Continent-based Comparative Study of Internet Attacks

Idris A. Rai and Matsiko Perez

Makerere University, Kampala, Uganda.
rai@cit.mak.ac.ug,mushura.perez@gmail.com

Abstract. We have deployed a honeypot sensor node in Uganda that is connected to a distributed honeypot system managed by Leurrecom.org Honeypot project, which constitutes of a large number of different *honeypot* sensors distributed across different continents. Once joined the project, the system allows access to the whole dataset collected by all sensors in the distributed system. We use the data collected by the honeypot sensors for a period of six months to compare the attacks that have been detected by honeypot sensors in Africa to the attacks detected by sensors in other continents. Our findings reveals that sensor nodes in Africa experience a significant number of attacks. In some cases, the number of attacks for African sensor nodes is significantly higher than many sensors in developed countries. This shows that network attacks are independent of location and Internet popularity in a country. That is, low Internet penetration level in African countries does not mean that networks in Africa are safe from external attacks. In fact, the results further indicate that some attacks are highly likely guided against specific networks.

Key words: Internet attacks, threats, honeypot sensor, distributed honeypot systems, SGNET

1 Introduction

In order to effectively protect the Internet, there is a need to have an indepth knowledge of Internet threats and attacks. To achieve this, it is very necessary to collect sound measurements about the existing and emerging Internet threats and their processes as observed on the Internet world over. Several initiatives have been in existence to monitor malicious activities or to capture malware information [1, 2, 10, 12, 13, 14]. In this paper we use one of the most recent similar initiative called Leurrecom.org project and its data collection infrastructure using SGNET deployment to study Internet attacks [4]. Leurrecom project is based on worldwide distributed system of *honeypot sensors* that are deployed in more than 30 countries covering five continents. The major objective of the project is to have a clear knowledge of the nature and behaviors of threats/attacks happening on the Internet by collecting data on the attacks on a long term perspective.

A *honeypot sensor* can be defined as a security resource whose value lies in being probed, attacked, or compromised [3]. The concept of honeypot was

introduced by L. Spitzner [3] in the late 1990's with the main goal of studying attacks/threats and their trends on a global scale across the whole Internet. In this paper, we will refer to a honeypot sensor as a *sensor*.

SGNET honeypot technology that is used in Leurrecom project differs from other honeypot systems in that it coordinates honeypot sensors and seamlessly integrates them into a distributed architecture through an overlay based on ad-hoc HTTP-like protocol called Peiros. The result of this integration is a distributed honeypot deployment that is able to automatically learn and handle server-based exploits, and emulate the code injection attacks up to the point of the malware download [15]. Leurrecom project uses ScriptGen technology to collect data from all sensors. In this paper, we use SGNET to investigate Internet attacks at continental level. We are specifically interested in comparing the nature of the attacks experienced by honeypot sensors in Africa to sensors in other continents.

Inspite of the existence of a few sub-marine cable initiatives to connect networks in developing countries to the Internet, most networks in Africa, mainly Sub-Saharan Africa are still connected to the Internet using low-speed satellite links. As a result, access to Internet is still not affordable to many, leading to stagnant or low Internet penetration in African countries. In turn, Internet traffic dynamics in developing countries are very simple and fairly predictable compared to traffic patterns in developed countries. As such, attackers might, perhaps rightfully, assume that the lack of wide spread Internet access in African countries is synonymous to lack of Internet security awareness and security expertise to secure and troubleshoot the networks. Others Africa might well think that attackers wouldn't be interested in simple networks in Africa. While the former is a very good motivation for attackers to test their newly developed attacks, we show that the later belief is practically very wrong and misleading.

We use the data gathered by all active sensors on Leurrecom project for a period of six months, mainly from Dec 2009 to May 2010. The analysis of the data reveals that the honeypot sensors in Africa experience a significant amount of attacks, in some cases surpassing the attacks reported by honeypot sensors located in developed countries. We therefore show that networks in Africa are as vulnerable, exposed, and at risk as networks in other continents. Our analysis also shows that some attacks are directed to specific continents or networks across the world.

In the remainder of the paper, we present an overview of honeypot systems in Section 2. In Section 3, we present deployment requirements for SGNET. In Section 4 we analyze the collected data and discuss on our findings. In Section 5 we discuss a summarised analysis of results from nodes in Africa and finally conclude the paper in Section 6.

2 A review of honeypot systems

We have presented a definition of honeypot sensor from [3] as a security resource whose value lies in being probed, attacked, or compromised. There are

however varying other definitions of honeypot, leading to some miscommunication and confusion amongst researchers. Some researchers refer to honeypot as an intrusion detection tool, whereas others think it is a deception tool. There are those who think it is a weapon to lure hackers, and still others believe a honeypot should emulate vulnerabilities, and some view honeypots as controlled production systems that attackers can break into.

One of the well-known research initiative on honeypot technologies was called *honeynet* project [5]. A *honeynet* is a type of honeypot designed primarily to gather information on the enemy for research purposes. The honeynet project began in 1999 for the purpose of gathering intelligence on attacker techniques, tools and motives that might help the security community identify new threats and weaknesses more effectively.

Honeypots are classified according to their level of interaction with the attacker. There are *low-interaction, mid-interaction, high-interaction* honeypots [6]. Low-interaction honeypots expose to attackers certain *fake* (emulated) services which are implemented by listening on specific port (services are limited to specific listening ports). With low-interaction honeypot, there is no real operating system target that an attacker can operate on. An example of low-interaction honeypot is *honeyd*, which is an open source designed to run primarily on Unix systems [7].

Mid-interaction honeypots provide more sophisticated *fake daemons* with deeper knowledge about the specific services they provide. With this, the attacker is able to detect a real operating system and has more possibilities to interact and probe the system. The daemons involved need to be as secure as possible. Examples of mid-interaction honeypots are *honeyd and Specter*. Finally, high-interaction honeypots involve a real operating system that is offered to the attacker, thus providing the attacker with capabilities to upload and install new services or applications. All actions are monitored and recorded in order to gather more information about the blackhat community. This means that the system must be under monitoring all the time.

Distributed honeypot systems are built using a number of connected honeypot sensors that are configured across the Internet. They therefore provide platforms to compare attacks experienced at different locations, and to study propagation of existing and newly emerging attacks.

There are a number of projects based on distributed honeypot platforms. For example, a research project called Collection and Analysis of Data from Honeypots (CADHo project) [10], DShield Project [12], MyNetWatchman [13], and Internet Telescope project CAIDA [14].

As earlier mentioned, the most recent distributed honeypot platform is SGNET [4]. SGNET is a low-interaction honeypot system that exploits the strengths of ScriptGen technology and dynamically combines with other existing solutions namely Argos and Nepenthes. ScriptGen is an automated script generation tool for honeyd, Nepenthes is a honeypot with specific objective to download malware from attacking sources, and Argos is an emulator for Fingerprinting Zero-Day Attacks. SGNET is capable of offering an overlay based on an

ad-hoc HTTP-like protocol called Peiros to coordinate its entities and integrate them into a distributed architecture. The ultimate result of this integration is a distributed honeypot deployment that automatically learns and handles server-based exploits, and emulates the code injection attacks up to the point of the malware download.

SGNET has been used by researchers to study various behaviors and attributes of Internet attacks and their underlying attack tools [6, 8, 9]. In this paper, we use SGNET to particularly compare Internet attacks between honeypot sensors in Africa with sensors located in other continents.

3 Deployment of honeypot sensor using SGNET

To be able to collect attack data and have access to data from other honeypot sensors across the world, we deployed honeypot sensor in our local network in Uganda and connected it to the SGNET platform. The process of deploying a honeypot sensor is fully automated. Interested parties in participating in the Leurrecom.org project provide a dedicated computing and networking environment with minimal stipulated requirements whereas the institute that oversees the honeypot project provides an installation CD, access to the collected data and analysis tools as well as integrity of data collected.

SGNET based deployment uses a collection of several tools and functional modules to build what we call a distributed honeypot system. Such tools are Argos, Nepenthese, VirusTotal, Anubis, Maxmind, and P0fv2. The data collected from all sensors in the network is automatically uploaded into a central database on daily basis. Different datasets that are collected from all the participating partners is accessible through a Web interface to all participating partners for easy analysis of the data.

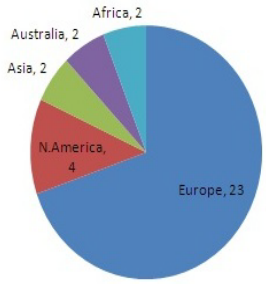

Fig. 1. Number of Active Honeypot Sensors per Continent

4 Comparative analysis of attacks

In this section we discuss the findings we derived from analyzing the collected data. Specifically, we compare and analyze the attacks experienced by honeypot

sensors on per continent basis. We use the data collected during six months period starting from Dec 2009 to May 2010 to compare how vulnerable are honeypot sensors (also networks) in Africa compared to honeypot sensors located in other continents. During the six months, a total of 33 honeypot sensors were active in five continents, namely Europe, North America, Asia, Australia, and Africa.

Figure 1 shows the number of active honeypot sensors by continent. We can see that Europe had by far the largest number of sensors; 23 active sensors which make about 70% of the total active sensors. This shows the involvement of European research community in network security issues. Perhaps it also shows how cautious the Europeans are on network security. The figure shows that North America had four active honeypot sensors while the rest of the continents had two active sensors each. The active nodes in Africa were located in Uganda and Egypt. Interestingly, during the whole six months of our project running, we have not observed any active sensor in South America. It was however also observed that there were honeypot sensors in the network that were part of the distributed honeypot system but were not active during the time when measurements were collected.

In the following sections, we present an indepth comparative analysis of specific attacks sources and actual attacks that were experienced by sensor nodes in each continent.

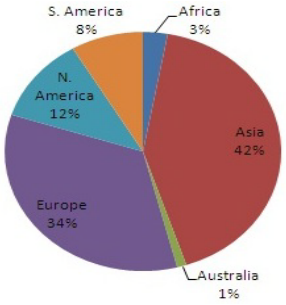

Fig. 2. Attack sources per continent

4.1 Analysis of attack sources

In this section we compare the sources of attacks originating from each continent. An attack source is identified by an IP address from which the attack is originated. We first look at the distribution of attack sources by continent and then investigate the countries that generate significant sources of attacks towards each continent.

In total, an overwhelming 1,042,282 attack sources have been recorded during a period of six months. Figure 2 shows the percentage of attack sources per continent. We can see from the figure that Asia generated 42% of all attacks which

amounts to 438,209 attack sources. Europe is also reported to have contributed to a large fraction of attack sources of 34%. This is a total of 354,056 attack sources originating from Europe. Attack sources from North America made 12% of total number of attack sources, which is equivalent to 120,455 attack sources. South America generated 88,236 attack sources, Africa 29,946 attack sources, and Australia 11,380 attack sources which contributed to 8%, 3% and 1% of the total attack sources respectively. Africa and Australia have the least number of attack sources as compared to the rest of the continents.

It is perhaps a fact that, the more Internet users in a continent, the more malicious users are likely to be as well. That is why we have fewer attacks from Africa and Australia compared to attacks from Europe, Asia, North America and South America. It is however interesting to observe that Africa generates a sizeable number of attacks which is more than in Australia. Detailed analysis of the data is not shown in Figure 2. Most of the attacks from Africa are observed to originate from Gabon and South Africa.

Origin	**EU1**	**EU2**	**AS1**	**AS2**	**NA1**	**NA2**	**AU1**	**AU2**	**AF1**	**AF2**
Russia	10	10	10	1	7	8	5	7	6	6
USA	9	9	9	3	10	9	9	9	8	9
China	8	8	8	9	9	10	10	10	9	10
Taiwan	7	5	6	7	4	5	4	5	-	7
Italy	6	6	5	-	-	4	-	-	-	-
Denmark	4	4	3	-	-	1	-	-	-	-
Romania	5	2	4	-	-	-	-	-	-	-
Brazil	-	7	7	4	5	6 -	6	6	-	4
Peru	-	3	-	6	-	-	-	-	-	-
Japan	3	-	2	8	3	-	3	4	-	-
Poland	2	-	-	-	-	-	-	-	-	-
T. Tobago	-	1	-	5	-	-	-	-	4	8
Argentina	1	-	-	-	-	-	-	-	-	-
Colombia	-	-	-	2	-	-	-	-	-	-
Gabon	-	-	-	-	1	-	1	1	3	-
Pakistan	-	-	-	-	2	-	-	-	-	-
France	-	-	-	-	-	-	-	2	5	2
Estonia	-	-	-	-	-	3	-	-	7	-
Portugal	-	-	-	-	-	-	-	-	2	1
India	-	-	-	10	-	-	-	-	-	-
Canada	-	-	-	-	6	2	2	3	-	3
Australia	-	-	-	-	-	-	7	-	-	-
Latvia	-	-	-	-	-	-	-	-	10	-

Table 1. Score of attack sources by country of origin

We further selected two most attacked sensors from each of the five continents and analyzed the recorded attack sources by each sensor. We denote EU1 and EU2 the two sensors located in Europe, AS1 and AS2 are located in Asia,

NA1 and NA2 are located in North America, AF1 and AF2 are located in Africa and AU1 and AU2 are located in Australia. For each honeypot sensor, we identified top 10 countries with the most attack sources that were recorded by each identified sensor. We then gave scores to each country such that the country with most attack sources is given the highest score of 10 and the one with least sources is scored 1. This enables us to compare the sources of attacks in terms of specific countries that attacked different continents. We present the results in Table 1.

We can observe a few patterns from the table. Firstly, the three countries with most attack sources (Russia, China, USA) seem to attack all continents almost equally, i.e., their scores in each continent don't vary too much. This shows that either there are many attack sources in these countries, and/or the sources indiscriminately broadcast their attacks on the Internet, i.e, without specific target in mind. In contrast, the table also shows that sources from some countries tend to target networks in specific countries. For instance, India has a score of 10 and appears only in Asia which means there are many attackers in India that target networks only in Asia. Similar pattern is observed for the case of Pakistan and Canada. We also observe that most attacks recorded by one honeypot node in Africa were from Latvia (score 10) and Estonia (score 7), which are not the most attacking countries to other continents. Trinidad and Tobago also oddly appears to strongly attack the other sensor in Africa. Also interesting to note is Portugal appearing only on the top 10 list of African honeypots. Other similarly odd observations are seen for Argentina, Trinidad and Tobago, Colombia, and France. These attacks may be due to compromised machines in those countries.

The results in Table 1 reveal that while majority of the attacks originate from specific countries that indiscriminately attack networks, some attacks seem to be targeting specific networks. In particular, we observe that African nodes are vulnerable to attacks that originate from isolated countries. This is a clear evidence of directed attacks to specific networks.

In the following section, we analyze in details the specific attacks registered by most honeypot sensors. Some of these attacks include code injection attacks, malware download, backscatter attacks and targeted ports.

Code injection attack sources Code injection is the exploitation of a computing system bug that is caused by processing invalid data. It can be used by an attacker to introduce (or "inject") code into a computer program to change the course of execution. The results of a code injection attack can be disastrous. In a period of six months, a total of 26,710 code injection attack sources was observed by all honeypot sensors. Europe recorded the largest percentage of code injection attacks with 23,171 code injection attack sources as seen from Figure 3. Asia follows in the second position with 1,868 attack sources.

It is perhaps surprising that honeypot sensors in Africa reported more code injection attack sources compared to sensors in continents such as North America where we expect more attacks because of more Internet activities. These results further asserts our previous observation that having less Internet activities does

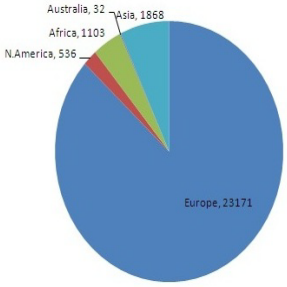

Fig. 3. Code Injection Attack Sources by Continent

not guarantee security against Internet threats, and it is highly likely that some attackers tend to target locations where Internet penetration is low assuming that security awareness level in those countries is low.

Backscatter attack sources Backscatter attacks occur when a flood of messages is received with a forged sender address as spam messages. A total of 25 Backscatter attacks sources were observed targeting those sensors. These attacks were recorded on only three honeypot sensors; two located in Europe and one located in Australia.

Analysis of targeted ports/port sequences Attacks tend to attack different ports based on a certain sequence while others target individual ports directly. The sequence of ports attacked is also an inherent feature for some attack tools. Analysis of attack ports sequences is used to understand the nature of the attacker. Table 2 shows the distribution of sources received by each port or port sequence per continent.

Port Sequence	Europe	Asia	Africa	N. America	Australia
445	380,233	70,450	54,783	1,460	347
135	836,099	3,321	3,367	1,141	203
23	35,048	3,280	544	329	436
44-139	120,039	21,458	2,740	904	112
445-80	95,920	20,411	2,028	31	2
445-139-80	22,815	16,873	1,707	31	14
80-445	10,842	2,423	147	0	0
80-445-139	12,625	2,349	141	0	0

Table 2. Attacked Ports

Attack process against these ports seems to be fairly regular. We also observe that most sources have sent their requests to port 445 (Microsoft-ds). Comparing to other continents, Europe recorded a much larger number of attacks at port 135 (Microsoft RPC), which is similar to port sequence 44-139. We again see from the

table that honeypot sensors in Africa experienced more cases of targeted ports than other continents notably North America and Australia. In some cases, for instance, port 445 attacks, the difference is significantly large.

5 Analysis of attacks for African nodes

In summary, we have observed that Africa honeypot sensors have been subjected to similar attacks with other honeypots located in different continents. Some of these attacks include code injection attacks, backscatter attacks, malware downloaded and specific ports being targeted by the attacks. In this section, we look at attacks for African honeypot nodes.

African sensors recorded a total of 29,946 attack sources and of these attack sources, 1,103 were code injection attacks, with Egypt sensor registering most attacks equal to 596 attack sources and Ugandan sensor experienced 507 code injection attack sources. From Table 1, we see that the two sensors in Africa are attacked by sources China, USA, and Russia at almost equal intensity. However, the sensor in Egypt was uniquely attacked by sources from Taiwan, Brazil, and Canada whereas the sensor in Uganda was uniquely attacked by sources from Latvia and Estonia.

We also observed that the honeypot sensor in Egypt experienced more attacks on targeted ports than the sensor in Uganda. However, the pattern isn't uniform when one looks at individual ports. For instance, the sensor in Uganda experienced more attacks on ports 135 and 23 recording 2,411, and 501 compared to 926 and 43 for Egypt respectively. It is difficult to know why there was more interest to attack some specific ports on the network in Uganda than on the network in Egypt.

6 Conclusion

In this paper, we deployed a honeypot sensor in a local network in Uganda and connected it to a distributed honeypot sensor system called SGNET. We collected data from the local sensor and all other active sensors in the distributed honeypot system to study and compare the attacks reported.

We observed from our analysis of the data that networks in Africa experience a significant amount of attacks in some cases surpassing the attacks experienced by networks in developed continents such as North America and Australia. We also discovered that some attacks sources constantly target specific networks such as networks in Africa and Asia. In summary, the Internet doesn't have any boarders to block against attacks. This demands for a special care to be taken when deploying networks anywhere in the world. There is a need to setup honeypot sensors in different locations in the world in order to collect data that will provide a complete picture and wider comparison of Internet attacks and their behaviors.

Acknowledgement. This work was partially supported by the Cisco University Research Fund, a corporate adviced fund of Silicom Valley Community Foundation.

References

1. Team Cymru, The darknet project. http://www.cymru.com/darknet/, accessed 04/03/2010.
2. Internet Motion Sensor. http://ims.eecs.umich.edu/, accessed 01/03/2010
3. Lance Spitzner, Honeypots, Tracking Hackers. Addison Wesley, Boston, 2002
4. C. Leita, V. H. Pham, O. Thonnard, E. Ramirez-Silva, F. Pouget, E. Kirda and M. Dacier, The Leurrecom.org Project, Collecting Internet threats information using a worldwide distributed honeynet. June 2008
5. L. Spitzner, Know your enemy, Honeynets. AusCERT2004 Conference, Technical Stream, 2004
6. F. Pouget, M. Dacier and H. Debar, Honeypot, Honeynet, Honetoken Terminological issues. Technical report, Institute Eurecom and France Telecom RD, France, 2003
7. N. Provos, Honeyd, A virtual Honeypot Daemon, Center for Information Technology Integration, University of Michigan, 2002
8. M. Kaaniche, E. Alata, V. Nicomette, Y. Deswarte and M. Dacier, Empirical analysis and statistical modeling of attack processes based on honeypots, In conjuction with the International conference of dependable systems and networks. June 2006
9. F. Pouget, M. Dacier and V. H Pham, On the advantages of deploying a large scale distributed honeypot platform on the Internet, First Workshop on Quality of Protection, 2005
10. E. Alata and M. Dacier, Collection and analysis of attack data based on honeypots deployed on the Internet, First workshop on Quality of protection, 2004
11. F. Pouget, M. Dacier, P. T. Chen and C. S. Laih, Comparative survey of local honeypot sensors to assist network forensics, SADFE'05, 1st International Workshop on Systematic Approaches to Digital Forensic Engineering, Institut Eurcom and National Cheng Kung University, 2005
12. M. Hofman, DShield distributed intrusion detection system, http://www.dshield.org, accessed 23/05/2010
13. L. Baldwin, myNetWatchman, Network intrusion detection and reporting, http://www.mynetwatchman.com, accessed 29/12/2009
14. C. Shannon and D. Moore, CAIDA Project, The UCSD Network Telescope, The spread of the witty worm, www.caida.org/publications/papers/2004/witty, accessed 15/04/2010
15. C. Leita and M. Dacier, SGNET, A worldwide, deployment framework to support the analysis of malware threat models, Proceedings of the 7th European Dependable Computing Conference, 2008