Information-leakage in NDN: Detecting Anomalous Names

Daishi Kondo[1,2], Thomas Silverston[3]*, Hideki Tode[4], Tohru Asami[5] and Perrin Olivier[1,2]

Information leakage is one the main security threats in today's Internet. It can have a significant impact on companies especially by reducing profits and destroying reputations. As Named-Data Networking (NDN) is a promising alternative for the future internet, it is essential to prevent this security threat.

NDN relies on a new networking paradigm based on content name. Indeed, today's users are interested in content and not location, and there is a need for a shift from a host-to-host communication paradigm to a host-to-content one. NDN content names are defined with the traditional URL format commonly used in the Internet.

In this work, we propose a novel filtering technique to detect packets with malicious names. Indeed, malicious names are more likely to be generated by malwares through "Targeted Attacks" in order to leak out information from legitimate networks. The filters will be used for the NDN firewall as they cannot rely on IP address anymore.

We have performed a comprehensive statistical study of URLs based on extensive crawling experiments of main Web organizations. From our experiments, we have derived filters, which were able to detect 15% of malicious names in our data set. This is an essential step towards preventing information leakage in NDN.

INTRODUCTION

Information-leakage is one of the main security threats in today's Internet [1]. Indeed, cyber espionage through targeted attacks have been viewed as major attacks against companies, whatever their size and profit. Recently, IT companies such as Sony and retailers such as Target suffered from massive information leakages. As a consequence, the Target data breach in 2013 caused the company to spend more than $100 million upgrading their system to prevent another breach and they also suffered from a 46% drop in profits after the attack itself [2]. These data breaches through targeted attacks relied mostly on malwares installed via emails, websites or external memory devices. They allowed attackers to obtain the confidential information of companies. Network administrators have had to rely on firewalls and filtering services to drop confidential information leaking packets.

Information Centric Networking (ICN) relies on a new networking paradigm based on content name. Indeed, as today's users are interested in content and not its location (as with TCP/IP), there is a need for a shift from a host-to-host communication paradigm to a host-to-content paradigm. In this context, Named-Data Networking (NDN) [3] has gained much attention and formed an increasing research community. It has been implemented into the NDNx architecture and relies on two messages: (i) Interest, a request sent by a user towards a named content, and (ii)

[1]University of Lorraine, LORIA (CNRS UMR 7503), France
[2]Inria Nancy – Grand Est, France
[3]The University of Tokyo, Dept. of Computer Science, JFLI, Japanese-French Laboratory for Informatics, Graduate School of Information Science and Technology 7-3-1 Hongo, Bunkyo-Ku, 113-8656 Tokyo, Japan
[4]Graduate School of Engineering, Osaka Prefecture University, Japan
[5]Graduate School of Information Science and Technology, The University of Tokyo, Japan
*Corresponding author (Thomas Silverston, e-mail: thomas.silverston@is.s.u-tokyo.ac.jp)

Data, a response to a request sent by any nodes possessing the named content. NDN content names are defined in traditional URLs as those existing in today's Internet. The network is responsible to cache content for further requests and helps in delivering the content to users. NDN messages exhibit the names of the contents to be exchanged and the network administrator can explicitly control the traffic through the firewall by their names [4].

This work aimed in investigating the information leakage security threat in NDN. Assuming that names in NDN will follow current URLs in the Internet, we have performed extensive web crawling experiments to collect URLs and provide a statistical analysis of web URLs. From these experiments, we have proposed a new filtering technique to detect anomalous names in NDN.

TARGETED ATTACKS

Targeted attacks enable the leaking of information from a network. The rationale of these attacks is to enforce a legitimate user of the network to set up malware on his computer. To this end, an attacker can prepare emails, websites or external memory devices in order to compromise a computer. After the attacker has succeeded in infecting computers, he can establish communication channels to a Command & Control server and can access network internal computers, and bypass firewall security to obtain information.

Target attacks in NDN can also be performed to set up malwares. Fig. 1 shows a targeted attack causing information leakages in a NDN network. Malwares in the network can send Interest packets to bots, by specifying bots' name prefixes into packet names.

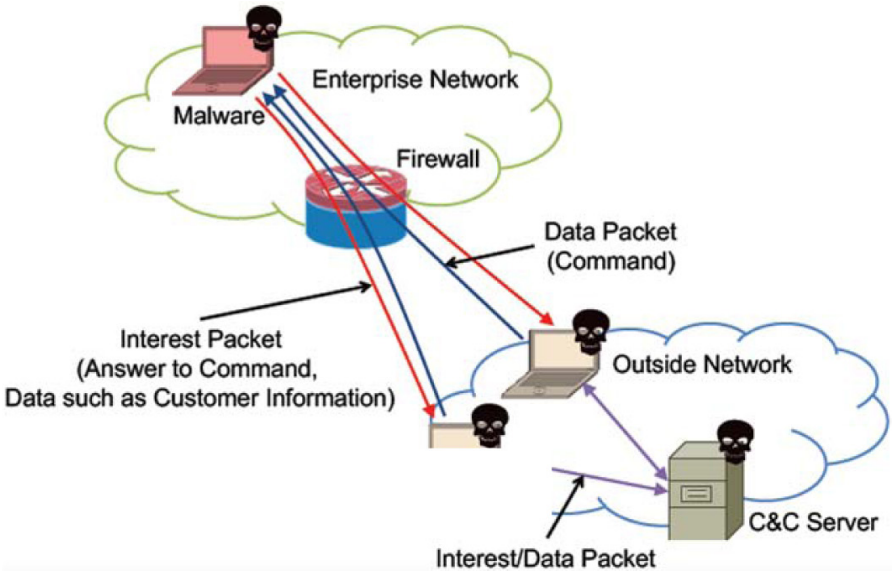

Figure 1: Targeted Attacks in NDN

WEB CRAWLING EXPERIMENT

As URLs are commonly used in the current Internet, it is highly probable that URLs will be used for names in NDN. Thus, in order to investigate names in NDN, we sampled URLs from seven main Internet organizations according to Alexa TOP 500 Global Sites [5] namely Amazon, Ask, Stack Overflow, BBC, CNN, Google. We used a crawler browsing URLs of each organisation from their homepage and went across all URLs in the page. We present in Table 1, the number of URLs collected from each organisation.

Table 1. Number of URLs collected from each organisation

Organisation	Number of URLs
Amazon	10,020
Ask	6,388
Stack Overflow	19,133
BBC	8,936
CNN	2,749
Google	6,738
Yahoo	5,744

URLs follow the format provided in RFC 1808 [6] and we have extracted several parameters as shown in Figure 2: Length of Path, Length of Query, Length of Fragment, Length of Directory Name, Length of File Name, Number of slash character in Path. We also computed the Cosine similarity for the Path, Query and fragment part of the URLs. This measure shows the similarity between parts of the URLs among all collected URLs of all organisations.

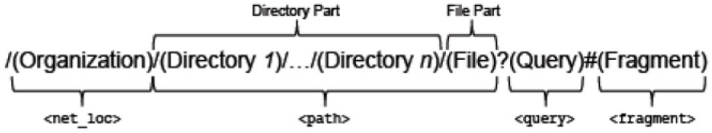

Figure 2. Format of URLs

RESULT OF EXPERIMENTS

For each URL we computed the cumulative distribution function for each attribute. For clarity, we have only shown in Figure 3, an example for the Length of Path of URLs. For instance, 95% of the URLs' Length Path did not exceed 100 characters.

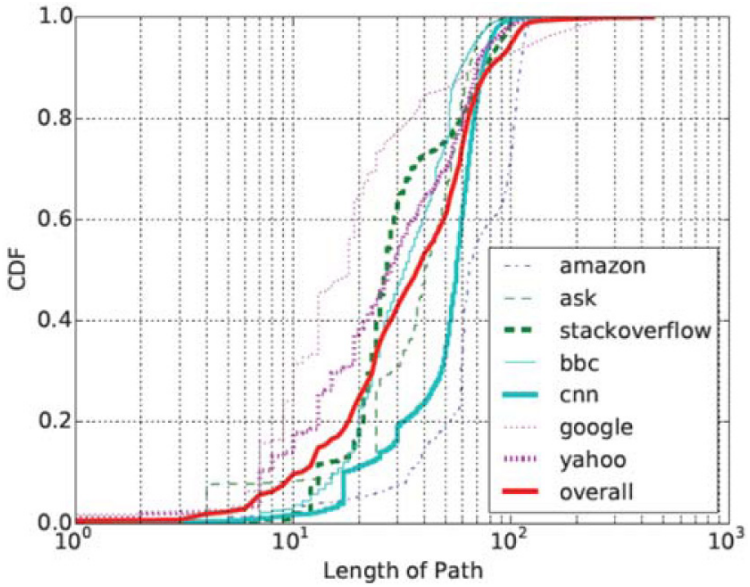

Figure 3. Length of Path of URLs (CDF)

Then, we created a filter based on our observation of all the URLs attributes and similarities. In other words, a URL was considered anomalous if one of its attributes did not belong to the 95th percentile of the measured variable, or if the similarity between parts of the URL was inferior to the average. Thus, our filter wasis able to detect that 15% of URLs were anomalous, i.e., one of the attribute values exceeded the 95th percentile, or the similarity with other URLs was below average for all URLs.

DISCUSSION

The proposed filter was based on statistics from our URL crawling experiments. Basically, filters could help determine if the name of an Interest packet is legitimate or anomalous. Indeed, in case of targeted attacks in an NDN network, if the names created by a malware were detected as anomalous, packets would be dropped and the attack would be prevented. As 15% of name anomaly detection was still important for practical operations, this was an important step toward filtering names into the NDN.

Note that a malware can also learn the statistics rules used by our filter (length of path, length of fragment, etc.) To overcome this limitation, each network operator should monitor names in their network and should tune the threshold for the attributes.

CONCLUSIONS

As it is already the case for the current Internet, Information-leakage will be one of the main security threats in NDN. We therefore propose a filter to detect anomalous names in NDN, as these packets are more likely to be created by malwares to leak information from the network.

Our filter is based on URL statistics obtained through extensive crawling experiments of major Internet organisations. Our filter can detect 15% of malicious names in our dataset which can be practically used in real-world network operations, thus it is an important step toward detecting anomalous names in NDN and preventing information leakage. We are now refining our filter to be practically deployed within firewalls for NDN.

REFERENCES

IT Security Risks Survey 2014: A Business Approach to Managing Data Security Threats, http://media.kaspersky.com/en/it_security_risks_survey_2014_global_report.pdf

Understanding Targeted Attacks: The Impact of Targeted Attacks, http://www.trendmicro.com/vinfo/us/security/news/cyber-attacks/the-impact-of-targeted-attacks

L. Zhang, A. Afanasyev, J. Burke, V. Jacobson, K. Claffy, P. Crowley, C. Papadopoulos, L. Wang, and B. Zhang, “Named Data Networking”, ACM SIGCOMM Computer Communication Review, vol. 44, no. 3,pp. 66–73, Jul. 2014.

D. Goergen, T. Cholez, J. Francois, and T. Engel, “A Semantic Firewall for Content-Centric Networking,” in Proc. 2013 IFIP/IEEE International Symposium on Integrated Network Management (IM 2013), pp. 478–484, May 2013.

Alexa Top 500 Global Sites, http://www.alexa.com/topsites

RFC 1808, https://tools.ietf.org/html/rfc1808