K. Rieck and P. Laskov, "Visualization and explanation of payload-based anomaly detection," in *Proc. of European Conference on Computer Network Defense (EC2ND)*, November 2009.

[12] Y. Song, M. Locasto, A. Stavrou, A. D. Keromytis, and S. J. Stolfo, "On the infeasibility of modeling polymorphic shellcode: Re-thinking the role of learning in intrusion detection systems," *Machine Learning*, 2009.

[13] R. Perdisci, D. Dagon, W. Lee, P. Fogla, and M. Sharif, "Misleading worm signature generators using deliberate noise injection," in *Proc. of IEEE Symposium on Security and Privacy*, 2006, pp. 17–31.

[14] G. Cretu, A. Stavrou, M. Locasto, S. Stolfo, and A. Keromytis, "Casting out demons: Sanitizing training data for anomaly sensors," in *Proc. of IEEE Symposium on Security and Privacy*, 2008.

[15] R. Perdisci, D. Ariu, P. Fogla, G. Giacinto, and W. Lee, "McPAD: A multiple classifier system for accurate payload-based anomaly detection," *Computer Networks*, vol. 5, no. 6, pp. 864–881, 2009.

[16] D. Denning, "An intrusion-detection model," *IEEE Transactions on Software Engineering*, vol. 13, pp. 222–232, 1987.

[17] J. Nazario, "A virtual client honeypot," in *Proc. of USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET)*, 2009.

[18] K. Rieck, P. Trinius, C. Willems, and T. Holz, "Automatic analysis of malware behavior using machine learning," *Journal of Computer Security*, 2011, to appear.