Study of Internet Threats and Attack Methods Using Honeypots and Honeynets

Tomas Sochor and Matej Zuzcak

University of Ostrava, Ostrava, Czech Republic
tomas.sochor@osu.cz
http://www1.osu.cz/home/sochor/en/

Abstract. The number of threats from the Internet has been growing in the recent period and every user or administrator should protect against them. For choosing the most suitable protection the detailed information about threats are required. Honeypots and honeynets are effective tools for obtaining details about current and recent threats. The article gives an introduction into honeypots and honeynets and shows some interesting results from initial 3-months period of the implementation of a small honeynet made of 3 Dionaea and one Kippo low-interaction honeypots. Basic conclusions regarding the amount of currently actively spread malware and their type are formulated.

Keywords: computer attack, Dionaea, honeynet, honeypot, Internet threat, Kippo, low-interaction honeypot, malware.

1 Honeypot Application Introduction

Despite growing popularity of honeypot and honeynet application only few detailed statistical reports from their application are publicly available. The main reason is probably the risk of misuse of such reports by potential attackers as well as financial aspect. Among few exceptions there are DenyHosts [1], Dshield.org [2], Honey Pot Project [3], Shadowserver [4] and HoneyMap [5]. All the mentioned project have a common drawback that is the lack of data available for third-party analyses. Some of them are not up-to-date as well. In the recent past there was a project called honeynet.cz publishing a lot of reports thus being widely appreciated in Europe and worldwide but it is no longer available.

The aim of the study described here was to identify suitable honeypot implementations (with focus to low-interaction server honeypots) and test them in practical implementation of our own honeynet. Subsequent evaluation of gathered data is inevitable part of the study as well. The main focus of the study is the protection of local IP-based networks using honeypots.

2 Honeypot and Honeynet Classification

Honeypots are classified into various categories and not all publications apply the same approach. The first and the most obvious classification is based on

A. Kwiecień, P. Gaj, and P. Stera (Eds.): CN 2014, CCIS 431, pp. 118–127, 2014.

the activity of the honeypot. Therefore honeypots can be either *passive* that simulates a server by offering some (vulnerable) services and is just waiting for an interaction from an attacker. The other approach (*active* honeypot) means that the honeypot simulates a client actively looking for available services. When an active honeypot finds a server then it interacts with the server simulating a client software like (vulnerable) www browser, e-mail client etc.).

Honeynet is a logical network of several honeypots (either connected to a single physical network or to multiple networks interconnected using the Internet). The purpose on honeynets is to improve monitoring of detected threats and to explore them in a more efficient way. The typical implementation of a honeynet for research purposes is shown in Fig. 1.

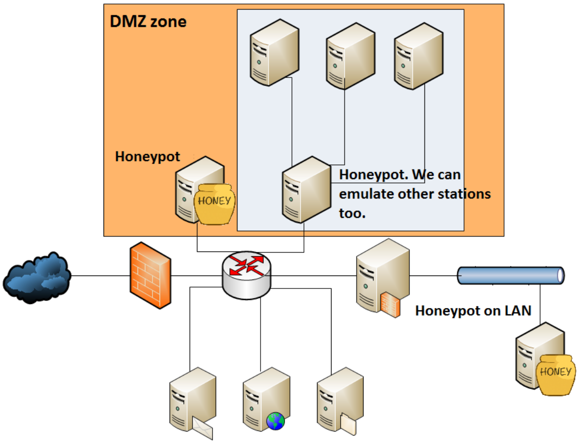

Fig. 1. Example of possible research honeypot implementation in a production network

Honeypots are commonly classified into two main categories according to the level of their interaction (in accordance to [6], Chap. 1.4.2):

- low-interaction honeypots,
- high-interaction honeypots.

2.1 Low-Interaction Honeypots

Low-interaction honeypots are based on network services emulation. Their aim is to achieve the active connection by an attacker. After the attacker's successful connection the honeypot will perform a predefined action (e.g. it reacts by displaying a predefined banner message, it downloads a malware specimen intended to store on the attacked server by the attacker etc. Low-interaction honeypots offer a limited scope of activities that can be performed and limited number of active connection as well. On the other hand it present very safe solution because an attacker is not offered with the whole operation system but they have an access to an emulated service only. This service is modified so that the attacker has a feeling that they attack a real running service with productive value. The main aim of such honeypot is to gain as much information about an attacker as possible.

This type of honeypots is the most frequently used as production honeypots, their administration is much easier than high-interaction honeypots, their operating costs are much lower as well, and the their application does not pose any risk. On the other hand they can catch only automated attacks that have been known in advance. Their typical use is for gathering statistical data, mapping threats, to distract attackers from production systems, or for fine-tuning of rules for safety tools.

2.2 High-Interaction Honeypot

High-interaction honeypot provide the access to the whole operating system including services and applications to an attacker. In most cases it is implemented as a virtualized system that allows easy and quick restore after possible modification. When this solution is used also some risks are present and their operation requires more administration. The system should be monitored in a detailed way so that any activity of an attacker could be recorded, analyzed and evaluated. The system must be well secured using a firewall and an IPS system that is adapted so that it cannot be abused by any attacker e.g. for a targeted attack or its integration into a botnet. On the other hand high-interaction honeypots are very beneficial for research, they can be used to identify zero-day vulnerabilities, to reveal a new malware as well as highly sophisticated targeted attacks.

2.3 Shadow Honeypots

Shadow honeypots present a specific subgroup of production (low-interaction) honeypots. The operate in combination with an ADS (Anomaly Detection System). If the ADS detects certain anomalies in the incoming network traffic, it is subsequently redirected to the honeypot instead of its blocking. The traffic is further analyzed on the honeypot. The legitimate traffic is segmented from a traffic containing certain anomalies this way. The benefit of this solution is in the fact that it can detect new or highly sophisticated attacks focused to real production systems.

3 Research Methods

There are various possible approaches applicable to the study of honeypots but the direct measurement of number of attacks and their nature prevails. One of recent studies [7] is a nice example of this approach that is applied by our study, too. The results presented in the paper [7] confirm the majority of our conclusions. Also other recent studies (e.g. [8] are in accordance to our results where applicable.

Two types of low-interactive honeypots, i.e. Kippo [9] and Dionaea [10], were used for gathering data about threats. Kippo is a low-interaction honeypot written in Python that emulates a SSH shell. It was inspired by the Kojoney honeypot. It contains false file system that is able to emulate adding and deleting files. It was inspired by debian Linux OS. The user can download new files (e.g. using wget command) and use cat command, too. The downloaded files are recorded. The most of Linux commands and tools are not implemented (emulated) however and if an attacker tries to use them an error message is produced. All attacks are logged with proper timestamps.

Dionaea was configured to support all emulated services available (including SMB, http, ftp, tftp, SIP, and database services of MS SQL and MySQL). Certain additional modules (e.g. VirusTotal, sandboxes, XMPP) were also used but with no influence to the measurements. After data gathering they were processed and statistics were prepared in the form of tables and diagrams (see below in the Sect. 5).

The reason for using two honeypots is the following. The Dionaea honeypot emulates the vulnerabilities of Windows operating system, especially focusing the SMBD protocol. The results from Dionaea are affected by the fact that it cannot emulate other services at the required level of quality, however.

On the other hand attackers focusing to Linux presented also an object of interest. Their common modus operandi is that they penetrate the system via a weak SSH password. This is what Kippo honeypot emulates: namely such system that looks like self-contained application. In Kippo only the SSH service (port 22) is emulated and after system penetration only activities in the system are analyzed. But thanks to the fact that Kippo is a low-interaction honeypot, the attacker is allowed to do hardly anything. No command is executed it the real system, some of them are available as dummy command while many others produce an error message. The measurement allows to create an overview of the most frequently tried after penetration through the SSH protocol. Thanks to fact that the level of interaction in the honeypot is high usually only simpler attacks (e.g. bots and script-kiddies) are caught.

4 Distribution of Sensors

Three sensors emulating Windows and one emulating Linux were used.

- The first sensor is located in the server center of the University of Ostrava (Czech Republic) where it is connected to the special VLAN where filtering is not applied. This network is directly connected to the Czech Academic Network CESNET.
- The second sensor is implemented at the virtual private hosting – VPS server in Prague.
- The third sensor is located in Spojena skola in Kysucke Nove Mesto (secondary school in the northwest of Slovak Republic). This network is directly connected to the Slovak Academic Network SANET.

- The sensor focusing to attacks against Linux servers where Kippo honeypot is used is located at the VPS hosting in Prague.

It should be noted that CESNET and SANET academic networks are peered. All honeypots deployed in the study through the area of the Czech Republic and Slovakia as shown in Fig. 2.

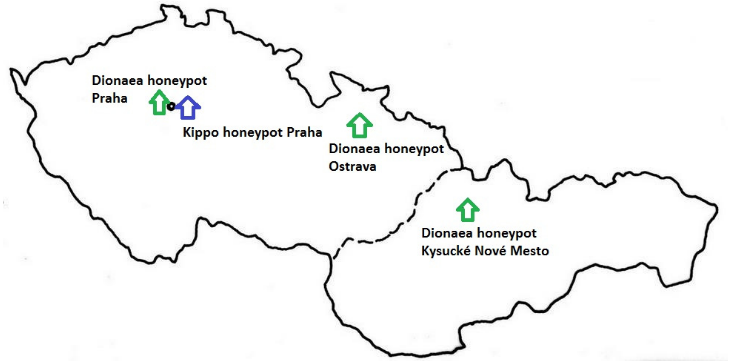

Fig. 2. Locations of Dionaea honeypots on the map of the Czech Republic and Slovakia

5 Results

The majority of results in this section is split into two categories in accordance with the OS that is emulated (i.e. Windows and Linux by Dionaea and Kippo honeypot, respectively). One of the results that can be gathered on both types of honeypots is the operating system of an attacker. Unfortunately this data (although gathered) turned out to be unreliable because of high level on “unknown” records due to outdated module for OS recognition in Dionaea. Therefore results from this type of analysis is not presented here in details. The total 99.956 % of attacks against Dionaea honeypots (Windows emulating) came from Windows and almost all remainder was represented by Linux (0.033 %) and SunOS (0.01 %).

5.1 Results for Kippo Honeypot – Linux Emulation

The main ’entry point’ to Linux-based systems is usually SSH protocol where attackers try to connect by a guessed password and username (that is often ’root’). One of quite interesting results is listed in Table 1 where the list of the most frequently tried combinations of the user name and password is shown.

The interesting and important result is the overview of activity of IP addresses that is shown in Fig. 3. Also the structure of SSH clients could be interesting as shown in Fig. 4.

Table 1. TOP10 combinations of usernames and passwords tried at Kippo honeypot

Username	Password	Tot. number of attacks
root	123456	215
root	admin	190
root	root	85
root	1qaz2wsx	78
root	cisco123	72
root	abc123	72
root	toor	60
root	1q2w3e	59
root	passw0rd	56
root	1	49

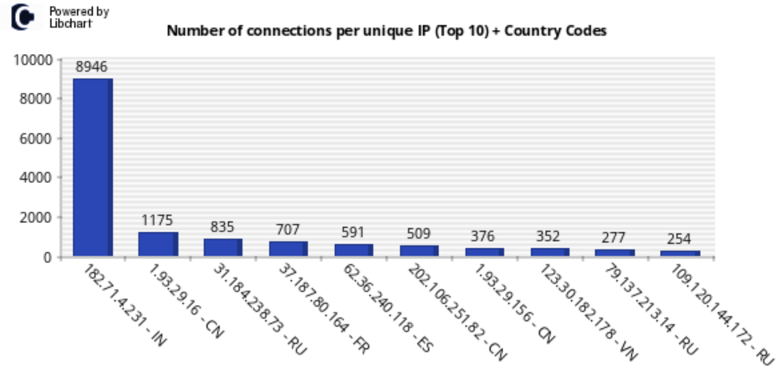

Fig. 3. Most active IP addresses on Kippo including assigned country codes

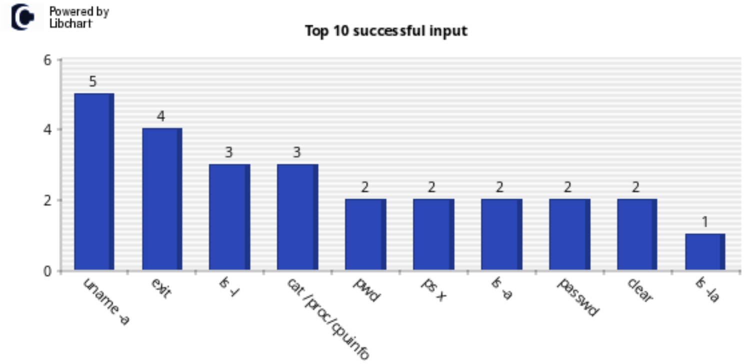

Fig. 4. The most frequently used commands on Kippo

5.2 Results for Dionaea Honeypots – Windows Emulation

One of the most interesting results from Dionaea honeypot is the overview of IP addresses activity throughout the period of measurement. The diagram in Fig. 5 shows the number of all attacks against the honeynet (3 Dionaea honeypots) from 10 the most active IP addresses as well as the number of connection when successful offering and subsequent downloading of a malware piece occurred. As one can see the number of connection towards the honeynet elements is quite high but only small part of attackers are able to offer a malware in a correct way so that it could be subsequently downloaded.The reasons were not investigated in details.

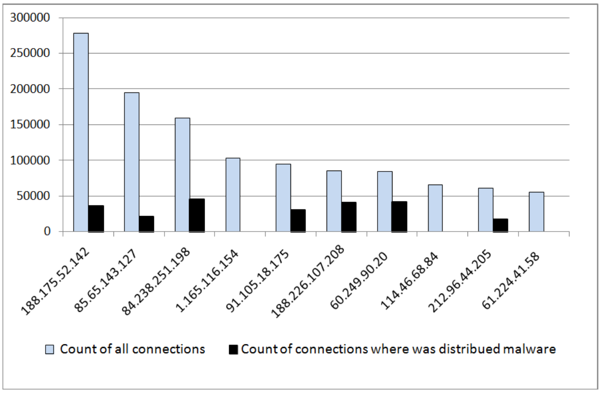

Fig. 5. Connections from the most active IP addresses to Dionaea honeynet and number of malware samples, which were distributed from these IP addresses for monitored time period

The main focus of honeypots is to identify attacks and threats coming from the Internet. Therefore one of the most important results of honeypots (especially those emulating Windows OS) is the identification of port numbers that are the most frequent object of attacks. The downloaded threats were identified according to the VirusTotal service with ESET-NOD32 as a primary antivirus database. The summary of results is shown in Table 2.

The apparent domination of the port number 445 is due to the fact that this port is registered for the SMBD service and it is known as the most frequently abused port by attackers against Windows OS. This is the port used primarily by the well known but still widespread worm conficker. The analysis of downloaded malware shows that conficker worm is still prevailing (almost 99.993 % of all malware downloaded). Because of this fact other threats (e.g. Win32/AutoRun.IRCBot.FC, Win32/Agent.UOT etc. whose occurrence was seldom – 81 and 40 cases, respectively) are not mentioned here in details.

The conficker dominance is despite the fact that this worm is quite old and patches avoiding its distribution is available for virtually every OS both for

Table 2. The most frequently attacked port numbers at Dionaea honeypots

Port number	Number of attacks
445	4279064
80	11806
1433	9220
3306	6818
135	135
21	64
5060	1

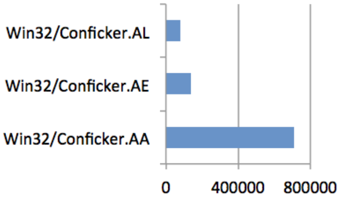

Fig. 6. Statistics of the most popular malware (conficker variants) according the number of connections that distributed malicious code

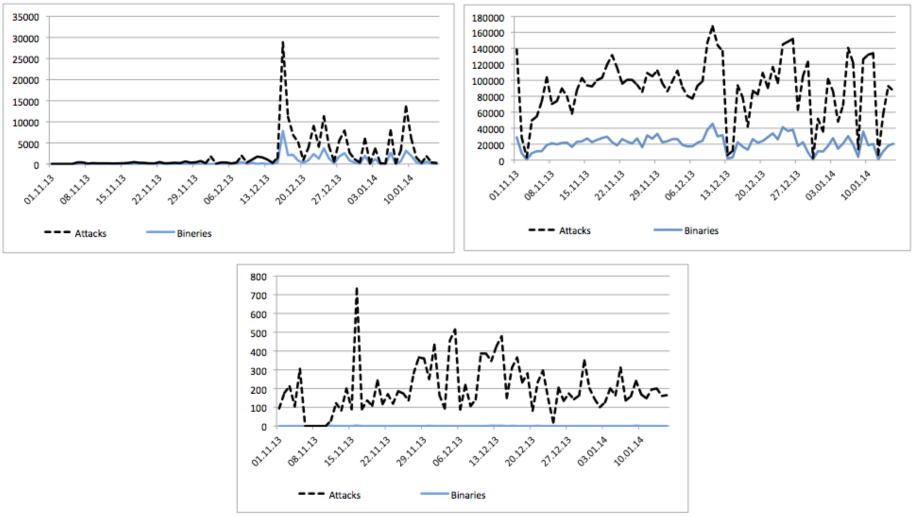

Fig. 7. Statistic outputs of attacks (upper dotted line) and downloaded binaries (lower solid line) for 90-day long period starting Nov 2, 2013, from all 3 Dionaea honeypots as described in Sect. 4 – 1st sensor in the upper left, 2nd in the upper right, 3rd one in the bottom

desktops and servers. On the other hand it should be noted that conficker variants are affluent and new ones are always emerging. The detailed analysis showed total 24 variants of conficker downloaded and another one suspicious to be so. Total 9.3387 million malware pieces were downloaded and only 665 were not identified as conficker (and only 162 pieces were identified positively as other malware. Regarding the occurrence of variants of conficker, the most frequently attacking variants are shown in the diagram in Fig. 6. The day-by-day history of number of attacks and downloaded binaries is shown in Fig. 7.

5.3 Attack Visualization

All attacks against the sensor described above being part of the low-interaction honeynet are visualized using HoneyMap [5] in real-time with places from where attacks were coming indicated. An example of our own experimental data visualized using HoneyMap is shown in Fig. 8. HoneyMap is based on the hpfeeds protocol and its social supplement HpFriends [11]. HoneyMap is a tool that is used here only to visualize our won results as it is used by other bodies (e.g. [12]).

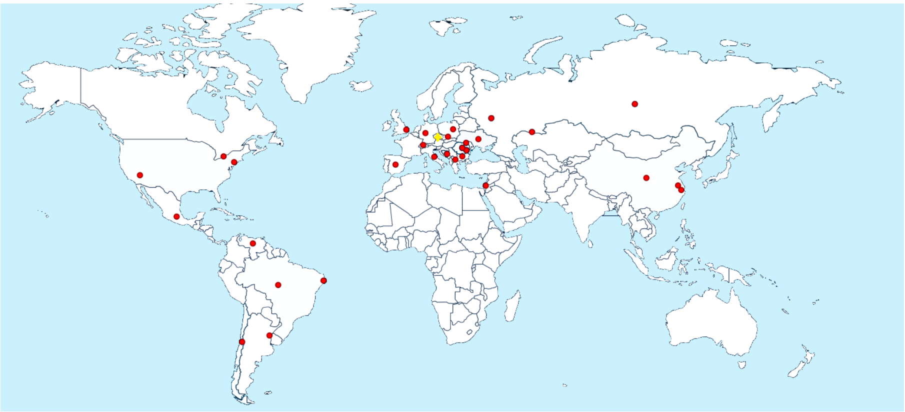

Fig. 8. Attack visualization from the data from our honeypots by HoneyMap in real time. The yellow dot indicates the sensor, the red dots indicate attackers.

6 Conclusions

Results obtained from our low-interaction honeynet resulted in a review of currently spreading threats and partially to activities of attackers as well. The most significant asset obtained is large amount of data in a scale and for that is not available elsewhere. The data show current spreading threats caught by honeypots. On the other hand the interpretation should be careful. For example the high amount of conficker worm attacks should not be interpreted as this is currently the most frequently attacking malware but it is still extremely frequent. This observation witnesses that the update approach of many computers is inappropriate. It indicates the growth of botnets as well. The most important conclusion from our measurements is that is was confirmed that honeypots allow to analyze the behavior of attackers and to reveal trends and changes. Using honeypots researchers are able to find more about techniques and procedures applied by attackers, to map current threats and malware and to apply such information into the improvement of safety tools and rules.

Acknowledgment. The authors thank for financial support for this paper publication to the Student Grant Programme 2014 (SGS) of the University of Ostrava. The authors thank to the secondary school Spojena Skola in Kysucke Nove Mesto and to the Centre for Information Technology of the University of Ostrava for providing a space and connectivity for our research honeypot sensor. Our thank belongs also to Mr. David Vorel, head of Czech Chapter of The HoneyNet Project, and to the staff of CZ.NIC Labs for expert consulting.

References

1. DenyHosts homepage, `http://denyhosts.sourceforge.net/index.html` (online, quoted January 21, 2014)
2. Internet Storm Center (ISC). Dshield.org, `http://www.dshield.org/` (online, quoted January 21, 2014)
3. Unspam Technologies, Inc. Project Honey Pot, `https://www.projecthoneypot.org/` (online, quoted January 21, 2014)
4. The Shadowserver Foundation. Shadowserver, `https://www.shadowserver.org` (online, quoted January 21, 2014)
5. Weingarten, F., Schloesser, M., Gilger, J.: HoneyMap, `https://github.com/fw42/honeymap/` (online, quoted January 21, 2014)
6. Joshi, R.C., Sardana, A.: Honeypots A New Paradigm to Information Security. Science Publishers (2011)
7. Kheirkhah, E., Amin, S.M.P., Sistani, H.A.J., Acharya, H.: An Experimental Study of SSH Attacks by using Honeypot Decoys. Indian Journal of Science and Technology 6(12), 5567–5578 (2013)
8. Sokol, P., Pisarcik, P.: Digital evidence in virtual honeynets based on operating system level virtualization. In: Proceedings of the Security and Protection of Information 2013, Brno. Univ. of Defence, May 22-24, pp. 22–24 (2013)
9. Kippo – SSH honeypot homepage, `http://code.google.com/p/kippo` (online, quoted January 21, 2014)
10. Carnivore.it. Dionaea project homepage, `http://dionaea.carnivore.it` (online, quoted January 21, 2014)
11. Schloesser, M., Gilger, J.: HpFriends homepage, `http://hpfriends.honeycloud.net` (online, quoted January 21, 2014)
12. Deutsche Telecom. SicherheitsTacho, `http://www.sicherheitstacho.eu/` (online, quoted January 21, 2014)