Name anomaly detection for ICN

Daishi Kondo, Thomas Silverston, Hideki Tode, Tohru Asami, Olivier Perrin

Name Anomaly Detection for ICN

Daishi Kondo*†, Thomas Silverston‡, Hideki Tode§, Tohru Asami¶ and Olivier Perrin*†
*University of Lorraine, LORIA (CNRS UMR 7503), France
†Inria Nancy - Grand Est, France
‡The University of Tokyo, JFLI (CNRS UMI 3527), Japan
§Graduate School of Engineering, Osaka Prefecture University, Japan
¶Graduate School of Information Science and Technology, The University of Tokyo, Japan
Email: {daishi.kondo, olivier.perrin}@loria.fr, thomas.silverston@is.s.u-tokyo.ac.jp,
tode@cs.osakafu-u.ac.jp, asami@akg.t.u-tokyo.ac.jp

***Abstract*—Information leakages are one of the main security threats in today's Internet. As ICN is expected to become the core architecture for Future Internet, it is therefore mandatory to prevent this threat. This paper proves that some ICN configuration prevents information leakages via Data packets and shows that it is an open problem to prevent interest packets from carrying encoded crucial information in their names. Assuming that names in ICN will follow the current URL format commonly used in the Internet, we get the statistics of web URL based on extensive crawling experiments of main internet organizations. Then we propose a simple filtering technique based on these statistics for firewall to detect anomalous names in ICN. The experiment shows that our filtering technique recognizes 15% of names in our dataset as malicious. As the false positive rate is still high for this filter to be used in a real world operation, this work is an important step for detecting anomalous names and preventing information-leakage in ICN.**

I. Introduction

Information-leakage is one of the main security threats in today's Internet [1]. Indeed, *Cyber Espionage through Targeted Attacks* is viewed as the major attack for companies, whatever their size and profit. Recently, IT companies such as Sony or the retailer Target suffered from massive information leakages. As a consequence, Target data breach in 2013 lets the company spend more than $100 million upgrading the system to prevent another breach and suffer from 46% drop in profits after the attack itself [2]. These data breaches through Targeted Attacks rely mostly on a malware installed via emails, websites or external memory devices. They allow attackers to obtain confidential information of companies.

As today's Internet traffic is going to be composed mostly of encrypted HTTPS traffic, it is therefore impossible for a company to know precisely the data that is transported. Network administrators have to rely on firewall and filtering services to drop confidential information leaking packets while a Targeted Attack consists of a malware to leak confidential information out of the network as legitimate data not to be dropped by security infrastructure of the company.

Information Centric Networking (ICN) relies on a new networking paradigm based on content name. Indeed, today's users are interested in content and not its location as with TCP/IP, and there is a need for a shift from host-to-host communication paradigm to host-to-content paradigm. In this context, Named-Data Networking (NDN) [3] has gained attention and forms an increasing research community. It has been implemented into NDNx architecture and relies on two messages: (i) Interest, a request sent by a user toward a named content, and (ii) Data, a response to a request sent by any nodes possessing the named content. NDN content names are defined in traditional URI as those existing in today's Internet. The network is responsible to cache content for further requests and helps delivering the content to users wherever they are. 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.

This paper aims at investigating the information-leakage security threat with names in ICN network. To the best of our knowledge, this paper is the first to investigate the information leakage security threat with names in ICN and provides a novel technique to detect anomalous names. Our contributions are fourfold: (1) For "Pull"-based ICN such as NDN, we prove there exists a firewall configuration that prevents information leakages through Data packets; (2) As the countermeasure of the above firewall, we propose a model of Targeted Attack using Interest packets, which encodes leaked information into content names; (3) To block anomalous names created by these attacks, we have to know the features of normal content names. Assuming that the relevant statistics of NDN-like content names can be predicted from those of the current URLs in the Internet, we conduct a comprehensive study of web URL based on extensive web crawling; (4) We propose a new filtering technique to filter out anomalous NDN names, using these statistics, such as length of path, etc.

The paper is organized as follows: Section II states the problem by presenting ICN architecture, ICN firewall, the Targeted Attack in ICN network, and the naming policy through URL. Section III presents our web crawling experiment on URL, while Section IV presents statistical analysis on URL name and our filtering techniques to detect anomalous name in NDN. Section V surveys the related work and Section VI concludes the paper and presents research perspectives.

II. Problem Statement

This section shows the overview of ICN, and explains that a configuration of NDN firewall can prevent information leakages through Data packets but it is possible for an attacker to bypass the NDN firewall through Interest packet. Then, we propose a model of Targeted Attack causing information

leakages in ICN network and discuss the future naming policy in ICN.

A. Information Centric Networking

ICN basically adopts a "Pull"-based network architecture. To retrieve information, at first a node sends an Interest packet (i.e., a request) to the network and then obtains the Data packet corresponding to the Interest packet, which is a reply packet from the network. Along the delivery path, nodes can store the transmitted information in order to serve further requests, making the network able to provide information and reply to Interest packet. In other words, nodes cannot send a Data packet unless they receive an Interest packet, and they can get a content from any nodes who has it such as the publisher and network cache. The following discussions assume that ICN has this property which is common in NDN.

In NDN, there are three components in the node; Forwarding Information Base (FIB), Pending Interest Table (PIT), and Content Store (CS). FIB has the routing information to forward the Interest packet. PIT maintains the return path state to send the Data packet corresponding to the Interest packet. CS is utilized to cache the Data packet based on the caching replacement policy.

B. ICN Firewall and its Policy Maker

The basic framework of our firewall assumes the following assumptions: A1) Operators have the naming policy and routing policy for all the information assets inside the enterprise network, A2) Operators can announce to the outside network what content name in the enterprise network can be accessed from the outside network, but do not have any right for the naming policies outside their network, A3) A malware has the list of clients of this company.

It is easy to prevent the information leakages through Data packets in NDN as follows: i) define the name prefix for public contents as "/company/pub/" (A1), ii) announce "/company/pub/" to the outside network (A2). In this case, if the malware names the leaked list as "/company/pub/ABC", then the Interest packet can reach the malware from the outside network via the gateway. This can be blocked by the next rule. iii) Any Interest packet from the outside network shall not be relayed toward the inside network (A1).

Even though the malware sends a Data packet toward the gateway, the gateway cannot send it toward the outside network since there are no PIT entries for that Data packet. If the malware announces the leaked list as "/company/pub/ABC" inside the network, the gateway may cache this content during the communications among the nodes inside the network. This threat can be blocked by the next rule. iv) The gateway shall not relay an Interest/Data packet whose name prefix is "/company/pub/" from the inside network back to the inside network (A2). To realize the communications inside the network, the nodes utilize the name prefix such as "/company/priv/" for the private content, which cannot be accessed from outside.

In this case, all the publicly-accessible contents are on the gateway. To control the content of its own, the following rule is required. v) All the publicly-accessible content must be listed in the white list (A1). These five rules can shut out the information leakages through Data packets.

This is one of the main advantages of ICN or NDN compared to HTTPS. The Internet architecture is based on host-to-host communication paradigm, and so does content delivery. NDN is a chance for each company to fully control their information assets. It is usual to declare the advantages of NDN such that it provides the information integrity by signature, it controls encryptions if required, and efficient content caching is available using Attribute-based Encryption [4]. Table I shows the comparison between IP firewall for HTTPS traffics and ICN firewall since HTTPS traffics are becoming the majority of Internet traffic. In the case of content distribution through HTTPS, we can identify the location of the attacker as his host IP address. It is, however, computationally difficult to identify the attack since all the traffic are encrypted with SSL. Another disadvantage of SSL is that it prevents content caching to reduce the traffic. In contrast, content names are transparent at forwarding an ICN/NDN packet and it is possible to develop a various countermeasures against information leakages using this advantage (e.g., "/company/pub/" and "/company/priv/").

TABLE I
COMPARISON BETWEEN IP FIREWALL AND ICN FIREWALL

	IP Firewall	ICN Firewall
Identifying content sender	Partially (host IP address)	Yes (ID)
Identifying content receiver	Partially (host IP address)	No
Confirmation of content and its name	No	Yes
Cache availability	No	Yes
Information leakage through content name	NA	Partially

To remedy the disadvantage of HTTPS, D. Naylor et al. propose mcTLS (Multi-Context TLS) extending TLS [5]. If there is some consent between content provider and user, we can adapt middle boxes for caching HTTPS traffics. However, without the consent we cannot utilize middle boxes. Therefore, it is difficult for middle boxes to obtain information from an attacker, who does not agree to use them.

In ICN/NDN, however, gatekeepers do not hold the naming policy right on contents outside their private network. This has another risk of information leakages caused by encrypting the internal content into the content names of Interest packets to pretend to access the outside contents. Since no information on the attacker is included in these Interest packets, it is more difficult for the gatekeeper to detect the location of the attacker than in the case of the current IP firewall. Thus, it is important to develop the detection method of this attack.

C. Targeted Attack in ICN network

1) IP Network: Regarding the Targeted Attack in IP network, first of all, an attacker targets a network such as an enterprise network, and prepares emails, websites or external memory devices with a malware and a Command and Control server (C&C) [6]. C&C server is used by the attacker to send commands to the malware and control the behavior of the compromised computers. After the attacker has succeeded infecting, these computers establish communication channels to C&C server, which are called connect-back channels.

Through connect-back channels, the internal computers request connections to C&C server and allow the attacker to gain access into the internal computers. Connect-back channels are mainly used to bypass firewall security. Via these connect-back channels, C&C server sends commands to infected computers and control them remotely. Similarly, C&C server obtains leaked data from infected computers.

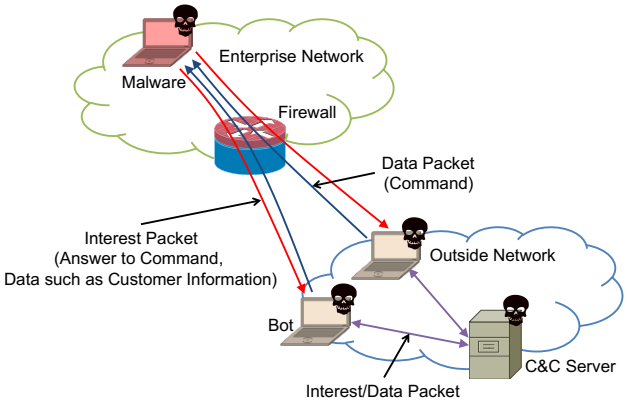

Fig. 1. Targeted Attack causing information leakages in ICN.

2) ICN Network: Fig. 1 shows a Targeted Attack causing information leakages in ICN. In advance, an attacker prepares C&C server, bots whose name prefixes are registered into the FIBs in the network, and a malware which knows the name prefixes to establish the communication channels with bots. Exploiting bots as springboards, C&C server does not need to register his name prefix into FIBs, which allows him to hide and protect himself. C&C server controls the malware remotely via bots.

Thus the malware in the enterprise network can send Interest packets to bots safely, by specifying bots' name prefixes into a part of content names for its first access. This Interest packet is relayed to C&C server via a connect-back channel after replacing the prefix into the one for C&C server. Then via this connect-back channel, C&C server responds by the corresponding Data packet with the appropriate commands to control the malware at will. After the connection setup, Interest packets are exploited to send the data such as customer information to C&C server.

D. ICN Content Naming Policy as Natural Extension from URL

In RFC 1808 [7], URL is defined as ⟨scheme⟩://⟨net_loc⟩/⟨path⟩;⟨params⟩?⟨query⟩#⟨fragment⟩. URL is created by a host shown by ⟨net_loc⟩ part. Each URL is unique in the world when a user or an organization owing this host makes this URL based on its own naming policy.

According to the report from Google [8], in 2008 there were 1 trillion unique URLs in the Internet. Users are accustomed to the current URLs. Considering the high affinity between the numerous contents named by URLs and users who are familiar with their URLs, it is highly possible for the future ICN content naming policy to become the one naturally evolved from the current naming policy of URLs. Naming syntax of CCNx 1.0 [9] as well as NDN [10] is similar to that of URL. Thus we can predict the naming policy of these ICN systems based on that of URL.

This future naming policy of CCNx or NDN may specify `ccn` into ⟨scheme⟩ part. Thus, an organization or a user described in ⟨net_loc⟩ part can define a name from the ⟨path⟩ part to the ⟨fragment⟩ part, independent from each other. This content naming policy is familiar with users and each generated content name is unique in the world. Moreover, it is very easy to translate the current numerous content names distributed in the Internet to the corresponding ICN names.

In CCNx 1.0 and NDN, they refer to URI written in RFC 3986 [11] and `lci` and `ndn` are specified into ⟨scheme⟩ part respectively. Above two naming policies, however, ignore ⟨authority⟩ part, which corresponds to ⟨net_loc⟩ part in RFC 1808. We consider that an authority described in ⟨net_loc⟩ part uniquely defines the name, even when the following parts of ⟨path⟩ may be defined independently. Therefore, this idea is different from the naming policies of CCNx 1.0 and NDN. However we assume the same discussions may be applied to those naming policies except ⟨authority⟩. Fig. 2 shows the content name naturally extending URL in this paper. However, `ccn`, which is specified in ⟨scheme⟩ part, is omitted. ⟨path⟩ part has two parts; directory and file part. We omit ⟨params⟩ part because mainly this part is the option in FTP.

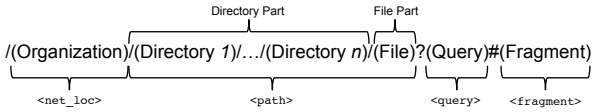

Fig. 2. Content name naturally extending URL [7].

III. Experiment

A. Web Crawling Experiment

In order to obtain statistics on name, we sampled the URLs created by 7 main organizations of the Internet, namely Amazon, Ask, Stack Overflow, BBC, CNN, Google, and Yahoo, which were high ranked in the Alexa Top 500 Global Sites [12]. To obtain the URL datasets, we utilized a breadth-first crawler [13]. Our breadth-first crawler starts from the seed URL of each organization (i.e., homepage) and goes across all the encountered URLs in the page; we stop the process when the crawler has reached 30,000 URLs. From these URLs, we keep in our dataset those returning the HTTP status code 200 (OK) with the name of each organizaition in ⟨net_loc⟩. Table II shows the number of URLs in dataset for each organization.

B. Dataset Presentation and URL Attribute

After obtaining our dataset, we divided it into two distinct sets: a training set and an evaluation set. The training set contains 90% of the URLs, and it is used for the statistical analysis of URL names (Section II-D). Moreover, from this training set we computed the average frequencies of characters in path, query, and fragment of the URLs in all the organaizations and defined them as *average frequencies in path, query, and fragment* respectively. The evaluation set (remaining 10%

TABLE II
NUMBER OF URLS IN DATASET OF EACH ORGANIZATION

Organization	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

of the URLs) were used for evaluating our Interest packet filtering as it will be presented in Section IV-B.

Table III shows 9 URL attributes. For some URLs, a file name sometimes lacks in file part because some default pages like index.html are often omitted. In such a case, we set the length of file name to 0. For each path, query, and fragment, we calculated the frequencies of characters in each URL and compared them with the corresponding parts in the whole URL datasets given by each organization using cosine similarity. We referred them as C_{Path}, C_{Query}, and $C_{Fragment}$.

TABLE III
9 ATTRIBUTES AND CORRESPONDING VARIABLES

Name	Variable
Length of path	L_{Path}
Length of query	L_{Query}
Length of fragment	$L_{Fragment}$
Length of directory name	$L_{DirectoryName}$
Length of file name	$L_{FileName}$
Number of slashes in path	$N_{Slashes}$
Cosine similarity of frequencies of characters in path with *average frequencies in path*	C_{Path}
Cosine similarity of frequencies of characters in query with *average frequencies in query*	C_{Query}
Cosine similarity of frequencies of characters in fragment with *average frequencies in fragment*	$C_{Fragment}$

IV. RESULT

A. Statistics on Name

Fig. 3, 4, 5, 6, 7, and 8 show the cumulative distribution function (CDF) of the length of path, query, fragment, directory name, file name, and the number of slashes, respectively, for each organization. Averaged CDF (i.e., overall in legend) has been normalized considering the differences between the number of URLs from each organization. Table IV shows the percentiles of each variable obtained from averaged CDF.

TABLE IV
PERCENTILES OF EACH VARIABLE

Variable	90%	95%	99%
L_{Path}	79	100	122
L_{Query}	127	162	283
$L_{Fragment}$	16	23	54
$L_{DirectoryName}$	15	21	41
$L_{FileName}$	46	57	90
$N_{Slashes}$	5	5	7

Table V shows C_{Path}, C_{Query}, and $C_{Fragment}$. Basically, the more closed to 1 the cosine similarity gets, the more similar those two vectors are. Average values of C_{Path}, C_{Query}, and $C_{Fragment}$ for the 7 organizations are 0.75, 0.68 and 0.55, respectively.

TABLE V
AVERAGE COSINE SIMILARITY OF FREQUENCIES OF CHARACTERS IN EACH URL TO CHARACTER FREQUENCIES IN PATH, QUERY, AND FRAGMENT PARTS AVERAGED OVER THE WHOLE DATASET IN EACH ORGANIZATION

Organization	Average C_{Path}	Average C_{Query}	Average $C_{Fragment}$
Amazon	0.76	0.73	0.5
Ask	0.76	0.86	0.57
Stack Overflow	0.77	0.76	0.4
BBC	0.74	0.56	0.6
CNN	0.81	0.54	0.63
Google	0.66	0.67	0.62
Yahoo	0.72	0.64	0.51
Average	0.75	0.68	0.55

In addition, we created another dataset with all the URLs in each organization, where we removed the first '?' identifying the start of query and also '#' identifying the start of fragment. Then, we calculated the cosine similarity of frequencies of alphabets in each dataset with typical English text. As for the frequencies of alphabets in typical English text, we use those defined in Ref. [14] (english-letter-frequency.dat). The highest cosine similarity of path was 0.98. This means that it is highly possible for the English words to be used in path. We confirmed that in actual URL datasets their readability was high (e.g., http://edition.cnn.com/specials/asia/on-the-road-japan). However, in the case of Stack Overflow, the cosine similarity was 0.51, and this value was the lowest because the word "tab-top" was often used in fragment.

B. Evaluation of Interest Packet Filtering

Assuming the naming policy of NDN is similar to that of URL, in this section, we propose a countermeasure against Targeted Attack with Interest packets, based on name anomaly. In this filter, we assume the normal name from our URL dataset belongs to the 95th percentile of each variable. Thus, based on the statistics from Section IV-A about L_{Path}, L_{Query}, $L_{Fragment}$, $L_{DirectoryName}$, $L_{FileName}$, and $N_{Slashes}$, we defined a filter ($F1$) as formula (1) referring to Table IV, where each variable is larger than the 95th percentile.

$$\begin{aligned} F1 = {} & (L_{Path} \geq 101) \vee (L_{Query} \geq 163) \vee (L_{Fragment} \geq 24) \\ & \vee (\min_{in\ path} L_{DirectoryName} \geq 22) \vee (L_{FileName} \geq 58) \\ & \vee (N_{Slashes} \geq 6) \end{aligned} \tag{1}$$

When an URL is filtered with $F1$ and the resultant $F1$ becomes true, this URL has an anomalous name and may cause information leakages. Thus Interest packet with this anomalous name has to be droped. We used our filter $F1$ on the evaluation dataset (10% of the remaining URLs), described in Section III-B. The false positive rate or the ratio of filtered names as anomalous in the datasets was 33%.

From the point of cosine similarity, we define another filter ($F2$) as formula (2) referring to Table V.

$$\begin{aligned} F2 = {} & (C_{Path} < 0.75) \vee (C_{Query} < 0.68) \\ & \vee (C_{Fragment} < 0.55) \end{aligned} \tag{2}$$

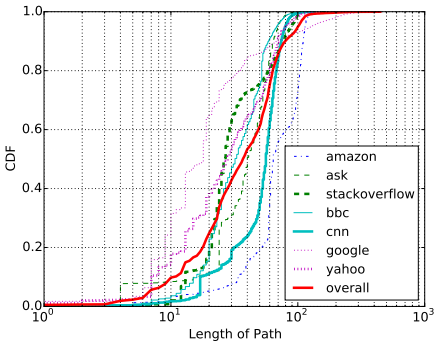

Fig. 3. The length of path.

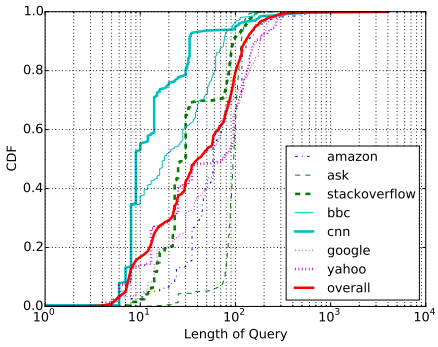

Fig. 4. The length of query.

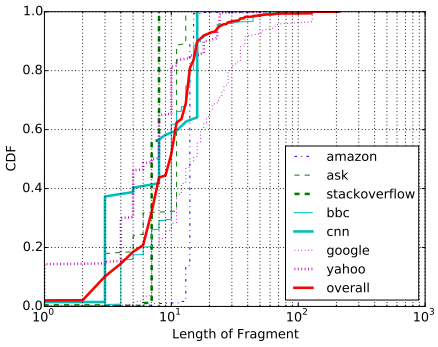

Fig. 5. The length of fragment.

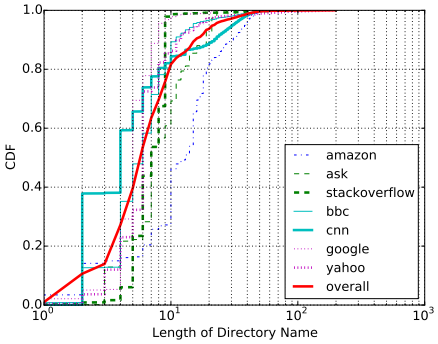

Fig. 6. The length of directory name.

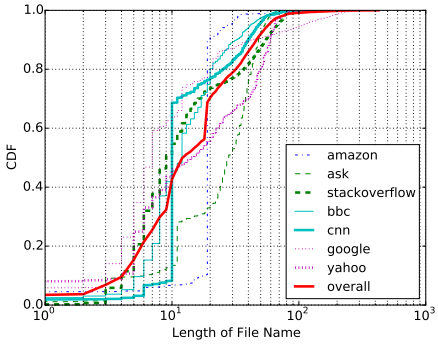

Fig. 7. The length of file name.

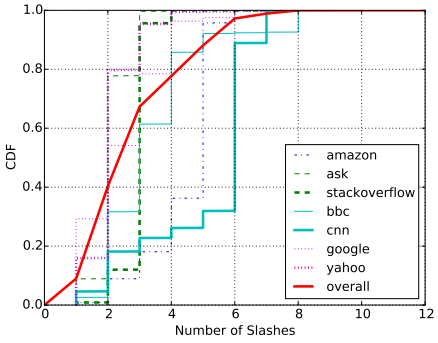

Fig. 8. The number of slashes.

The false positive rate was 52%.

$F1$ is mainly based on the length to detect anomalous names. The longer name increases the possibility that $F1$ is true. But, if the name is not anomalous, the longer it is, and the better similarity it has since it has more characters to compare. Thus we define the third filter as shown in formula (3).

$$F3 = F1 \wedge F2 \tag{3}$$

In this case, the false positive rate was 15%. It means that 15% of names are considered as anomalous to be dropped. But the false positive rate was still high. To solve this problem, we must further investigate better measures with a smaller false positive rate as well as the false negative rate.

C. Discussion

The proposed filters are based on statistics from our URL crawling experiments. Basically, filters can help stating if the name of an Interest packet is legitimate or anomalous. Indeed, in case of Targeted Attacks in an ICN network, if the names created by a malware are detected as anomalous, packets will be dropped and the attack will be prevented. As 15% of name anomaly detection is significant, it is still possible to improve the accuracy of the filters.

Note that a malware can also learn the statistics rules used by our filters (length of path, length of fragment, etc). In fact, to overcome this limitation, each network operators should monitor names in their network and should adapt lively the threshold of the filters. This also requires traffic anomaly detection techniques, where network operators monitor continuously the traffic to detect slight changes and set up alarm if detected.

Even in ICN world, SEO (Search Engine Optimization) business will exist. SEO is an optimization to improve the rank of the web site in search engines such as Google and to move this web site in the upper level of retrieval results by search engines. One of the methods to improve the rank is to modify the structure of URL according to the recommended policies. Moz, which is one of the SEO companies, reports the policies to modify the structure of URL [15]. Specifically, these policies are “to become human-readable”, “to add the keywords”, “to shorten the length”, “to decrease the number of directories”, “to avoid utilizing the hash value”, “to utilize “-” or “_” to separate words”, “to avoid utilizing the keywords repeatedly”, and so on. Likewise, in ICN/NDN the names will also follow some policies in terms of SEO, and they will narrow the degree of freedom to make names. Then, we can utilize these limitations in filtering.

V. Related Work

There are a few research activities about ICN Firewall. D. Goergen et al. [16] focus on filtering Data packets in ICN Firewall, and propose 7 filters. They mainly focus on the situations to pull the contents outside/inside the enterprise network. Basically, there are three approaches for countermeasure; entry

control, inside operation prevention, and exit control [6]. Entry control prevents attacks from getting into the system. Inside operation prevention enhances monitoring and prevents deeper infiltration into the internal system. Exit control enhances monitoring and blocks the suspicious outbound communications. In this paper, we focus on exit control and propose Interest packet filtering.

One of the other examples than NDN in ICN is DONA (Data-Oriented Network Architecture) [17]. In contrast to NDN, the name in DONA has a flat structure. Specifically, the name is constructed by the pair of P and L (P:L). P is the hash value of publisher's public key and L is a unique label assigned by publisher. This name, which includes the hash value of public key, is called Self-Certifying Names. A. Ghodsi et al. [18] discuss a human-readable name and Self-Certifying Names in terms of the relationship between Real-World Identity (RWI), name, and public key. RWI refers to a person or organization. Considering three parts; security, scalability, and flexibility, they argue the efficacy of Self-Certifying Names. Moreover, as one method not to use public key needed for data authentication, M. Baugher et al. [19] propose Self-Verifying Names, which utilize a hash value in names got from Catalog.

According to these naming policies, there are three kinds of content names; human-readable name, non-human-readable name, and the combination of human-readable and non-human-readable name. However, considering information leakages through Interest packets, which we mention in Section II-C, we cannot distinguish whether encrypted non-human-readable name is the real content name or the name created by a malware. When Self-Verifying Names are exploited to realize information leakages, the volume of the leaked data is low because in the names the length of hash value is limited. The name in DONA has also limitation about the length of the hash value. According to the packet format in Ref. [17], the length of P:L is 40 Byte. However, the name in NDN has variable length. Therefore, an attacker can increase the volume of the leaked data easily. NDN has a higher risk about information leakages through the names.

VI. Conclusion

Information leakages through Interest packets will be one of the main security threats in ICN. It is therefore mandatory to propose countermeasures. In this context, we design Interest packet filtering to detect anomalous names in ICN and prevent information leakage. Indeed, malicious names are more likely to be created by malwares to leak information of the network through Targeted Attack. Our name anomaly detection filters are based on statistics from URLs that we obtained through extensive crawling experiments of main Internet organizations. Our filter is currently immature and recognizes 15% of names in our dataset as malicious since they are not compliant with the statistical properties of the collected URLs. This false positive rate is still too high to use this filter in a real field of operation. However, this work is an important step for detecting anomalous names and preventing information-leakage in ICN.

The actual performance of our filter must be evaluated against real attacks which encode leaked information into their interest names. Invention of such methods is one of our future works. We will take into account SEO rules followed by most of the top websites to improve the accuracy of our filters.

Acknowledgment

This work is supported by DOCTOR Project, funded by French National Research Agency (ANR-14-CE28-0001), and the GreenICN project (GreenICN: Architecture and Applications of Green Information Centric Networking), a research project supported jointly by the European Commission under its 7th Framework Program (contract no. 608518) and the National Institute of Information and Communications Technology (NICT) in Japan (contract no. 167).

References

[1] 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, (Accessed 8/3/2016).
[2] Understanding Targeted Attacks: The Impact of Targeted Attacks, http://www.trendmicro.com/vinfo/us/security/news/cyber-attacks/the-impact-of-targeted-attacks, (Accessed 8/3/2016).
[3] 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.
[4] M. Ion, J. Zhang, and E. M. Schooler, "Toward Content-Centric Privacy in ICN: Attribute-based Encryption and Routing," in *Proc. the ACM SIGCOMM 2013 Conference on SIGCOMM*, pp. 513–514, Aug. 2013.
[5] D. Naylor, K. Schomp, M. Varvello, I. Leontiadis, J. Blackburn, D. R. López, K. Papagiannaki, P. R. Rodriguez, and P. Steenkiste, "Multi-Context TLS (mcTLS): Enabling Secure In-Network Functionality in TLS," in *Proc. the 2015 ACM Conference on Special Interest Group on Data Communication*, pp. 199–212, Aug. 2015.
[6] System Design Guide for Thwarting Targeted Email Attacks, http://www.ipa.go.jp/files/000035723.pdf, (Accessed 8/3/2016).
[7] RFC 1808, https://tools.ietf.org/html/rfc1808, (Accessed 8/3/2016).
[8] We knew the web was big... (Google Official Blog), https://googleblog.blogspot.jp/2008/07/we-knew-web-was-big.html, (Accessed 8/3/2016).
[9] CCNx 1.0 Protocol Architecture, http://www.ccnx.org/pubs/CCNxProtocolArchitecture.pdf, (Accessed 8/3/2016).
[10] NDN Packet Format Specification, http://named-data.net/doc/ndn-tlv/name.html, (Accessed 8/3/2016).
[11] RFC 3986, https://tools.ietf.org/html/rfc3986, (Accessed 8/3/2016).
[12] Alexa Top 500 Global Sites, http://www.alexa.com/topsites, (Accessed 8/3/2016).
[13] B. Liu, "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," *Springer-Verlag Berlin Heidelberg*, 2011.
[14] Frequency analysis, https://en.wikipedia.org/wiki/Frequency_analysis, (Accessed 8/3/2016).
[15] 15 SEO Best Practices for Structuring URLs, https://moz.com/blog/15-seo-best-practices-for-structuring-urls, (Accessed 8/3/2016).
[16] 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.
[17] T. Koponen, M. Chawla, B. Chun, A. Ermolinskiy, K. H. Kim, S. Shenker, and I. Stoica, "A Data-Oriented (and Beyond) Network Architecture," in *Proc. the 2007 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, pp. 181–192, Aug. 2007.
[18] A. Ghodsi, T. Koponen, J. Rajahalme, P. Sarolahti, and S. Shenker, "Naming in Content-oriented Architectures," in *Proc. the ACM SIGCOMM Workshop on Information-Centric Networking (ICN'11)*, pp. 1–6, Aug. 2011.
[19] M. Baugher, B. Davie, A. Narayanan, and D. Oran, "Self-Verifying Names for Read-Only Named Data," in *Proc. 2012 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 274–279, Mar. 2012.