Network Security: Experiment of Network Health Analysis At An ISP

Perpetus Jacques Houngbo *jacques.houngbo@imsp-uac.org*
Institut de Mathématiques et de Sciences Physiques (IMSP)
Dangbo, Benin

Abstract

This paper presents the findings of an analysis performed at an internet service provider. Based on netflow data collected and analyzed using nfdump, it helped assess how healthy is the network of an Internet Service Providers (ISP). The findings have been instrumental in reflection about reshaping the network architecture. And they have also demonstrated the need for consistent monitoring system.

Keywords: Network Monitoring, Information Security, Netflow, Nfdump, Nfsen, Worm, Bandwidth, Threat, Bot Botnet, Trojan, Behavior Analysis.

1. INTRODUCTION

Network health

Computer networks are nowadays at the heart of most information systems. They are valuable assets in those information systems. For the sake of the smooth continuity of the business, it is of paramount importance that computer networks operate and perform at their best. This includes their health in terms of information security.

Network health can be evaluated by the mean of network monitoring that cover activities like high-level health checks, diagnostics and capability assessments. Beyond those instant checks, network health expands to awareness raising, consensus creation and generation of commitment to act [1] [2]. These are strong recommendations from COBIT 5.

Views from ISP side

Internet Service Providers (ISPs) are in charge of networks of various sizes that generally share some few characteristics:

- large number of computers,
- multiple border routers,
- traffic from multiple administrative domains that are relatively or completely independent.

In such environment, it is not unusual tendency to not take responsibility of information security and to disclaim that that responsibility relies on the shoulders of administrators who take care of autonomous domains. Nevertheless, the super-fast speed of the spreading of worms, the nuisances they cause and the damages resulting from Distributed denial-of-service (DDoS) are introducing more responsible attitudes. All actors, especially ISP, are paying more attentions to cybersecurity threats [3].

Current situation in Benin: "little" care about network security, about network health
Very few information is currently available on how secure are computer networks in Benin. Discussions with network administrators reveal that security is not yet in their top ranked priorities, if one excludes the point of installing virus protections. Some vendor solutions are in place at some well equipped nodes.

There is a kind of culture of secrecy about any event happening in computer networks as the negative impact of any disclosure of information is perceived as very damaging. For *Omnium des Telecom et de l'Internet (OTI)* to accept to host this study on information security in Benin based on its own computer network is special mark of openness and also an indication of change in mind towards more focus on information security. This acceptation is highly appreciated and saluted. OTI was kind enough to validate the initial goals of the study and to highlight the two results that are of importance for them.

2. OBJECTIVES

The initial request for this work has announced the following main goals:

- establishment of user profiles:
    - unusual patterns will trigger investigation as potential evidence of threats
    - those benchmarks for normal traffic may also help in defining new services to be offered to clients
- detection of potential bots or command & control in the network
- listing types of malware moving around in the network, mainly identification of worms overusing bandwidth
- detection of signs of attacks.

The ability to detect worms that are poisoning bandwidth was one of the most exciting and expected results as it is of interest for the ISPs. ISPs are almost overwhelmed by their clients complaining about actual bandwidth and many times it turns out that there are unmanaged traffics going through.

3. RELATED WORK

This section briefly presents some of the research literature related to "network security". Authors interesting in "network security" have mostly focused their attention on using netflows to detect intrusions, anomalies, botnets. Detecting adverse events [4] [5] [6] [7] is very important for network security. Some mathematical models have been elaborated for that purpose and they have been experimented using fuzzy data or real data [8] [9] [10]. The raw material used in all those research is netflow data. An overview of the research on network anomaly detection methods and network intrusion detection methods and systems is presented in [11] and [12]. The tools used to capture flows of traffic are Cisco Netflow V.9 [1], Nfdump [13] [14], and NfSen. Nfdump/Nfsen are presently the most widely used tools [15]. Experts are offering a variety of plugins to extend the functionalities of those two tools[2].

This paper will present practical application to the situation of an ISP in Benin.

4. HYPOTHESIS

The hypothesis of this paper is labeled as follows: known behaviors of infested hosts can be detected by analyzing some components of the netflow they generate namely kind of traffic,

1 http://www.cisco.com/en/US/products/ps6645/products_ios_protocol_option_home.html

2 http://www.muni.cz/ics/research/projects/4622/web/?lang=en

volumes, number of packets per flow, packets sizes, ports used, DNS queries. Based on the fact that this experimentation is conducted at an ISP, it also relies on the assumption that in such network, there are very few servers in the ISP's clients hosts. ISP's clients usually mainly have workstations.

5. METHODOLOGY

5.1 Data Collection

5.1.1 Tools and Concepts

netflow
netflow is a protocol developed by Cisco. It is used for the collection and monitoring of network traffic flow data generated by routers and switches. The protocol evolved to be largely adopted by the industry. The NetFlow Version 9 is the one standardized by IETF[3].

nfdump
nfdump is a set of tools developed for the purpose of capture, storage and analysis of netflow data. nfcapd is the netflow capture daemon. It reads the netflow data from the network and stores the data into files. nfdump is the netflow dump. It reads the netflow data from the files stored by nfcapd; it performs many types of calculations and displays netflow data based on IP addresses, ports, volumes, etc.

nfprofile
nfprofile reads the netflow data from the files stored by nfcapd, filters the netflow data according to the specified filter sets (profiles) and stores the filtered data into files for later use.

nfreplay
nfreplay is used to read the netflow data from the files stored by nfcapd and to send it over the network to another host.

nfsen
nfsen is a graphical web based front end for the nfdump set of tools. Its goal is to ease (i) the display of netflow data (IP addresses, ports, volumes, etc.), (ii) the navigation through the data, (iii) the process of the netflow data within specified time span and (iv) the definition of alerts.

5.1.2 Configuration of The Capture of The netflow Data

Omnium des Telecom et de l'Internet (OTI) has graciously dedicated a computer for the capture, storage and analysis (partly) of the netflow data. That computer has been installed in a sub-LAN designed for that purpose. The netflow data has then been mirrored from a Microtik router to the collection computer.

The capture covered the time period from 02 July 2013 to 17 October 2013. The result of which is 222 GB of data, average of 3 GB daily.

5.2 Data Analysis Methods

The analysis performed a flow analysis. It would be somehow nice to run the analysis when having clear identifications of what are to be sought. In the conditions of this study, the difficulty was to define what to search. The option then comes up to let the data themselves crop up information and thereafter to dig about those information. The next step was then to apply some queries collected from related work.

[3] https://tools.ietf.org/html/rfc7011 and https://tools.ietf.org/html/rfc7012

The information that pops-up was about volume of traffic and this lead to question of knowing and listing those IP that are on the top of volumes of traffic, number of flows, number of packets, bandwidth usage.

In order to get closer to the actual scheme of the real activity on the network, the first instance observations have been confronted and validated by the network administrator of OTI. Some special questions have triggered fine tuning analysis.

For specific data to be analyzed, the following have been used.

Port scans & Port sweeps
Ask and Skrautvol [16] have established a table of anomalies defined by packet size and the number of flows. Port scanner processes by probing a host for open ports while port sweeping scan multiple hosts for a specific listening port.

Anomaly	Defined by packet size and number of flows
Dos	Small packet sizes, usually < 100 bytes. Often only one packet per flow.
Port scans & Port sweeps	Small packet sizes, often < 55 bytes. Often one packet per flow.

TABLE 1: Description of anomalies by packet size and number of flows.

For the purpose of analysis, this study has translated the table above into filters to be applied to the data collected.

Anomaly	Translation into nfdump filter in case of local network being victim	Translation into nfdump filter in case of local network being perpetrator
Dos attack	(dst net local_network) and (bytes <100) and (flows 1) and (packets < 2)	(src net local_network) and (bytes <100) and (flows 1) and (packets < 2)
Port scans & Port sweeps	(dst net local_network) and (bytes <55) and (flows 1) and (packets < 2)	(src net local_network) and (bytes <55) and (flows 1) and (packets < 2)

TABLE 2: Filter component of script - detecting anomalies by packet size and number of flows

Botnet scanning
Pavel Čeleda [15] has published a best practice document that offers some useful points:

Horizontal scans usually are performed by initiating new connection against every possible IP address in the targeted network. The connections use SSH port (22) or Telnet port (23) or HTTP port (80). Because the connection do not complete the three way handshake of TCP, they contain only SYN (S) or RESET (R) TCP flags, but not ACKNOWLEDGE (A) flag.

This study has derived the following filters:

Anomaly	Translation into nfdump filter in case of local network being victim	Translation into nfdump filter in case of local network being perpetrator
Botnet scanning potentially for Chuck Norris botnet	(dst net local_network) and (dst port 22 or dst port 23) and (proto TCP) and ((flags S and not flags ARPUF) or (flags SR and not flags APUF))	(src net local_network) and (dst port 22 or dst port 23) and (proto TCP) and ((flags S and not flags ARPUF) or (flags SR and not flags APUF))
Botnet scanning potentially for Conficker worm	(dst net local_network) and (dst port 445) and (proto TCP) and ((flags S and not flags ARPUF) or (flags SR and not flags APUF))	(src net local_network) and (dst port 445) and (proto TCP) and ((flags S and not flags ARPUF) or (flags SR and not flags APUF))

TABLE 3: Filter component of script - detecting botnet scanning

Analysis of the bot code [17] is a complex activity; the same apply to the botnet communication [18] and to the topologies of command-and-control (C&C) infrastructure [19]. Those analyses were not performed for this study.

DNS spoofing attacks

Detecting possible DNS spoofing attacks is performed by tracing flows with source port 53 that go to IP addresses outside the local network and different from well known Public DNS servers with good reputation like the Google Public DNS 8.8.8.8 and 8.8.4.4 or the OpenDNS 208.67.222.222 and 208.67.220.220.

Anomaly to be detected	Filter component of nfdump script
DNS spoofing attack	(src net local_network) and (dst port 53) and (proto UDP)

TABLE 4: Filter component of nfdump script - detecting DNS spoofing attack

The result of that query will be checked against the list of well known Public DNS servers with good reputation provided by the namebench utility [20].

Long lived flows are characterized to last more than 24 hours [21], hence the combination of timestamp and filtering of lines where the third column, the duration of the flow, is over 86400 seconds.

In their research on detecting botnets, Vojtech Krmíček and Tomas Plesník [22] [23] have presented five detection methods:

- telnet scan detection
- connections to botnet distribution sites detection
- connections to botnet C&C centers detection
- DNS spoofing attack detection
- ADSL string detection.

This paper has already explored telnet scan detection and DNS spoofing attack detection. As for the connections to botnet distribution sites detection, the connections to botnet C&C centers detection and the ADSL string detection, they imply availability of some specific elements such as

IP addresses of attacker's botnet distribution web servers or IP address of an attacker's IRC server (Botnet C&C center). Those elements are not yet available at this moment and the related detections have not been applied.

6. DATA ANALYSIS RESULTS

All the data analysis was done manually. Some command lines were flashed against sample data from the large amount of data collected. The command lines are often tuned with additional parameters and filters. When the output is expressive enough, a script is edited that will be executed on the whole data, one script per month of data.

At the end of data collection process, the computer that OTI has provided for that purpose has been devoted to other activities. The main data analysis has then been performed on a simple laptop. Due to the limitations in performances, the analysis unfortunately suffered from long time consumption before result of script are finally written in output file for further reading.

6.1 Statistical Observations

Indicators like top 10 volume of traffic, top 10 hosts in bandwidth usage have created the different categories of users of OTI's network. When paired with time of observation, that information have been confirmed by the network administrators as related to expected behaviors of the identified sub-networks based on the IP in use.

6.2 Anomalies Detections

The results of data analysis presented below only focus on four types of anomalies detection: trojan activities, network scanning, botnet scanning, DNS spoofing.

Trojan activities

Many trojans has been reputed to be using well known TCP/IP ports for their specific communication when they get into a system. Unfortunately, there is no authoritative body responsible of updating a master list of trojan ports. But several users offer lists that present more or less identical list of trojan ports. This study relied on the port list provided by http://www.emsisoft.com/en/kb/portlist/.

Before jumping on conclusions about the eventual trojan activities in the network, it is important to talk about the port address translation (PAT) system that OTI has implemented.

Port address translation (PAT) is defined to translate multiple real addresses to a single mapped IP address. The real address and source port are translated to the mapped address and a unique port. Port address translation (PAT) uses available ports above 1024.

Because PAT uses available ports, it is important to understand the way the system in place at OTI is working: what is the pool of ports available for use, how can some ports be excluded from mapping, etc. Getting further information about PAT usage at OTI was not easy as it implies some high degree of availability of the IT personnel for presenting diagrams and explaining deeper their network architecture. Moreover, disclosing that information could also open doors for breaches in confidentiality of their information assets. This is the main reason why this study has not really expressed the need for deep understanding of PAT usage at OTI.

As a consequence, even though many activities have been recorded using port that are on the list of well known as potential trojan ports, those “suspicious” activities have not been investigated further.

Network scanning: port scanning and port sweeping

The kind of DoS attacks, port scans and port sweeps described by [16] seem to be very rare in the data collected. This does not mean that no port scans have been executed; “perpetrators” may have used other techniques.

Botnet scanning
The netflow data showed that some few host were performing intense horizontal scans activities as if there were infected with Chuck Norris or Conficker worm. Three hosts were particularly aggressive and at least for one of them, checking against the whois database produced useful information that has been handed to the network administrator.

On the other side, the number of host attacking using those two worms is very high, more than 108085 hosts, while only 845 hosts in the network concerned by the data collected are under those horizontal scans at the time of data collection.

DNS spoofing
The analysis revealed kind of massive DNS activities: volume, number of hosts involved. 193 hosts have solicited more than 52790 DNS servers for name resolution. That list of probable DNS servers has been checked against the 4515 name servers used by namebench. Only 697 of them match. This means that there are 193 hosts in the network that are using the name service on more than 52000 name servers that may not be trustworthy! There is no way to guarantee that none of those more than 52000 name servers is malicious server.

7. STUDY RESULTS

7.1 Most Effective Ways of Detection

Methodologies of detection usually come after perpetrators have acted and those in charge of defense have some indicators they can track. Moreover of detecting systems and hosts experiencing bad health conditions, ideal situation is to be able to recognize elements of advert events before they damage the business. The few analysis performed by this paper reinforce the concept of constant and rigorous monitoring. This implies resources that the manager will devote to a wise and wide implementation of information security because the manager understands the value the business is getting from that implementation.

7.2 Usability of The Results

Sharing with the network administrator has shown clear usability of the results. The first impact is the reflection induced that can be compared to the “check” stage of the Plan - Do - Check - Act of the Deming wheel. Some discussion about the network architecture and also about some usage from the clients has started and it is expected that they lead to a new level of fine tuning the architecture of the network. One of the most important concerns[4] of an ISP is the perception the client has of its services. Availability and reliability come immediately in mind.

The usable output from this study is a set of recommendations to the administrators on their request. These recommendations clearly mentioned actions that are expected to improve their mastering of their network. They read as:

- fine tuning of the network architecture in order to clarify
    - situation of hosts that are having lot of traffic on port 23 Telnet
    - situation of hosts that seem to be recursively trafficking with themselves
- set the parameters of the PATing system so that identification of some malicious activities like Trojan by the ports used can be easier
- investigate further on sub networks where potential Conficker worm and Chuck Norris worm activities have been detected
- participate in controlling botnets and their command and control (C&C) by implementing a

4 Generally speaking, concerns of ISPs seem to be only focused on lowering capital expenditure while ensuring the highest benefits. ISPs think in terms of (i) high capacity networks at lower prices, (ii) subscribers per tower to keep CAPEX to the least possible, (iii) extension of service footprint to new geographies quickly and cost-effectively and (iv) quickest payback time or time-to-revenue. And eventually, they also think of customer satisfaction.

monitoring of connexions between hosts and known C&C IP addresses as confirmed by famous vendors like Trend Micro or Team Cymru.

7.3 Accuracy of The Results

This document has presented results based on the data collected. The quality of the findings obviously depends on the conditions of the collection and the accuracy of tools used for the analysis. The current situation is that there are very few statistics for usage as baseline one can refer to. The current findings display the characteristics of that network at a time. Improvements are definitely needed in collection, storage and fast analysis of the data so that findings can immediately lead to appropriate actions. The long delay (almost nine months) between the time of collection of the netflow data and the release of this analysis reinforce the need for fast analysis.

7.4 Results Compared To Research Objectives

Establishments of user profiles

The user profiles have not been established at this stage, they have not been disclosed for this paper. Before this can be proceeded, there is a need to access some classified information that OTI was not willing to disclose.

Detection of potential bots or command & control in the network

This analysis clearly showed some potential bot activities. Even if the issue of delay between the time the netflow data have been collected and the moment where the findings are disclosed is ignored, OTI is not yet in the mood of providing further assistance to its clients on such issues as there is a significant cost impact. And there is at this time no guarantee that clients will accept to cover those “additional” costs.

Listing types of malware moving around in the network

There are clearly signs of malware operating in the network. Once again, the issue is the cost of assistance from OTI and the willingness of the clients to cover the costs.

Detection of signs of attacks

Even though potential bots and malware are identified, at this stage of the analysis of the netflow data, it is difficult to present a clear signs of attacks where perpetrators are identified and victims as well.

8. CONCLUSION

From the findings of this analysis, it clearly pops-up the need for further work on different axis. The first one is the implementation of a consistent monitoring system that will operate in a way of contribution to long term analysis as well as flashing alerts that will trigger immediate actions. The second axis concerns fine tune of analysis by, for example, cross-referencing analysis of DNS spoofing with observations of Passive DNS for bot detection.

The collection of the netflow data for this analysis has started July 2013 and ended September 2013. Since that time, the landscape of the network OTI is managing has certainly evolved a lot. But this analysis has contributed to pinpoint some key elements for improvement. It has also demonstrated the possibility to implement a layer of monitoring at very low cost. The way forward is to elaborate on how, while using the findings of that monitoring, OTI can convert potential assistance to its clients into cash.

9. FUTURE RESEARCH

As mentioned in the conclusion paragraph, future research can be in term cross-referencing analysis of DNS spoofing with observations of Passive DNS for bot detection. The objective is to be able to detection suspicious systems before they impact negatively the whole network. This kind of services will be of increasing importance as the quality of internet services is improving and as it is envisioned that more private companies will have their production system online.

10. REFERENCES

[1] Information Systems Audit and Control Association. ISACA, *Cobit 5: A business framework for the governance and management of enterprise IT.* Rolling Meadows. IL, 2012.

[2] K. Singh, R. S. Yadav, and Ranvijay, "A review paper on ad hoc network security," *Int. J. Comput. Sci. Secur.*, vol. 1, no. 1, p. 52, 2007.

[3] C. Gates, J. McNutt, J. B. Kadane, and M. Kellner, "Detecting Scans at the ISP Level," DTIC Document, 2001.

[4] A. H. M. M. Uddin, "Detecting Botnets Based on their Behaviors Perceived from Netflow Data," 2009.

[5] R. Schoof and R. Koning, "Detecting peer-to-peer botnets," *Univ. Amst.*, 2007.

[6] Y. Singh, Y. Chaba, and P. Rani, "Integrating–VPN and IDS–An approach to Networks Security," *Int. J. Comput. Sci. Secur.*, vol. 1, no. 3, p. 1, 2007.

[7] M. Gandhi and S. K. Srivatsa, "Detecting and preventing attacks using network intrusion detection systems," *Int. J. Comput. Sci. Secur.*, vol. 2, no. 1, pp. 49–58, 2008.

[8] J. Vykopal, "Flow-based Intrusion Detection in Large and High-Speed Networks," PhD thesis, 2010.

[9] J. Vykopal, "A Flow-Level Taxonomy and Prevalence of Brute Force Attacks," in *Advances in Computing and Communications*, Springer, 2011, pp. 666–675.

[10] M. Elich, "Flow-based Network Anomaly Detection in the Context of IPv6," Sep. 2013.

[11] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Network Anomaly Detection: Methods, Systems and Tools," *IEEE Commun. Surv. Tutor.*, pp. 1–34, 2013.

[12] S. H. C. Haris, G. M. Waleed, R. Ahmad, and M. Ghani, "Anomaly detection of IP header threats," *Int. J. Comput. Sci. Secur.*, vol. 4, no. 6, p. 497, 2011.

[13] P. Haag, *nfdump and NfSen*. 2010.

[14] P. Haag, "Watch your Flows with NfSen and NFDUMP," 2005.

[15] P. Čeleda, "Network Security Monitoring and Behavior Analysis," 2011.

[16] M. B. R. Ask and H. Skrautvol, "Anomaly Detection and Identification in Feature Based Systems: An Empirical Evaluation," Norwegian University of Science and Technology, 2011.

[17] R. Link and D. Sancho, "LESSONS LEARNED WHILE SINKHOLING BOTNETS - NOT AS EASY AS IT LOOKS," 2013, pp. 106–110.

[18] G. Ollmann, "Botnet communication topologies," *White Pap. Damballa*, 2009.

[19] J. Vania, A. Meniya, and H. Jethva, "Association Rule Based Data Mining Approach to HTTP Botnet Detection," Sep. 2013.

[20] *namebench - Open-source DNS Benchmark Utility - Google Project Hosting*. 2014.

[21] M. Campbell, *Collecting and Analyzing Flow Data for Large Networks*. .

[22] V. Krmıcek, “Hardware-Accelerated Anomaly Detection in High-Speed Networks,” Sep. 2013.

[23] V. Krmíček and T. Plesník, *Detecting Botnets with NetFlow*. 2014.