

インターネットから収集した URLデータの統計量に基づく NDN情報漏洩防止方法の検討

近藤 大嗣†, †† シルバーストン トーマス†††
戸出 英樹†††† 浅見 徹†††††

† University of Lorraine, LORIA (CNRS UMR 7503)
†† Inria Nancy – Grand Est
††† The University of Tokyo, JFLI (CNRS UMI 3527)
†††† 大阪府立大学大学院工学研究科
††††† 東京大学大学院情報理工学系研究科

19/5/16

ICN Technical Committee Meeting

Outline

- Problem statement
 - Threat of targeted attack in the Internet
 - Future targeted attack in ICN
- Objective and proposal
 - Objective: Prevent information leakage through Interest packet
 - Proposal: Interest packet filtering
- ICN content naming policy as a natural extension from URL
- ICN content naming analysis based on URL
- Evaluation of Interest packet filtering

19/5/16

ICN Technical Committee Meeting

Threat of Targeted Attack

- IPA reports that targeted attack is one of the 10 major security threats [1] [1] 10 Major Security Threats 2015, <https://www.ipa.go.jp/files/000048018.pdf>
 - 2014: 3rd, 2015: 1st
 - In targeted attack,
 - Attacker infects PCs within target network with malware via email, etc.
 - Attacker probes internal networks and steals data such as customer information
 - Countermeasure
 - Train employees not to access suspicious media
- Almost impossible to extinguish human error!!**

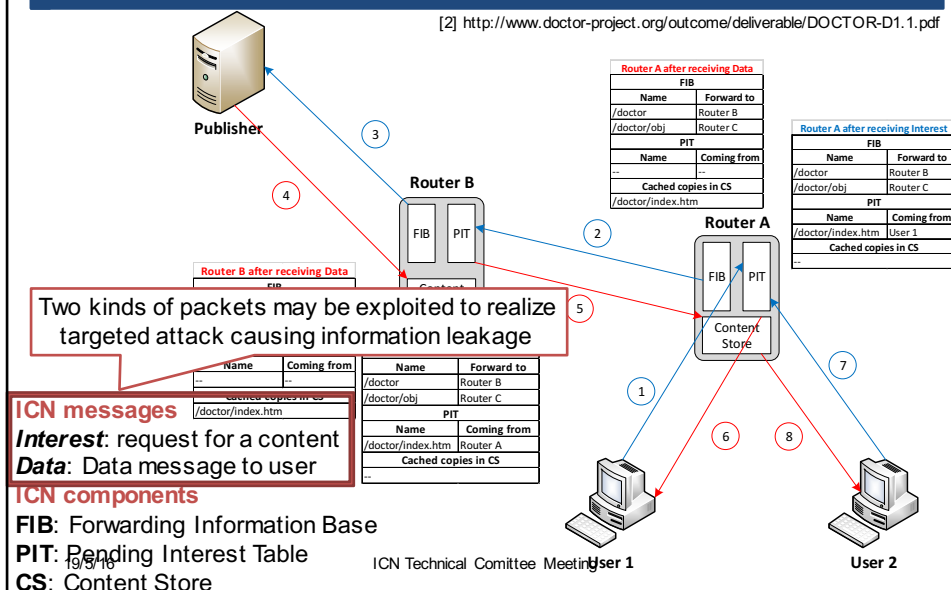
Important to investigate how to prevent information leakages after infection

19/5/16

3

Overview of ICN

[2] <http://www.doctor-project.org/outcome/deliverable/DOCTOR-D1.1.pdf>




4

5

Information Leakage through Data Packet

- Data packet includes
 - Data, content name, etc.
- Characteristic of Data packet
 - Unless user receives Interest packet, he cannot send Data packet corresponding to the Interest packet

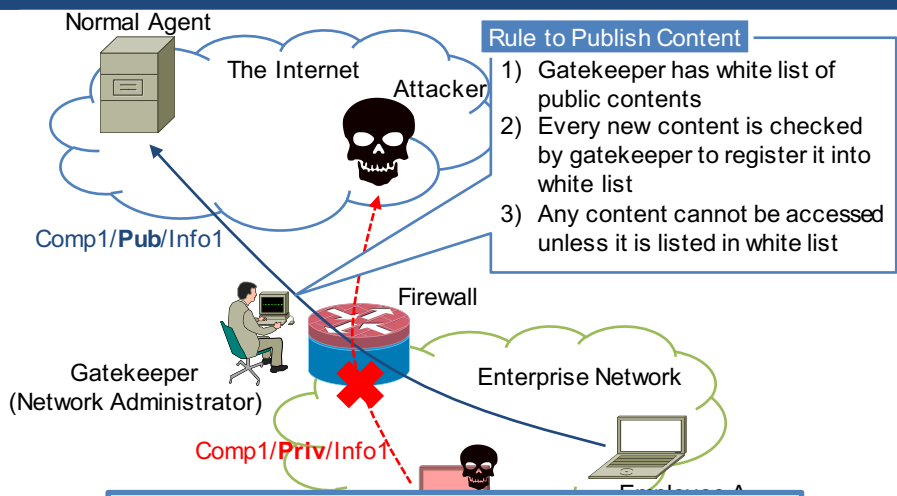


To steal information, attacker must send Interest packet to target network

19/5/16 ICN Technical Committee Meeting

6

Prevention for Information Leakage through Data Packet



Rule to Publish Content

- 1) Gatekeeper has white list of public contents
- 2) Every new content is checked by gatekeeper to register it into white list
- 3) Any content cannot be accessed unless it is listed in white list

Gatekeeper can prevent information leakage through Data packet with name

19/5/16

Information Leakage through Interest Packet

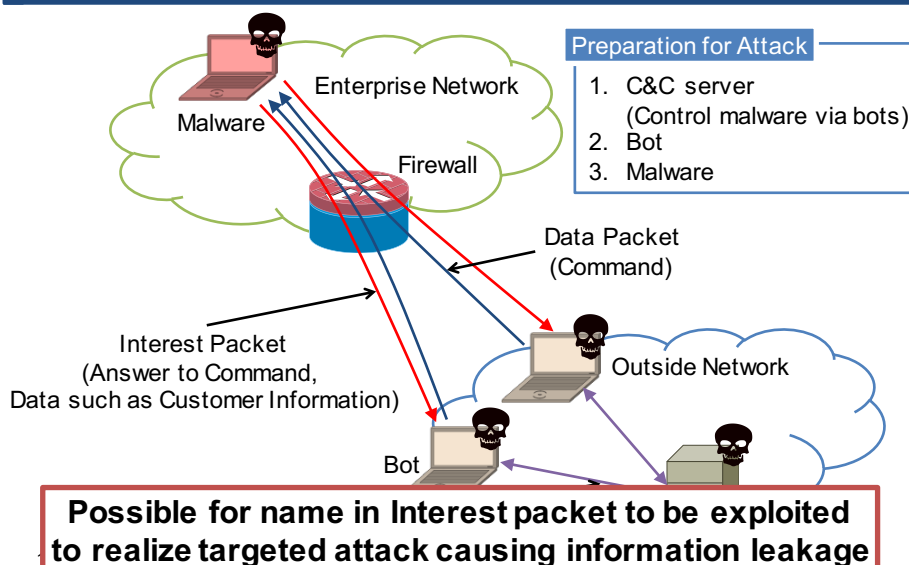
- Interest packet includes
 - Content name, etc.
- Characteristic of Interest packet
 - User sends out content name, which is named by outside publisher

Possible for malware to encrypt information, specify it into name, and send it outside as if malware requested outside content

19/5/16

ICN Technical Committee Meeting

Future Targeted Attack in ICN



9

Objective and Proposal

Objective

Prevent information leakage through Interest packet

Proposal

Interest packet filtering to detect malicious names causing information leakages

➔ **Necessity of realizing what is normal name**

Step to define normal name

- 1) Define ICN content naming policy
 - ICN content naming policy as a natural extension from URL
- 2) Assume statistic of ICN content name
 - Analysis of URL obtained from the Internet

↓

Interest packet filtering based on statistic of normal name

10

ICN Content Naming Policy as a Natural Extension from URL

- In RFC 1808, URL is defined as
`<scheme>://<net_loc>/<path>;<params>?<query>#<fragment>`
- Future ICN naming policy may specify ccn into <scheme> part in stead of http(s)
- Why natural extension from URL?
 - Organization described in <net_loc> part can define name from <path> part to <fragment> part, independent from each other
 - Easy to translate current numerous content names in the Internet to corresponding ICN names

19/5/16 ICN Technical Committee Meeting

11

URL Dataset

- Sampled URLs created by 7 organizations (Amazon, Ask, stackoverflow, BBC, CNN, Google, Yahoo)
- Obtained 30,000 from each organization, and then collected URLs returning status code 200
- Divided URLs into 9-to-1
 - 90% were used for statistical analysis
 - 10% were used for evaluation of Interest packet filtering
- Calculated average frequencies of characters in path, query and fragment of the URLs in all the organizations
 - *Average frequencies in path*
 - *Average frequencies in query*
 - *Average frequencies in fragment*

19/5/16 ICN Technical Committee Meeting

12

Average Frequencies in Path, Query, and Fragment

- 3 histograms were obtained from URL datasets of 7 organizations

Average Frequencies in path

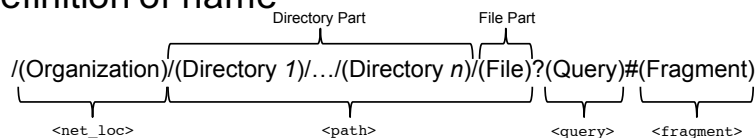
Average Frequencies in query

Average Frequencies in fragment

19/5/16 ICN Technical Committee Meeting

9 Statistics Obtained from URL

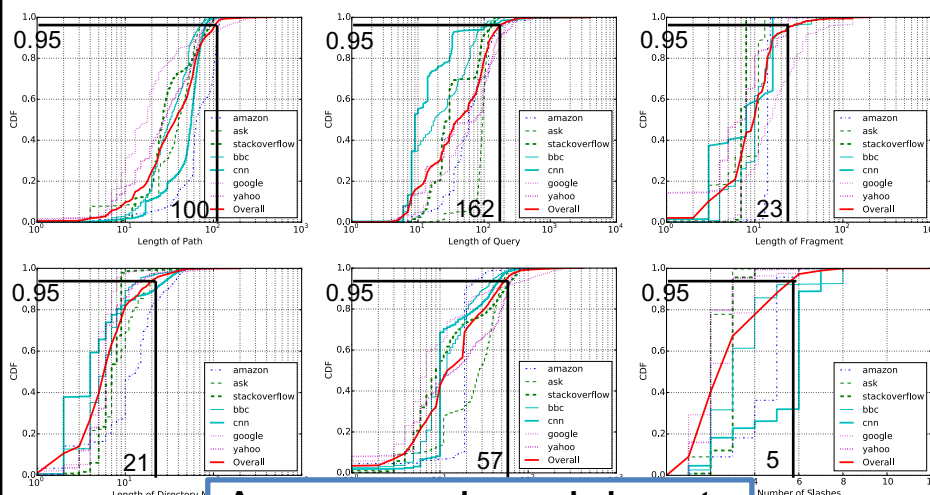
- Definition of name



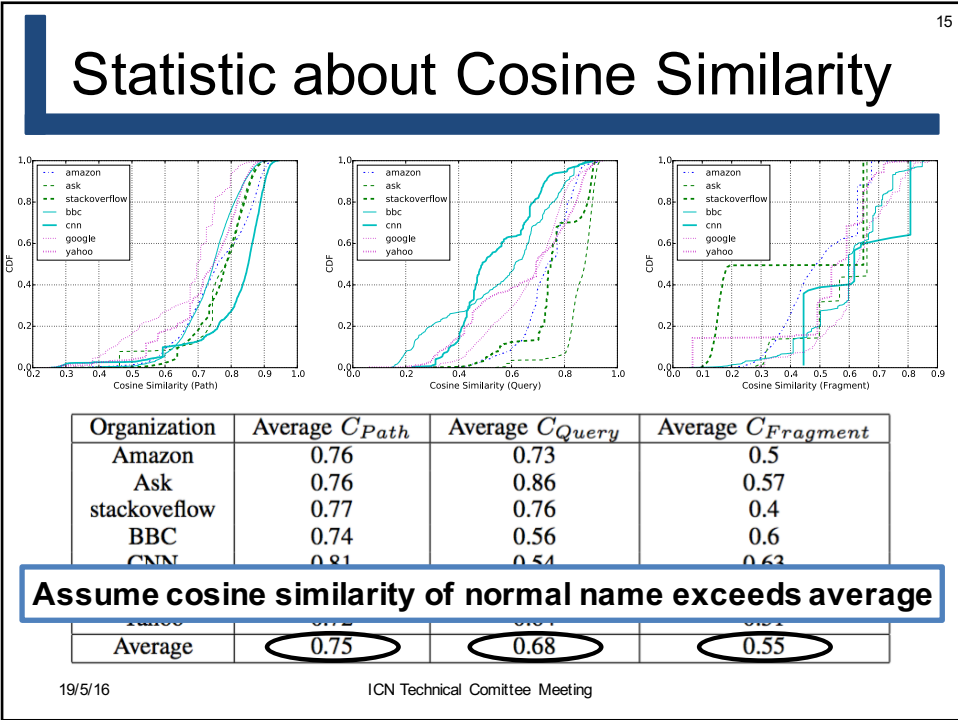
- 9 statistics

Statistic	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	$C_{Fragment}$

Statistics about Length and Number



Assume normal name belongs to the 95th percentile in each variable



Evaluation of Interest Packet Filtering

- Filter 1 (F1) from point of length and number**

$$F1 = (L_{Path} \geq 101) \vee (L_{Query} \geq 163) \vee (L_{Fragment} \geq 24)$$

$$\vee (\min_{in\ path} L_{DirectoryName} \geq 22) \vee (L_{FileName} \geq 58)$$

$$\vee (N_{Slashes} \geq 6)$$

False positive rate: 33% (1)
- Filter 2 (F2) from point of cosine similarity**

$$F2 = F1 \wedge ((C_{Path} < 0.75) \vee (C_{Query} < 0.68)$$

$$\vee (C_{Fragment} < 0.55))$$

False positive rate: 15%

Reduce false positive rate because filtering can judge character frequencies in name are similar to one in normal name even if length is over threshold

Discussion

- Proposed filtering rules are not enough to prevent information leakages
 - E.g., **Malware can create name for information leakages to avoid filtering, which decreases volume of leaked data**
- Conventional naming policies have risk about information leakages
 - Perhaps not only improvement of filtering but also new naming policy or other countermeasures are needed

19/5/16

ICN Technical Committee Meeting

Conclusion and Future Work

- Conclusion
 - Name in Interest packet will be exploited to realize future targeted attack
 - Considering migration path from current Internet to ICN, highly possible for future naming policy to become one as natural extension from URL
 - False positive rate of proposed Interest packet filtering was 15%, but too high as practical filter
- Future work
 - Propose method to reduce false positive rate in Interest packet filtering
 - Consider encryption method for malware to hide information into name and evaluate performance of each method

19/5/16

ICN Technical Committee Meeting