Detection, prevention and simulation approaches to address anomalies in cyber security

Luciano Argento
Supervised by: Fabrizio Angiulli and Angelo Furfaro
University of Calabria, DIMES, P. Bucci, 41C I-87036 Rende (CS), Italy
{luciano.argento,f.angiulli,a.furfaro}@unical.it
http://www.unical.it/portale/international/

Abstract. The cyber security landscape is heavily affected by a significant number of severe problems, from developers that do not properly follow guidelines for the development of secure coding best practices to worldwide diffusion of malware toolkits that allow even users with little expertise to perform a cyber attack. A very important aspect concerning cyber security is the steady evolution cyber threats are subjected to. As a consequence, such threats become more and more sophisticated, attack after attack. A lot of efforts are required in order to improve the state of cyber security and keep up with the evolution of cyber threats, and that means adapting existing security solutions in response to the ever-changing security landscape and devising new solutions when needed. In this paper we present the research activities conducted during the Ph.D., which find their place in the complex scenario previously described. We investigated relevant cyber security problems related to data analysis and anomaly detection, in the following area of research: Hybrid Anomaly Detection Systems; Intrusion Detection Systems; Access Control Systems and Internet of Things.

Keywords: Cyber Security · Anomaly Detection · Hybrid Anomaly Detection Systems · Intrusion Detection Systems · Access Control Systems · Internet of Things.

1 Introduction

With the massive adoption of the Internet both our private and working life has drastically changed. The Internet has introduced new ways to communicate and complete every day tasks. Organisations of any kind have taken their activities online to achieve many advantages, e.g. commercial organisations can reach more customers with proper marketing. However, the Internet has also brought various drawbacks and one of these concerns cyber security issues. Whenever an entity (e.g. a person or company) connects to the Internet it immediately becomes a potential target of cyber threats, i.e. malicious activities that take place in cyberspace. Examples of cyber threats are theft of intellectual property and denial of service attacks. Many efforts have been spent to make the
Internet perhaps the most revolutionary communication tool ever created, but unfortunately little has been done to design it in a secure fashion. Since the massive adoption of the Internet we have witnessed a huge number of threats, perpetrated by many different actors such as criminal organisations, disgruntled workers and even people with little expertise, thanks to the existence of attack toolkits. On top of that, cyber threats are constantly going through a steady evolution process and, as a consequence, they are getting more and more sophisticated. Nowadays, the cyber security landscape is in a critical condition. It is of utmost importance to keep up with the evolution of cyber threats in order to improve the state of cyber security. We need to adapt existing security solutions to the ever-changing security landscape and devise new ones when needed.

The research activities presented in this paper find their place in this complex scenario. We investigated significant cyber security problems related to data analysis and anomaly detection, in different areas of research: Hybrid Anomaly Detection Systems; Intrusion Detection Systems (or IDS); Access Control Systems and Internet of Things. We designed and developed a generic framework for anomaly detection, called HALF, to build Hybrid Anomaly Detection Systems. HALF is not bound to any specific application domain and its objective is to provide a flexible and more effective detection capability. In the context of intrusion detection, we conducted two research activities. Specifically, the activities focused on issues related to the employment of the $n$-gram technique in anomaly-based intrusion detection systems for detecting content-based attacks. The first activity led to the design and development of a novel anomaly-based intrusion detection technique, called PCkAD, while in the second activity we devised an approach to build compressed $n$-gram-based classification model. We proposed a machine learning approach to dynamically refine and update policies, respectively, to prevent insider threats and to automate policy administration. At last, with regard to IoT, we showed how to effectively assess cyber security scenarios involving IoT settings by combining novel virtual environments, agent-based simulation and real devices.

The paper is organised as follows. Section 2 presents HALF. Section 3 discusses the contributions related to intrusion detection. Section 4 talks about the machine learning approach proposed for access control and Section 5 our contribution for the Internet of Things.

2 Hybrid anomaly detection systems

Anomaly detection approaches are very relevant in the field of cyber security. Fraud and intrusion detection are well-known research areas where such approaches are very important. A lot of techniques have been devised, which can be categorised in anomaly and signature based detection techniques. Researchers have also spent much effort on a third category of detection techniques, i.e. hybrid anomaly detection, which combine the two former approaches in order to obtain better detection performances. However the existing works on hybrid anomaly detection from the literature suffer from a few limitations, in that they
are typically bound to specific techniques and types of data. To address these limitations, we designed a generic framework, called HALF [2], whose goal is to accommodate multiple mining algorithms of a specific domain and provide a flexible and more effective detection capability. HALF can be easily employed in different application domains such as intrusion detection and steganalysis due to its generality and the support provided for the data analysis process. We analysed two case studies in order to show how HALF can be exploited in practice to implement a Network Intrusion Detection System and a Steganalysis tool.

3 Intrusion Detection Systems

The concept of anomaly is a core element of the research activity conducted in the context of intrusion detection, where an intrusion can be seen as an anomalous activity that might represent a threat to a network or system. Intrusion detection systems constitute a very important class of security tools which have become an invaluable defence wall against cyber threats. In the following two sections, we present two research results that stem from issues related to IDSs that resort to the $n$-grams technique for the detection of content-based attacks.

**PCkAD: Packet Chunk Anomaly Detector.** Anomaly-based IDSs that employ the $n$-gram technique have achieved satisfactory results in detecting intrusions, over the years. However, counting on the $n$-grams alone might not be sufficient to create a reliable intrusion detection system. Indeed, legitimate $n$-grams may also appear in malicious data instances and their presence might make the instance appear as normal. This is an aspect that has not properly addressed by existing solutions. Another very important aspect to consider is the presence of an adversary attacker. Unfortunately, in literature many works do not address this aspect with sufficient efforts while others do not take it into account at all. We devised a novel anomaly-based intrusion detection technique [1], called PCkAD, that models legitimate traffic on the basis of the spatial distribution of the legitimate $n$-grams and has been designed to be resistant to blending evasion techniques. Indeed, we demonstrate that evading is an intrinsically difficult problem. The experiments conducted to evaluate PCkAD show that it achieves state-of-the-art performances in real attack scenarios and that it performs well against blending attacks.

**An approach to compress $n$-gram-based models for novelty detection.** The second contribution concerning intrusion detection investigates issues that may be brought by the employment of the $n$-gram technique. Many approaches using $n$-grams have been proposed in literature which typically exploit high order $n$-grams to achieve good performance. However, because the $n$-gram domain grows exponentially with respect to the $n$-gram size, significant issues may arise, from the generation of huge models to overfitting. We developed an approach aimed to reduce the size of $n$-gram-based models, which is able build models that contain only a fraction of the original $n$-grams with little impact on the detection accuracy. We evaluated the effectiveness of the compressed models on a real word dataset, in the context of intrusion detection. The experimental
results are promising: the models are able to distinguish between normal and malicious packet payloads in a satisfactory way, by using only 5% of the total n-grams that characterise the network traffic.

4 Access Control Systems

The research concerning access control systems focused on anomalies that represent attempts of exceeding or misusing access controls to negatively affect the confidentiality, integrity or availability of a target information system. Access control systems are nowadays the first line of defence of modern computing systems. However, their intrinsic static nature hinders autonomously refinement of access rules and adaptation to emerging needs. Advanced attributed-based systems still rely on mainly manual administration approaches and are not effective on preventing insider threat exploiting granted access rights. We proposed a machine learning approach to refine attribute-based access control policies based on behavioural patterns of users’ access to resources. The designed system, called ML-AC, tailors a learning algorithm upon the decision tree solutions. We analysed a case study to show the capabilities of ML-AC and conducted an experiment to evaluate the effectiveness of the system. The results were promising.

5 Internet of Things

IoT is showing the potential for impacting several domains, ranging from personal to enterprise environments. IoT applications are designed to improve most aspects of both business and citizens’ lives, however such emerging technology has become an attractive target for cybercriminals. A worrying security problem concerns the presence of many smart devices that have security holes. Researchers are investing their efforts in the evaluation of security properties. Following this direction, we show that it is possible to effectively assess cyber security scenarios involving IoT settings by combining novel virtual environments, agent-based simulation and real devices [3] and then achieving a means that helps prevent anomalous actions from taking advantage of security holes for malicious purposes. We demonstrated the effectiveness of the approach through a case study regarding a smart home setting.

References