Deploying a University Honeypot: A case study

Rasmi Vlad Mahmoud and Jens Myrup Pedersen

Department of Electronic Systems, Aalborg University, Aalborg East, DK-9220, Denmark
rvm@es.aau.dk, jens@es.aau.dk

Abstract. The cyber threat against all parts of our societies is constantly growing, and while some attacks are carried out by cyber criminals for financial gains, others have strong strategic values and are likely to be carried out by nation state actors. One group of institutions that experience the growing cyber threat is universities: Universities are attractive targets because they often possess valuable research knowledge, and because universities traditionally have promoted an openness culture. At the same time, they face challenges in maintaining a high level of cyber security, since many people have system and physical access, and because there are many legacy systems in use. In order to build an efficient cyber defense it is crucial to understand the always changing threat picture, so the countermeasures can be adapted accordingly. However, doing so requires updated information about the attacks from a variety of sources. While many of these sources, e.g. threat assessment reports from intelligence agencies, come with regular intervals or in case of significant changes, this paper explores a way of getting real-time information about current attack attempts towards a specific university: Honeypots. The paper contributes by discussing advantages and disadvantages of different kinds of honeypots in a university setting, and it demonstrates how results can be achieved through actual honeypot implementations. Our conclusion is that honeypots are a valuable supplement to other sources of intelligence, but it is crucial to choose the right types and architectures.

Keywords: Honeypots · Cyber Security · Risk assessment · Universities

1 Introduction

The increasing cyber risks in general are documented through multiple sources. In the USA alone, official sources have estimated that malicious cyber activity cost the U.S. economy between $57 billion and $109 billion in 2016, a number that is only expected to grow. Universities are dealing with specific cyber security threats due to to their handling of research data, as well as data required for their normal operation, e.g. information about students and researchers. While universities carry out a large variety of activities, they do not have typical organisational boundaries and therefore need to establish custom security policies to accustom their needs [12].

A report published by Cisco in 2018 focuses on cyber security in the public sector, with a particular focus on education. More than half of the higher educational institutions (58%), reported that they had experienced at least one security breach, a percentage which is the highest over all public sector industries. This type of breach is most of the time identified with damage to the institution reputation. Nearly 51% of the attacks resulted in loss of money, in total over 500.000$ for the universities [2].

The challenge has also been identified in a Danish context. The Danish Defence Intelligence Service Center for Cyber Security (CFCS) stated in February 2017 that foreign states are conducting acts of espionage against Danish research. The curiosity is generated both from political and commercial motives, but nonetheless assailants were interested also in the institutions infrastructures that can be used for attacking other public Danish institutions [3]. It is an ongoing discussion within the research and educational sector how these challenges can be met without compromising on the openness culture of the universities.

As in other organisations, a number of initiatives are currently being taken to heighten the level and matureness of cyber security in universities. Danish universities are no exception, and the initiatives include awareness campaigns, updated password policies, installation of IDS/IPS systems and so on. Ideally these initiatives are taken based on an analysis of risks and consequences, but we claim that often such analysis is based on a combination of assumptions and outdated/partial information: Little is known about the actual and current threat picture faced by each organisation.

In this paper, we investigate how honeypots can be used to achieve current information about actual attempts of attacking universities. The work is based on a master thesis project [8], and contains two main contributions. First it is analysed which honeypots are more suited for a university setting, and second the results of an actual honeypot deployment is presented and discussed.

The rest of the paper is organised as follows: In Section 2 we provide a background on different kinds of honeypots and analyse which honeypots are most suited for a university setting. Next, selected honeypots are deployed as described in Section 3, and in Section 4 we present the results obtained. Section 5 presents the conclusion and discussions.

2 Background

A honeypot can be defined as a trap, where potential attackers are lured into a seemingly operational network, which is really established in order to attract hackers and study their behaviors. To serve this purpose, it is created to look as realistic as possible from the outside, while at the same time containing relevant tools that allow the operator of the honeypot to monitor and analyse the behavior of visitors. An example of a honeypot is depicted in Figure 1.

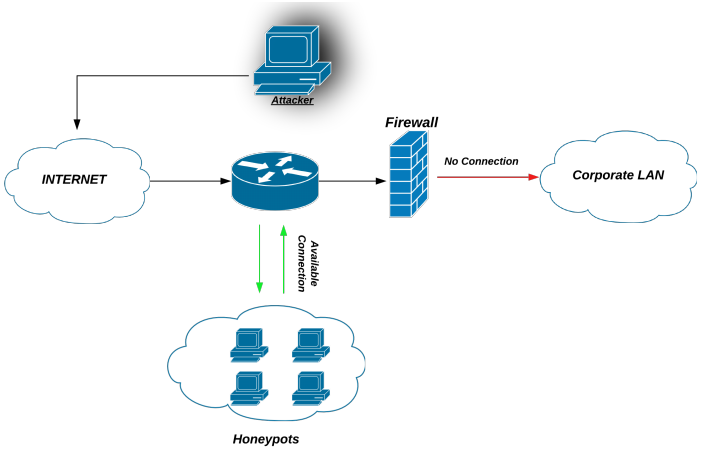


Fig. 1. Graphical display of how honeypots are working.

Based on the level of interactions determined by the availability of commands and the feedback that an attacker is experiencing when he is trapped inside, honeypots can be classified into three groups:

- Low-Interaction Honeypots (LIHP)
- Medium-Interaction Honeypots (MIHP)
- High-Interaction Honeypots (HIHP)

These will be described in the following.

LIHP - emulate a limited range of available services for the attackers to use. The main characteristic is that these types of honeypot are not having an operating system. Their main advantages are the facts that they are easy to deploy and maintain, yet they are excellent statistical tools. However, they are limited when it comes to detection of new attack patterns. [9]

MIHP - no operating system is present, but in contrast to LIHP they are capable of keeping the attacker engaged by answering to his commands. Nonetheless, the emulated services are more complex than LIHP. Both LIHP and MIHP present a low risk of being compromised, determined by the possibility of commands.[9]

HIHP - this type of honeypots is offering a real and unrestricted operating system to the attacker, and therefore is more complex in deployment and maintenance. Due to a large variety of information: monitoring services, attack logs, data access and file traversing that are stored in these types of honeypots, the processing of data needs to be done manually, and consequently it requires more time for deployment and maintenance. Moreover, these types of honeypots provide a high risk of compromise. [9]

In Table 1 the information in relation to LIHP, MIHP and HIHP is grouped into four levels *Low, Medium, High and Very High* based on the required initial knowledge, amount of information, maintenance time and risk of being compromised.

	LIHP	MIHP	HIHP
real operating system	no	no	yes
risk of compromise	low	mid	high
wish of compromise	no	no	yes
information gathering	low	mid	high
knowledge to deploy	low	mid	high
knowledge to develop	low	high	high
maintenance time	low	low	very high

Table 1. The table provided in [9] is summing up the principal characteristics of LIHP, MIHP and HIHP.

In addition, honeypots can be also classified based on their function as either *Production* or *Research* honeypots:

Research honeypots - are mainly used by educational, military or governmental institutions to gain information about the attackers' tactics, techniques and procedures(TTPs). This type of honeypots are not bringing direct value to the organization and require high maintenance time, on the other hand the type of information that they are providing is important for organizations to develop new policies to stay ahead of the cyber threats.[9]

*Production honeypots*s - are specific for companies and are mostly placed inside production networks, since they present a lower risk of being compromised. These honeypots are used to increase overall security, as well for decoy systems which works by deceiving the attackers and alerting the administrators about the activity.

2.1 Review of Honeypot Usage

In this section, a short review of existing honeypot studies is provided. There are several studies that were directed towards honeypots in the last years. Two projects worth highlighting are a project carried out by German Telekom [4] as well as the Leurre project [7]. Telekom used the information collected to protect their own systems, but also shares the data with security vendors. Unfortunately little has been published, and as such there are no further publications where the data is evaluated [6]. The Leurre project was stopped in 2008, but their data related to observed attacks was published.

Other large honeypot strategies include the NoAH project [1] and the Honeynet project [11], while more recently a number of projects concentrate on a

small number of sensors and a short period of time [6]. The honeynet project is a collaborative project, that is focusing the research on the black-hat community tools, tactics and procedures and then sharing the insight knowledge. The organization is composed from international security professionals who have deployed honeynets with the goal of further analysing the results.

The Finnish security company F-Secure is using honeypots to determine the landscape and the threat model of the years [5].

The data from these different projects provide useful statistics on the general threat picture, and can be a help in determining for example which protocols and countries to pay particular attention to. Also, much of the activity that is caught comes from automatic scanners, which tries to scan all possible Internet hosts without differentiating between companies, universities, and other organisations. However, they provide little insight into the particular threats towards universities. This challenge is studied further in the next section, where we analyse the requirements for establishing a university honeypot.

2.2 University Honeypots

Educational and research institutions are special types of organizations due to their variety of activities. Nonetheless, universities have claimed that due to budget limitations and lack of trained personnel they are facing huge impediments in relation to cyber security policies. In addition, the 2018 Cisco report over the public sector is stating that universities have employed only half of the medium number of security personal that other comparable organisations employ [2]. Therefore, a solution is needed which do not require a lot of time for deployment and maintenance, while the information provided by the system should still be valuable for the organizations.

Universities are sitting with large amounts of data that come in many forms such as intellectual property, personal data from employees and students, as well as research data from third parties, some of which can be business critical. For this reason security is an important aspect. Also, while honeypots can provide large amounts of data, in order to become valuable for universities this data need to be organized in a manner that facilitates automatic processing.

LIHP are good choices. Not only do they have the lowest necessary time of deployment and maintenance, they also present a low risk of being compromised due to the fact that they are not having an actual operating system. Nonetheless, the LIHP are able to provide an overview of the attacked protocols, the sources of the attacks as well as combinations of passwords and user-names used when services are probed. In addition, these type of honeypots are not requesting a long time to be spent daily for monitoring, and data generated by them can be structured for automatic processing. In addition to LIHP some MIHP might be relevant to consider as well, given an individual assessment of added value in terms of information gain versus the additional efforts in maintainance and risk management.

In the next section, the paper will go more in depth with the actual deployment of honeypots in a university environment.

3 Honeypots on AAU's Network

The following section will present the honeypots of choice together with the architecture and how they are integrated with the network of Aalborg University(AAU). There are many available honeypots, but some of them are already outdated and can present security risks, and so we narrowed down the choice to honeypots that are maintained and still in use. The chosen honeypots can emulate common services used at universities such as SSH, FTP, HTTP, HTTPS, SMTP and RDP.

Cowrie is a MIHP honeypot that is emulating *SSH* and *Telnet* protocols. It is recording the interactions that an attacker is having with the honeypot. Cowrie is chosen despite being a MIHP since it is the newest honeypot to emulate SSH and Telnet, and since it is easy to setup and maintain. Furthermore, it is considered secure to run since it is not running a real operating system, and since the amount of offered commands is limited.

Dionaea is a LIHP honeypot capable of offering a number of protocols including: Server Message Block (SMB), HTTP, FTP, Microsoft SQL Server (MSSQL), and Voice over Internet Protocol (VoIP).

Heralding is a LIHP designed to store the used credentials. The emulated protocols are Hypertext Transfer Protocol (HTTP), Hypertext Transfer Protocol Secure (HTTPS), Post Office Protocol 3 (POP3), Post Office Protocol 3 Secure (POP3S) and Internet Message Access Protocol (IMAP).

Mailoney is a LIHP that is mimicking a classic mail server by exposing the Simple Mail Transfer Protocol(SMTP).

RDPY is a LIHP that imitates the Windows protocol, Remote Desktop Protocol(RDP).

3.1 Honeypots Deployment

Honeypots were deployed using Docker which is an open-source platform for running, developing and distributing applications. Docker offers the possibility of grouping all the necessary dependencies inside one package, named *container*, which is offering isolation, abstraction and security [10]. Therefore, the honeypots were deployed into individual containers to ensure isolation and also to avoid any errors to escalate. In Figure 2 the honeypots relation with Docker is demonstrated, and the approach is presented horizontally.

For bringing more value to the architecture, the honeypots were integrated with AAU's existing network, and the following subsection will present a general overview of how the integration was made.

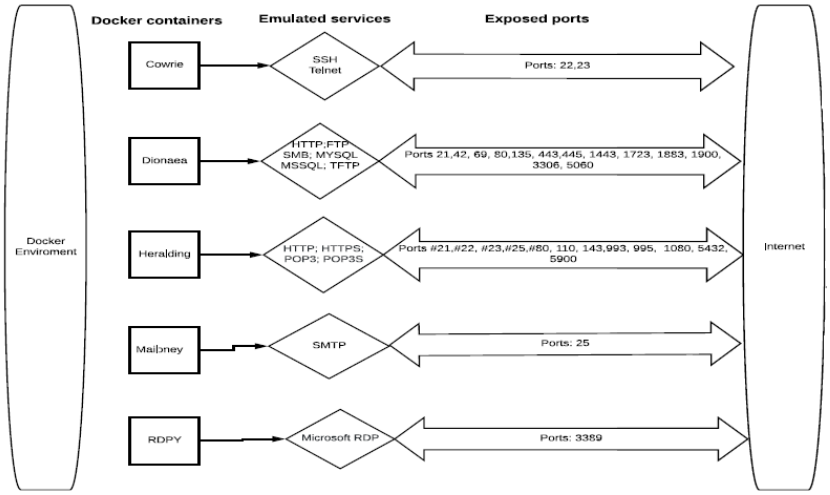


Fig. 2. Overview of the honeypots integration with Docker.

3.2 Honeypots Integration with AAU's Network

Docker was hosted on a Virtual Private Server (VPS) provided by AAU. Figure 3 gives an overview of the architecture. The VPS was configured on a small subnet ($X.X.X.113/29$) that was sitting outside AAU's main firewall, and a firewall (see the figure) was setup by us to control the traffic to and from the VPS. For this reason, the VPS had a principal network interface used exclusively for administration purposes, the connection was established over SSH, and a number of secondary network interfaces were created in software for the honeypots.

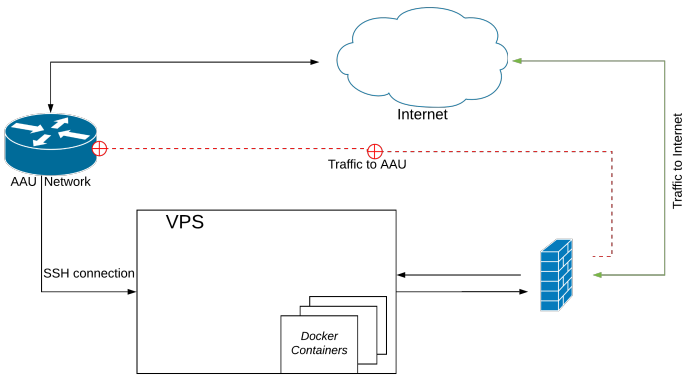


Fig. 3. Honeypots integration with AAU's Network.

Considering the design from a security perspective, traffic originated from the VPS was not allowed to go back to AAU's network. As a consequence all traffic to the Internet was freely allowed, since it was also important to keep the appearances and to not raise any suspicions for the attackers. In addition, four secondary network interfaces were created and the IPs were distributed across the docker containers. The primary containers behaviour was modified by assigning a public static IP address to every container in order to integrate them with the existing AAU's network.

All the interactions with the honeypots are stored individually in log files and therefore a method to structure and analyze the files is presented in the following subsection.

3.3 Logs Processing

Given this container based architecture, where the log files are saved individually, a method to centralize the logs was adopted. Graylog has been used as the log management tool, and together with its dependencies it was deployed inside Docker containers on the same VPS. Once the log files are sent to Graylog, their management is handled by storing the information in Elasicsearch and the neccessary Graylog settings in MongoDB. In Figure 4 the proposed architecture for this part is presented in order to offer a better understanding of the components and their relation:

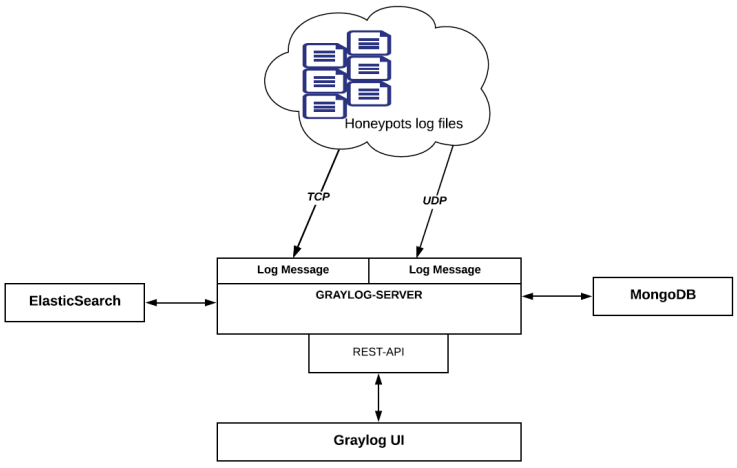


Fig. 4. Overview of log files parsing architecture.

The following section will present and analyze the results recorded by the honeypots by focusing on the probed protocols.

4 Results Of Honeypots

The honeypots were initially started on May 13 2019, and their activity was monitored for a period of 30 days. They were configured to monitor different kinds of intrusions including scanning and login attempts: In the following, we refer to these altogether as "connection attempts". During the recording period a total number of 780.305 connection attempts were registered. The *SSH* protocol was the one recording most of the attacks, but the following protocols were also targeted: *RDP*, *HTTP*, *TELNET* and *HTTPS*. The majority of the sessions created came from Ireland (72%), followed by Netherlands (20%), China (2%), Jordan (1%), and Germany (1%).

In Figure 5 the top 10 countries based on the number of attempts are displayed. The countries presented are considered to be the last hop of the attacks, since is not possible to identify from the logs if the attackers are using proxies.

Country Code	Percentage from Total Attempts	Total No. of Attempts
IE	71.90%	559,207
NL	19.95%	155,116
CN	2.17%	16,912
JO	1.20%	9,331
DE	0.88%	6,865
US	0.64%	4,951
RU	0.41%	3,182
TH	0.37%	2,857
PA	0.34%	2,650
TW	0.21%	1,643

Fig. 5. Top 10 most active countries based on the number of attempts.
IE = Ireland, NL = Netherlands, CN = China, JO = Jordan, DE = Germany, US = USA, RU = Russia, TH = Thailand, PA = Panama, TW = Taiwan

In the following subsections the honeypots that were deployed will be presented one by one and their results will be described and analyzed.

4.1 Cowrie

Considering that the majority of the connections were addressed to the *SSH* protocol generated by Cowrie it is valuable to get an overview of the activity. There were a total number of 725.993 connection attempts oriented to the *SSH* protocol. From the total number of connections nearly 240.000 were direct tcp/ip

requests from the honeypot to a company based in Russia that is offering different Internet services, *Ya.ru*. Therefore, there is a clear tendency of using the attacked machine as a proxy and launch attacks from that to other devices. However, it is beyond the scope of this study to analyse whether the attacker actually knows that he is inside a honeypot, which he could then deliberately try to use as a proxy for his attacks.

4.2 Dionaea

Dionaea is the honeypot that is emulating the highest number of protocols. Figure 6 provides an overview of the most attacked protocols, and the relation with the source country is presented.

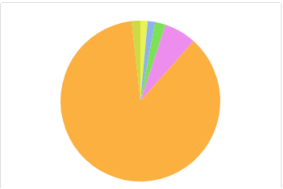
Value	%	Count
Top 5 values		
mssqld — CN	58.82%	7,779
mysqld — IE	4.11%	544
mssqld — IN	3.86%	510
httpd — NL	2.35%	311
mssqld — PH	2.34%	309

Fig. 6. Top 5 most attacked protocols and source countries.

Additionally, this honeypot was also registering usage of generic usernames such as *sa*, *root*, *admin*, *mssqla* or *server*, together with passwords like *12345678, password, 1qaz2wsx, abc123, and qwerty*. Nonetheless, binary and bitstream files were collected, but no thorough analysis was performed during the development of this project.

4.3 Heralding

This honeypot demonstrated a trend in the attacks towards less secure protocols such as *HTTP*. Moreover, Figure 7 shows the representation of top 5 protocols together with the total number of attempts.



Value	%	Count
Top 5 values		
http	86.62%	21,131
https	6.49%	1,584
postgresql	2.12%	518
pop3	1.52%	370
imap	1.46%	356

Fig. 7. Top 5 most attacked protocols in the Heralding honeypot.

From the total number of attempts, a percentage of 32.30% were using combinations of usernames and passwords such as *admin-admin,root-root, postgresql-postgresql and Cisco-Cisco*. As such, attackers were inclined to use sequences where the service name was used both as a username and password. Moreover apart from the service names, usernames also include *admin, super, superadmin, user or support*. Also additional passwords were used between the groups such as *root, cisco, admin, or support*.

4.4 Rdpy and Mailoney

For Rdpy and Mailoney the number of connections were lower than for the other honeypots, but nonetheless interesting: Rdpy was registering attacks from Jordan, USA, Russia, Ivory Coast and China. The main number of the attacks were originating from an USA IP address that was already known as an infected device primarily used for $DDOS$ attacks via IoT devices.

Furthermore, for Mailoney the attacks were linked back to USA, China, and Ireland. The origins of the connections were the IP addresses that are known to host malicious botnet activities. The activity that was most recorded was to use this honeypot as an open relay by trying to connect to other mail-servers for sending spam emails. Domains such as *sh-chi-us-gp1-wk108.internet-census.org* and *zx2.quadmetrics.com* were contacted, all of which are already blacklisted domains due to spam activity.

5 Conclusion

In the last years, cyber security has gotten on top of the agenda in many public and private organisations: The risks are increasing both from cyber criminals driven by profit and nation states driven by more strategic interests. In this situation, universities are facing themselves with a high risk due to their valuable research data as well as personal data for their operations, while at the same time trying to maintain an open culture. Having a good and actual view of the current attack picture is crucial in order to take the right countermeasures. This paper investigated how honeypots can contribute to achieving such a better picture: Through an analysis of different honeypot techniques it was found that low interaction honeypots can provide valuable information while at the same time keeping the risks low and minimizing the exposure of critical information. Among the results were that during a 30 day period more than 725.000 connection attempts were oriented towards SSH with the majority coming from Ireland, Netherlands and China. Looking at other protocols, many connections were coming from also China, India and Philippines. It was also revealed which protocols were most often targeted, along with the most commonly guesses of usernames/passwords. The results demonstrate that honeypots can provide valuable and timely information to universities about the current threat picture. There is however a trade-off between ease of use, configuration

and analysis on one hand, and the amount of information that can be achieved on the other. However, it does not help anyone that information is collected if it is not actually used. In order to make the information operational, future research could focus on how to present and integrate the results for the risk management organisation, but also on how a collaborative approach could be taken among universities to identify trends and ongoing attacks as early as possible.

References

1. Noah - a european network of affined honeypots (2018), https://cordis.europa.eu/docs/publications/1201/120142541-6_en.pdf, [Online; accessed 18-August-2019]
2. Cisco: 2018 annual cybersecurity report impacts on public-sector (2018), https://www.cisco.com/c/dam/m/digital/elq-cmcglobal/OCA/Assets/Federal/2018-Annual-Cybersecurity-Report-Impacts-on-Public-Sector.pdf?ccid=cc000126&oid=rptsc008809&elqTrackId=64397bd4bdfd4bf6a6cde2dee70f3e6e&elqaid=4518&elqat=2, [Online; accessed 08-March-2019]
3. for Cyber Security (CFCS), D.D.I.S.C.: Foreign hackers threaten danish public research (2017), https://fe-ddis.dk/cfcs/publikationer/Documents/TV%20forskning%20ENG.pdf, [Online; accessed 08-March-2019]
4. DTAG, T.: Fruhwarnsystem, sicherheitstacho (2013), http://www.sicherheitstacho.eu/, [Online; accessed 11-March-2019]
5. F-Secure: Attack landscape h1 2018 (2018), http://images.secure.f-secure.com/Web/FSecure/%7Ba1352f14-be26-4fd1-bcc8-3c9bd6b20bd3%7D_Attack_Landscape-H1-2018.pdf, [Online; accessed 11-March-2019]
6. Fraunholz, D., Zimmermann, M., Hafner, A., Schotten, H.D.: Data mining in long-term honeypot data. In: 2017 IEEE International Conference on Data Mining Workshops (ICDMW). pp. 649–656. IEEE (2017)
7. Leita, C., Pham, V., Thonnard, O., Ramirez-Silva, E., Pouget, F., Kirda, E., Dacier, M.: The leurre. com project: collecting internet threats information using a worldwide distributed honeynet. In: 2008 WOMBAT Workshop on Information Security Threats Data Collection and Sharing. pp. 40–57. IEEE (2008)
8. Mahmoud, R.V.: Honeypots on aau's network (2019), https://projekter.aau.dk/projekter/files/306402738/NDS10_Gr1024_Report.pdf, [Online; accessed 08-March-2019]
9. Nawrocki, M., Wählisch, M., Schmidt, T.C., Keil, C., Schönfelder, J.: A survey on honeypot software and data analysis. arXiv preprint arXiv:1608.06249 (2016)
10. Rad, B.B., Bhatti, H.J., Ahmadi, M.: An introduction to docker and analysis of its performance. International Journal of Computer Science and Network Security (IJCSNS) **17**(3), 228 (2017)
11. Spitzner, L.: The honeynet project: trapping the hackers. IEEE Security Privacy **1**(2), 15–23 (March 2003). https://doi.org/10.1109/MSECP.2003.1193207
12. Universities, U.: Cyber security and universities; managing the risk. Retrieved December **31** (2013), https://www.universitiesuk.ac.uk/policy-and-analysis/reports/Documents/2013/cyber-security-and-universities.pdf, [Online; accessed 03-March-2019]