A proposal for a new way of classifying network security metrics. Study of the information collected through a honeypot

Alejandro Carrasco, Jorge Ropero. Paulino Ruiz de Clavijo, Jaime Benjumea
Department of Electronic Technology
Universidad de Sevilla
Spain
acarrasco@us.es, jropero@dte.us.es, paulino@dte.us.es, benjumea@dte.us.es

Amalia Luque
Department of Design Engineering
Universidad de Sevilla
Sevilla
amalia@us.es

***Abstract*—Nowadays, honeypots are a key tool to attract attackers and study their activity. They help us in the tasks of evaluating attacker's behaviour, discovering new types of attacks, and collecting information and statistics associated with them. However, the gathered data cannot be directly interpreted, but must be analyzed to obtain useful information. In this paper, we present a SSH honeypot-based system designed to simulate a vulnerable server. Thus, we propose an approach for the classification of metrics from the data collected by the honeypot along 19 months.**

Keywords—Honeypot, Metric, Kippo, Network Security, IDS

I. Introduction

Year after year, the security attacks on the Internet have been increasing notably. Security sytems as IDS (Intrusion Detection System) or IPS (Intrusion Prevention Systems) have been developed to detect and prevent attacks and malicious behaviors [1]. For this purpose, a honeypot is a tool that provides an active defense against intruders [2], creating servers that are attractive for attackers. Besides, honeypots allow intruders to have access to ficticious information in order to analyze their main goal, and finally come to conclusions about their nature. This monitorization prevents network administrators from future attacks.

Our main goal is to identify and classify attacks from the large amount of data gathered through the honeypot. Additionally, direct information may be obtained with simple queries to the database; for instance, the number of total sessions or different pairs user/password used. However, how to interpret that information is a key aspect, considering the large amount of selected data. Data mining tecniques and graphics help us to obtain a quick and detailed global vision of what can be found in a honeypot [3]. In spite of the fact that honeypots are currently widely spread, only a few detailed statistical reports are publicly available [4]. Fortunately, most of these available works follow common patterns in order to analyze the results. From now on, we propose to call these patterns metrics.

In this paper, we suggest a general method for data classification that may be obtained from IDSs. We also recommend the creation of a test scenario with a real implementation of the most significant metrics in our opinion. Furthermore, we study and compare our results with previous studies of the basic metrics, and we also implement and interpret some of the new metrics.

The final data, resulting from a 19-month collection work, makes this paper a good reference for future studies.

II. Definition of metrics

As mentioned above, only a few detailed statistical reports about IDSs are publicly available. The metrics we propose represent the metrics for the collected data. Metrics showing the most used combinations user/password, the number of attackers' attempts, and the IP addresses of the intruders are repeated in most of the analysis [3-7]. We can also find studies about the executed commands [3, 4, 6] as well as metrics about downloads and geolocation [4, 6-9].

After analyzing all these unstructured sources of information, we decide to propose a set of metrics to assess the data collected in our scenario, which is descibed in Section 3. The new metrics are grouped in four groups, based on the distribution of Kippo database tables. The four identified metrics are: Login (LOG), Inputs (INP), Downloads (DOW) and Geolocation (GEO). We have named all the metrics that are extracted from the Kippo database tables "Basic metrics", and they could be a reference for future a more homogenous future analysis.

III. Description of the scenario

In this paper, we describe a scenario where a honeypot simulates a vulnerable SSH server, named Kippo. Kippo offers a remote access, as well as logs how attackers access the server and records the credentials they have used together with their activity, once into the server. Kippo is a low-

This research is supported by the *Telefónica Chair in "Intelligence in Netwoks" of the University of Seville*.

interaction honeypot written in Python that emulates a SSH shell [5].

A Virtual Private Server (VPS) is also installed and the system runs into an OpenVZ container. The server does not contain any aditional security components, what makes it vulnerable to achieve more intrusions. SSH server configuration is changed to use port 22. This also happens in Kippo configuration file (kippo.cfg), which is listening at port 2222 for testing. Authbind is used to run kippo at port 22 with user permissions. Otherwise, users should execute Kippo as root to access the low-numbered ports. So our server is more visible for network users. Besides, a weak password is set (user: root; password: 123456). The events collected by Kippo are recorded into a MySQL database, instead of using the default text files, in order to achieve a more flexible way of managing data.

This scenario is continuously operating from 20th July 2014 to 4th February 2016, thus we have 19-month attack information. Figure 1 shows the network topology.

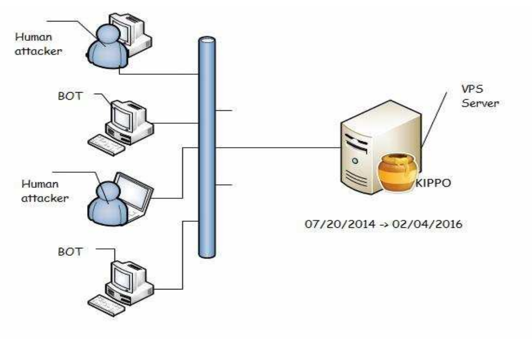

Fig. 1. Network Topology

IV. Data analysis

In this paper, we have firstly defined a set of metrics, in order to study the nature of the attacks. Then, we have classified metrics in four groups: login metrics, input metrics, download metrics and geolocation metrics. Besides, we have built a system that could simulate a SSH shell, based on Kippo, in order to study the attacks. The scenario has been functioning for 19 months what has enabled us to obtain a relevant amount of data. We have recorded up to 340,572 login attempts, consequently our study constitute the most complete one in the literature about this topic.

To summarize, more than 340,000 login attempts have been registered in 19 months, making possible to evaluate the attacks in different seasons. We found that attacks usually increase in holiday seasons. Techniques used for attacks vary, but organized brute force attacks have been common. As prior existing studies let us know, the password and the username coincide in many cases, though the results are quite heterogenous, due to the aforementioned brute force attacks. “Root” has been the most frequent username by far. Despite a later increase of Windows users, Linux has been the most utilized operating system.

According to the input metrics, an average of 1.08 commands per user have been registered, what represents very specific missions from attackers. Moreover, we have observed that the higher success rate is, the fewer executed commands are. Particularly, *Passwd* (in order to change the password) has been the most used command, while *ls* was the second one. *Echo* commands, the modification of Linux firewall stand among the most common commands, too. Getting information about the directory structure (*cd* and *ls*) is quite important for attackers, but it is curious how intruders try to modify the system with commands like *usermod, yum, password, useradd, chmod* or *service*. This means that there is a lot of “bad intention” on their part. Instructions to install malicious sofware have been detected (for example, denial-of-service attacks or adding our system to a botnet). They have even tried to install game server software or packs related to python, php, or teamviewer. Intruders have entered 694 passwords using the command *passwd*. 600 out of them were not filled in.

Another typical action is file downloading. Download metrics show that executable files have been the most downloaded ones. HTML or ZIP files are also quite present.

Finally, it is worth mentioning that most of the attacks have come from China, although the system has identified intruders from 74 different countries.

References

[1] W.Li, W. Meng, X. Luo, L.F. Kwok. (2016). MVPSys: Toward practical multi-view based false alarm reduction system in network intrusion detection. *Computers & Security*, Vol. 60, pp. 177-192.

[2] C. De Faveri, A. Moreira (2016). Desining Adaptive Deception Strategies. Proceedings of the 2016 IEEE International Conference on Software Quality, Reliability and Security Companion, QRS-C 2016, Vienna (Austria), pp. 77-84.

[3] I. Koniaris, G. Papadmitriou, P. Nicopolitidis P (2013). Analysis and visualization of SSH attacks using honeypots. *Eurocon 2013*, Zagreb, Croatia, pp. 65-72.

[4] T. Sochor, M. Zuzcak (2014). Study of Internet Threats and Attack Methods Using Honeypots and Honeynets. *Communications in Computer and Information Science,* Vol. 431, pp. 118-127.

[5] D. Ramsbrock, R. Berthier, M. Cukier (2007). Profiling attacker behavior following SSH compromises. *37th Annual IEEE/IFIP Int Conf on Dependable Systems and Networks, DSN 2007*, Edinburgh, UK, pp. 119-124.

[6] C. Valli (2012). SSH – Somewhat Secure Host. *4th Int Symposium on Cyberspace Safety and Security, CSS 2012*, Melbourne, Australia, pp. 227-235.

[7] M. Marchese, R. Surlinelli, S. Zappatore (2011). Monitoring unauthorized internet accesses through a honeypot system. *International Journal of Communication Systems*, Vol. 24, n. 1, pp. 75-93.

[8] V. Visoottiviseth, U. Jaralrungroj, E. Phoomrungraungsuk, P. Kultanon (2011). Distributed Honeypot log and visualization of attacker geographical distribution. *8th Int Joint Conf on Computer Science and Software Engineering, JCSSE 2011*, Nakhron Pathom, Thailand, pp. 23-28.

[9] K. Goseva-Popstojanova, G. Anastasovski, A. Dimitrijevikj, R. Pantev, B. Miller (2014). Characterization and classification of malicious Web traffic. *Computers and Security*, Vol. 42, pp. 92-115.