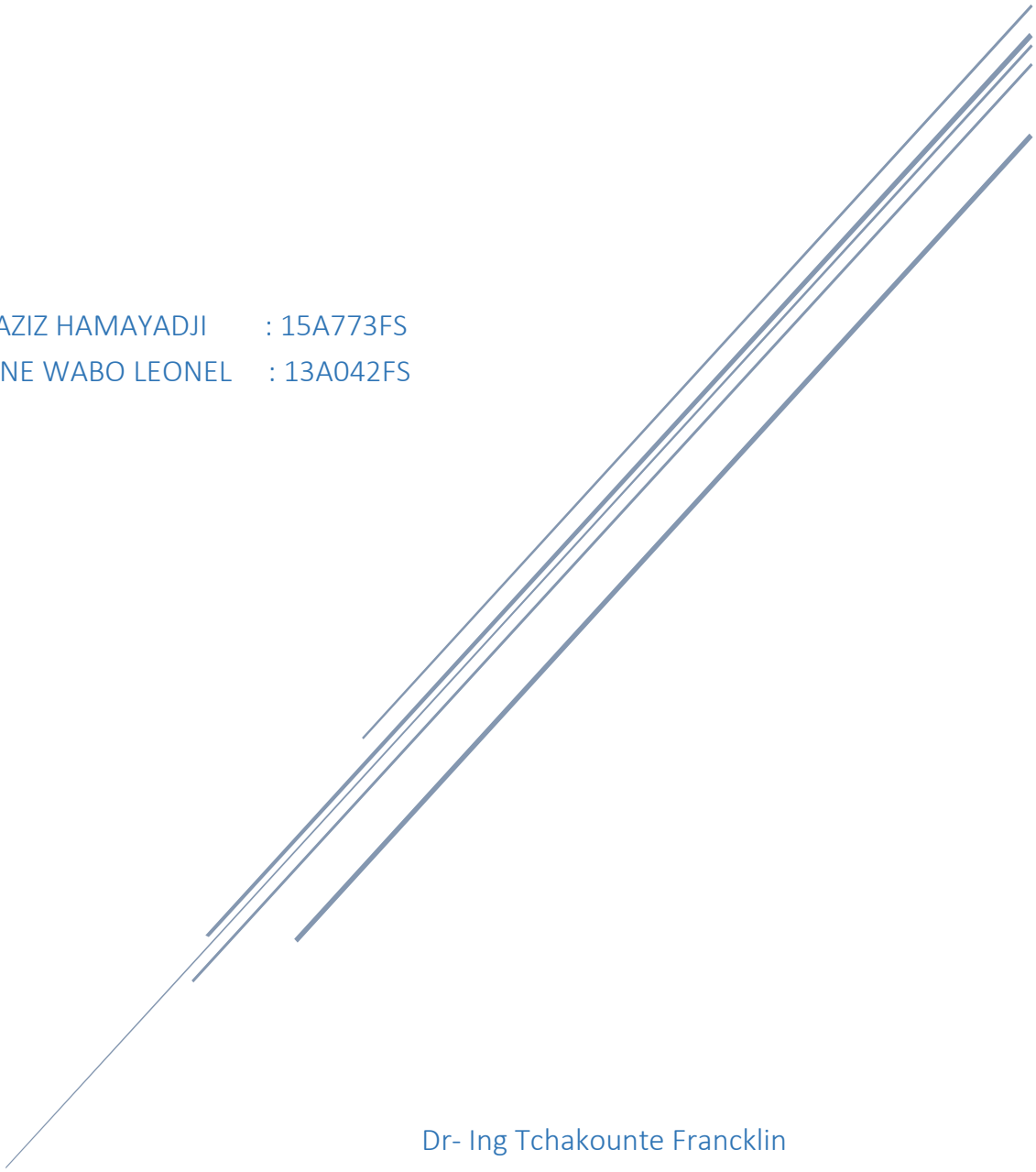
STUDENT PERSONAL WORK

Topic: Machine Learning Methods for Computer Security

REVISITED DOCUMENT

25/03/2017

ABDOUL AZIZ HAMAYADJI : 15A773FS
KANMOGNE WABO LEONEL : 13A042FS



Dr- Ing Tchakounte Francklin

University of NGAOUNDERE 2016/2017
Systems and Networks security

1. Introduction

Context: Nowadays, computation trend and data stream generated become more and more important and complex due to the democratization of internet which permits among other the high availability of many range of services. Therefore, human being is rapidly overflowed and need help of machine learning technology. Learning methods are being used more and more to obtain a better understanding of various data collected from complex systems. The document "*Machine Learning Methods for Computer Security*" on which this resume is produced, is the result of an inquiry made by a workshop between two big research domains which are **computer security** and **machine learning**; these two communities convened to bring together interested researchers to discuss **techniques**, **challenges**, and **future research directions** for secure learning and learning-based security applications, the topic of secure learning is emerging as a major direction of research that offers new challenges to both communities. The main researchers are: *Anthony D. Joseph, Pavel Laskov, Blaine Nelson, Fabio Roli, and J. Doug Tygar*.

Background: In the purpose to well understand the heart of the topic, it is important that the concept of **machine learning method** which is an approach of learning, that companies and enterprises which control or manage a huge stock of data stream use to predict and provide good quality of service and response. Therefore, the machine have learn at certain time some patterns of recognition before start. Another important concept is **computer security**. Security can be define in general as a state of being protected against malicious program, or all other elements that also try to attempt the safety of an entity. Behind those two main concept of our topic there are deep concept as **data analysis**, **vulnerabilities**, **profiling**, and **recommendation system** among others.

Problem and research questions: We know that Internet owes its success to the enormous amount of data it generates and to novel decision-making instruments based on data analysis. Unfortunately, the ubiquity of the Internet has also stimulated its abuse and the rise of sophisticated cyber-crimes. Therefore, some people are interested to this kind of crime and build a business around, that rely on the exploitation of security vulnerabilities. Moreover, everyday attackers produce a lot of new exploitation techniques in the aim to avoid the detection mechanism of security. That reveal the problem associated to the data analysis, because, originally machine learning was conceived under the assumption of "faithful" data and did not explicitly account for potential data manipulation by adversaries. In the aim to deal with this problem, researchers come from various disciplines ranging from machine learning and security to spam filtering, online advertisement and computer forensics.

The workshop divided the topic in three themes of discussion and carry those following research questions: **What security problems can machine learning best help to solve? What scenarios are they ill-suited for? What are the theoretical limitations of worst-case attacks against learning algorithms under different constraints? How can these constraints be used in practice for protecting learning methods against adversarial data? What are existing and emerging non-security applications where learning techniques are used and can potentially be exposed to adversarial data? What experience from these applications can be used for development of general methodology of secure learning?**

Before moving to the next part, the authors underline that most of security-related decisions involve a human operator.

2. Related works

Authors	Topics	Machine learning for security (overview)	Secure machine learning (algorithms)	Different methods for security in machine learning	Security requirements	Open issues	Year
Anthony D. Joseph; Pavel Laskov; Fabio Roli; J.Doug Tygar; Blaine Nelson	Machine learning methods for computer security	✓	✓	✓	✓	✓	2016
Sandeep V. Sabnani Supervisor: Professor Andreas Fuchsberger	Computer Security: A Machine Learning Approach	✓	✓	-	-	✓	2016
Pavel Laskov And Marius Kloft.	A Framework for Quantitative Security Analysis of Machine Learning	-	✓	-	✓	✓	2016
Anna L. Buczak And Erhan Guven	A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection	✓	✓	-	✓	✓	2016
Tadeusz Pietraszeka And Axel Tannera	Data Mining and Machine Learning Towards Reducing False Positives in Intrusion Detection	-	✓	-	✓	✓	2016
Marco Barreno; Blaine Nelson; Anthony D. Joseph And J.D. Tygar	The security of machine learning	✓	-	✓	-	✓	2016
Philip K. Chan And LIPPMANN@LL.MIT.EDU	Machine Learning for Computer Security	✓	✓	-	-	✓	2016
Ankit Kumar Jain, And B. B. Gupta	Comparative analysis of features based machine learning approaches for phishing detection	✓	✓	✓	-	✓	2016
Michael Bailey, Jon Oberheide, Jon Andersen, Z. Morley Mao, Farnam Jahanian, and Jose Nazario.	Automated classification and analysis of internet malware. In Recent Adances in Intrusion Detection (RAID), pages 178–197	✓	✓	-	-	-	2007

Marco Barreno, Blaine Nelson, Anthony D. Joseph, and J. D. Tygar.	The security of Machine learning. Machine Learning, 81(2):121–148,	✓	✓	✓	✓	✓	2010
Steffen Bickel and Tobias Scheffer	Dirichlet-enhanced spam filtering based on biased samples. In Neural Information Processing Systems (NIPS), pages 161–168	-	✓	-	✓	✓	2007
Battista Biggio, Giorgio Fumera, Ignazio Pillai, and Fabio Roli	A survey and experimental evaluation of image spam filtering techniques. Pattern Recognition Letters, 32:1436–1446	-	✓	✓	-	✓	2011
Nedim Šrndić and Pavel Laskov	Evasion resistant detection of malicious PDF files based on hierarchical document structure. In Network and Distributed System Security Symposium (NDSS),	-	-	✓	✓	✓	2013

✓ : This symbol denote that the specific domain is covered.
- : This symbol denote that the specific domain is not covered.

3. Methodology employed

In this article “Machine Learning Methods for Computer Security”, the authors focused on three major points which are:

The first is “**machine learning for security**” in this part, authors emphasize the rapid development of security exploits in recent years which has fueled a strong interest in data analysis tools for computer security. On the one hand, the sheer number of novel malicious software observed by security researchers transcends the limits of manual analysis, for example according to AVTEST, more than 200 000 of new malware sighted daily. They mention that the conventional methods based on hashes, signatures, or heuristic rules cannot deal with certain timely threats that need to turn to other methods such as detection based on anomalies that appear to be the best alternative for such cases, although they inevitably cause some false positives. In this part also they present some application of security based on the machine learning in particular:

- Detection of malicious activity in operating system data, or network traffic: "intrusion detection systems". Much of the work in intrusion detection has followed various learning-

based approaches, in particular, the anomaly detection rule inference and supervised learning (eg: Snort and Bro)

- Another critical contribution of learning-based systems is in the area of dynamic analysis of malware. Such systems have succeeded in gathering large amounts of data, which leads to an urgent need for tools to automatically analyze new malware.

- Automatic signature generation (ASG), which is based on automatic learning and malware analysis. The first ASG ad hoc systems combined the extraction of String tokens with frequency analysis, these ASG methods have proved very useful for several related problems, including botnet network detection and reverse engineering network protocol which Are less subject to contradictory data.

-Static analysis of JavaScript token sequences has been successfully deployed to detect malicious PDF documents driven by the reader and documents with JavaScript. Dynamic analysis of JavaScript channel allocations combined with the payload classification of channels has been successfully used in the browser to detect browser downloads.

The Second is "**Secure Learning: Theory and Methods**" In this section, the authors focused on the limitations and weaknesses of some learning algorithms:

They start with the limitations of traditional learning methods that are not designed to function in the presence of adversaries because these methods rely on statistical assumptions about the distribution of input data and are based on training data derived from Input data to construct analysis models. An attacker can exploit these features to interrupt the scan, cause the scan to fail, or engage in malicious activities that are not detected.

Then they explain that the learning algorithms can be surprisingly tolerant of bad labels attributed by an attacker during training based on the work of *Angluin* and *Laird* who have shown that for some classes of boolean functions an attacker must return Labels of nearly half of the training data to cause incorrect classification of the selected data points. On the other hand, by controlling the attributes of the data, an attacker can construct intrinsically difficult learning problems, in which the error rate can be almost as high as the data fraction under attacker control. These contrasting results emphasize the necessity of assumptions about contradictory power. Finally they emphasize that recent work on the safety of machine learning have provided more evidence of potential attacks but have given little details on how they can be avoided.

And the third is "**Secure learning beyond security**" In this last part of the theme, the author presents the analysis of the main factors that can make the specific application areas attractive for new attacks against learning methods and investigate the situation in several areas where learning technologies play a crucial role:

- Spam filtring: They draw attention to the fact that many modern email clients have automatic spam filtering that partially integrates automatic learning techniques, demonstrating both its scientific relevance and the analysis of Cost effectiveness of this application.

- Social media spam: With recent developments in spam detection, spam senders have changed their targeted communication medium. Nowadays, social networks, social media websites and referral services such as Facebook, Google+, MySpace, Twitter, YouTube, Flickr, SoundCloud, TripAdvisor, Amazon and so on are the main target of spammers.

- Detection of plagiarism and identification of the author: Plagiarism is a long-standing scourge in academics and the media that can ruin a person's scientific and even political career (here the author highlights the demonstration of Karl-Theodor zu Guttenberg and Annette Schavan)

4. Result found and Perspectives

4.1 Result found

i. machine learning for computer security

A synergy between machine learning and malware analysis has been exploited for **automatic signature generation** (ASG), the key weakness of such methods was their susceptibility to attacks, which made it possible to evade a deployed system or increase its false alarm rate.

The recent **emergence of JavaScript-based** exploitation schemes such as drive-by-downloads and malicious non-executable files raises a number of new challenges for detecting such threats.

ii. Secure learning: theory and methods

Evasion attacks against anomaly-based intrusion detection systems (IDS) have demonstrated two main evasion strategies: ***poisoning*** and ***mimicry***. In the development of learning algorithms with robustness to worst-case noise, two groups of methods should be noted: The **min-max approach** and the **more complex game-theoretic approach** such problems can be solved using: the Nash equilibrium approach. So the most progress in the study of security of machine learning has been made in identifying the security weaknesses of learning methods.

It is not possible to know in advance what kinds of attacks a given learning algorithm or classifier system will be subject to, as well as their characteristics. Any attack scenario implies that training and testing datasets follow different distributions. Any attack scenario implies that training and testing datasets follow different distributions.

Ideally, security metrics for ML systems will provide:

- A mechanism for inter-algorithm comparison,
- The ability to provide strong performance guarantees, and
- A mechanism for determining whether an algorithm is appropriate for use in a particular security setting.

iii. Future application of secure learning

The key factor in the assessment of the relevance of secure learning to specific applications is the existence of **a business case**. As experience shows, serious security threats arise mostly when miscreants can make money by exploiting technical vulnerabilities.

Beside distributing "classical" spam, these platforms are also used

- to attract attention by heavily interacting with other users,
- to create fake reviews and thus manipulate recommendations,
- to build fake friend networks to manipulate rankings,
- to start rumors and harass.

Social media spam differs from traditional email spam in that spammers can exploit many different functions of these websites including posting/reposting comments, sending private or chat messages, following/re-following other users, favoring items, and uploading images/videos/sounds/documents/applications.

4.2 PERSPECTIVES

The big motivation of attacker is the existence of a business case, but attack targets may also be **intellectual property** or **classified information**, as the cases of cyber-espionage. The future application is:

- **Spam filtering**: this is the most popular example of machine learning applications that has to deal with adversarial inputs. Many modern email clients have an automatic spam filtering function that partially incorporates machine learning techniques, thus proving both its scientific relevance of and the business case for this application.
- **Social media spam**: With the recent developments in junk email detection, spam senders changed their targeted communication medium. Nowadays, social networks, social media websites and recommendation services such as Facebook, Google+, MySpace, Twitter, YouTube, Flickr, SoundCloud, TripAdvisor, Amazon, and so forth are the major target for spammers.
- **Plagiarism Detection and Authorship Identification**: it has been a long-standing plague in academics and media. The discovery of plagiarism may ruin a person's scientific and even political career.
- **Copyright Enforcement**: Apart from detecting plagiarism and duplicated content in textual documents, social media platforms facilitate the sharing of media files such as images, videos and music, which also must be monitored for copyright infringements.

Recommendation and research priority

- **Poisoning**
  Poisoning refers to attacks against machine learning, in which an attacker can insert manipulated data before or during the training phase of a learning system. Poisoning has been identified as a serious threat for specific intrusion detection techniques.
- **Penetration Testing**
  This type of security assessment has become a de facto requirement for the majority of web-applications, especially in the financial industry.
- **Case Studies**
  Being largely an empirical field of science, machine learning typically entails large-scale case studies for the verification of results.
- **Privacy issues**
  Many datasets in computer security and related areas are subject to serious privacy restrictions.

- **Non-stationarity of the data**
  As attackers respond to a particular predictive model, static datasets are insufficient; however, dynamic data is difficult to generate under realistic conditions.
- **Lack of ground truth**
  In contrast to many other applications of machine learning, ground truth data is hard to obtain in adversarial environments. To identify common applications and use cases, it is necessary to formally describe non-adversarial activities for each case, potentially using domain-expert knowledge, and to treat deviations from these allowed activities as adversarial events.

5. Most important publication for this work

[10] Marco Barreno, Blaine Nelson, Russell Sears, Anthony D. Joseph, and J. D. Tygar. Can machine learning be secure? In ACM Symposium on Information, Computer and Communications Security (ASIACCS), pages 16–25, 2006.

[44] Prahlad Fogla, Monirul Sharif, Roberto Perdisci, Oleg Kolesnikov, and Wenke Lee. Polymorphic blending attacks. In USENIX Security Symposium, pages 241–256, 2006.

[60] Marius Kloft and Pavel Laskov. Online anomaly detection under adversarial impact. In International Conference on AI and Statistics (AISTATS), pages 405–412, 2010.

[96] Roberto Perdisci, David Dagon, Wenke Lee, Prahlad Fogla, and Monirul Sharif. Misleading worm signature generators using deliberate noise injection. In IEEE Symposium on Security and Privacy, pages 17–31, 2006.

[122] Olivier Thonnard. A Multicriteria Clustering Approach to Support Attack Attribution in Cyberspace. PhD thesis, Ecole Doctorale d'Informatique, Télécommunications et Electronique de Paris, 2010.

[126] Ke Wang and Salvatoro J. Stolfo. Anomalous payload-based network intrusion detection. In Recent Adances in Intrusion Detection (RAID), pages 203–222, 2004.