Name-Anomaly Detection in ICN

Information-leakage in NDN

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

[1]Université de Lorraine, LORIA (CNRS UMR 7503)

[2]Inria Nancy – Grand Est

[3]National Institute for information and Communication Technology

[4]Graduate School of Engineering, Osaka Prefecture University

[5]Graduate School of Information Science & Technology, University of Tokyo

3rd FRA-JPN meeting, April 24-26 2017, Tokyo

Information-leakage

- One of the main security threat in Internet
  - *IT Security Risks Survey 2014: A Business Approach to Managing* http://media.kaspersky.com/en/IT_Security_Risks_Survey_2014_Global_report.pdf
- Cyber Espionage
  - Targeted Attacks (malware, website, external memory device)
- Examples: Sony, Target
  - $100 M upgrading systems
  - 46% drop in benefits [*Understanding Targeted Attacks: The Impact of Targeted Attacks*]

Targeted Attacks

- Infects PC via emails
- Probes network
- Steals Information

Countermeasures

Train employees?

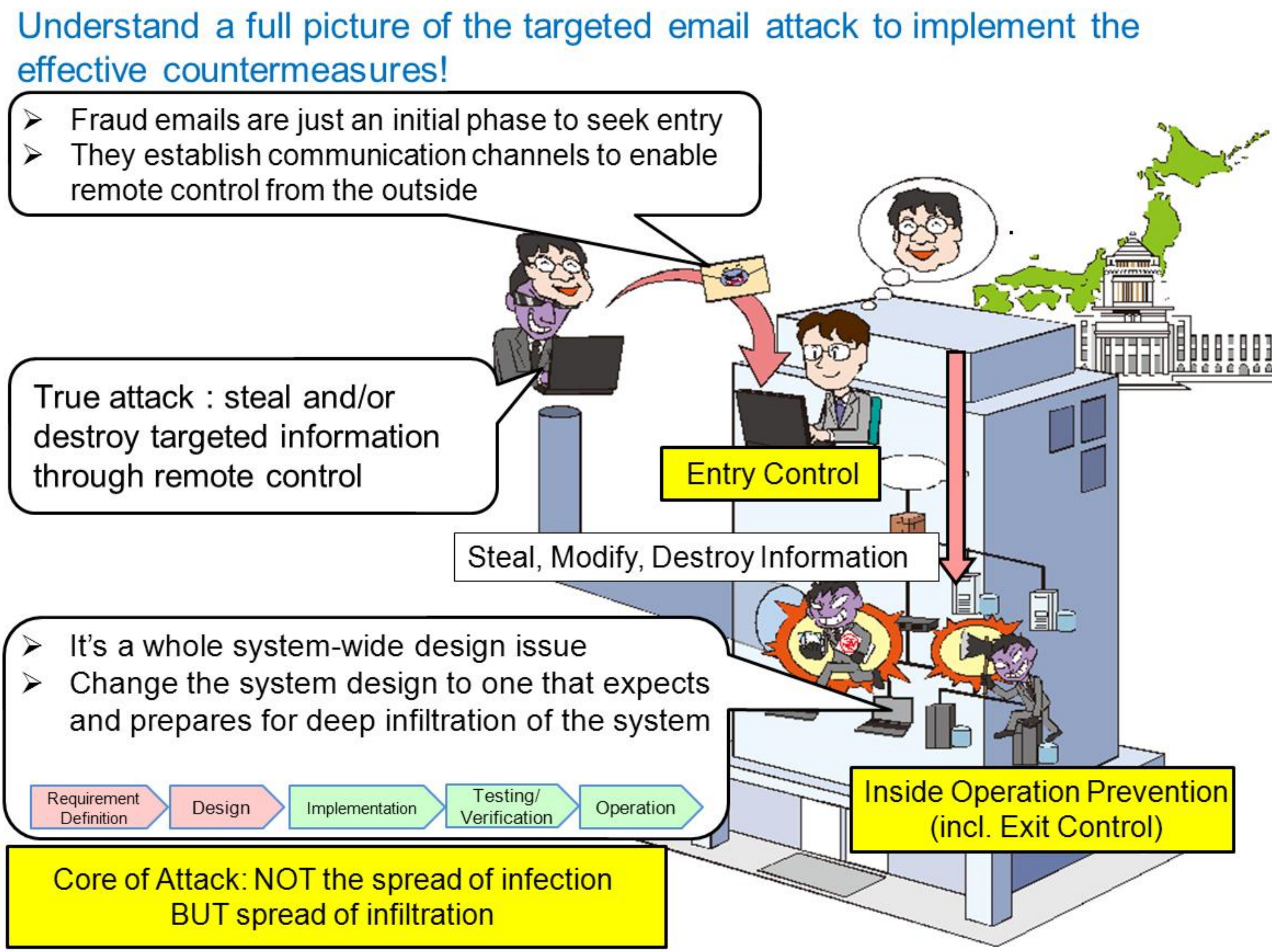
Human errors

Source:

IT Security Center

IPA: IT Promotion Agency

http://www.ipa.go.jp/security/english/newattack_en.html



Understand a full picture of the targeted email attack to implement the effective countermeasures!

- ➢ Fraud emails are just an initial phase to seek entry
- ➢ They establish communication channels to enable remote control from the outside

True attack : steal and/or destroy targeted information through remote control

Entry Control

Steal, Modify, Destroy Information

- ➢ It's a whole system-wide design issue
- ➢ Change the system design to one that expects and prepares for deep infiltration of the system

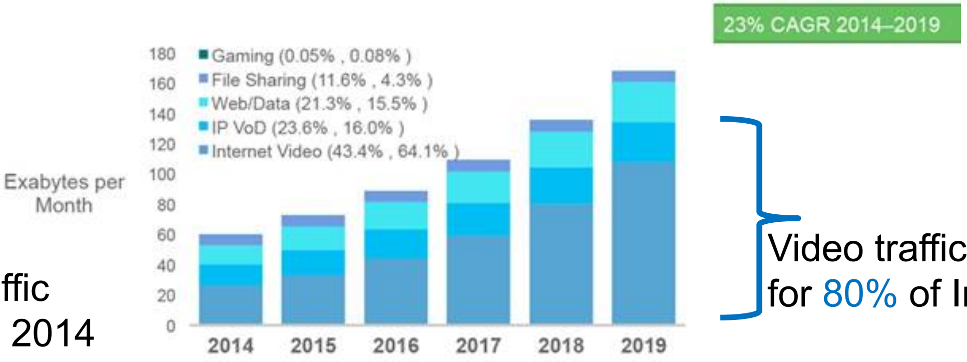
Requirement Definition → Design → Implementation → Testing/ Verification → Operation

Inside Operation Prevention (incl. Exit Control)

Core of Attack: NOT the spread of infection BUT spread of infiltration

Information-Centric Networking

- Internet is mostly used to access content
  - Video: 80% of global consumer traffic by 2019
    - [Cisco VNI 2015]
  - TCP/IP: *host-to-host* communication paradigm
- Users are interested with content not location
- Information-Centric Networking
  - ***Named-Data Networking*** (NDN) [CoNext 2009]
  - *Host-to-content* communication
    - Packet address *refers* to **content name** and not location (*host*)
- New « Network layer » for Future Internet
  - Data at the *core* of the communication



67% of Internet traffic was video traffic in 2014

Video traffic will account for 80% of Internet traffic

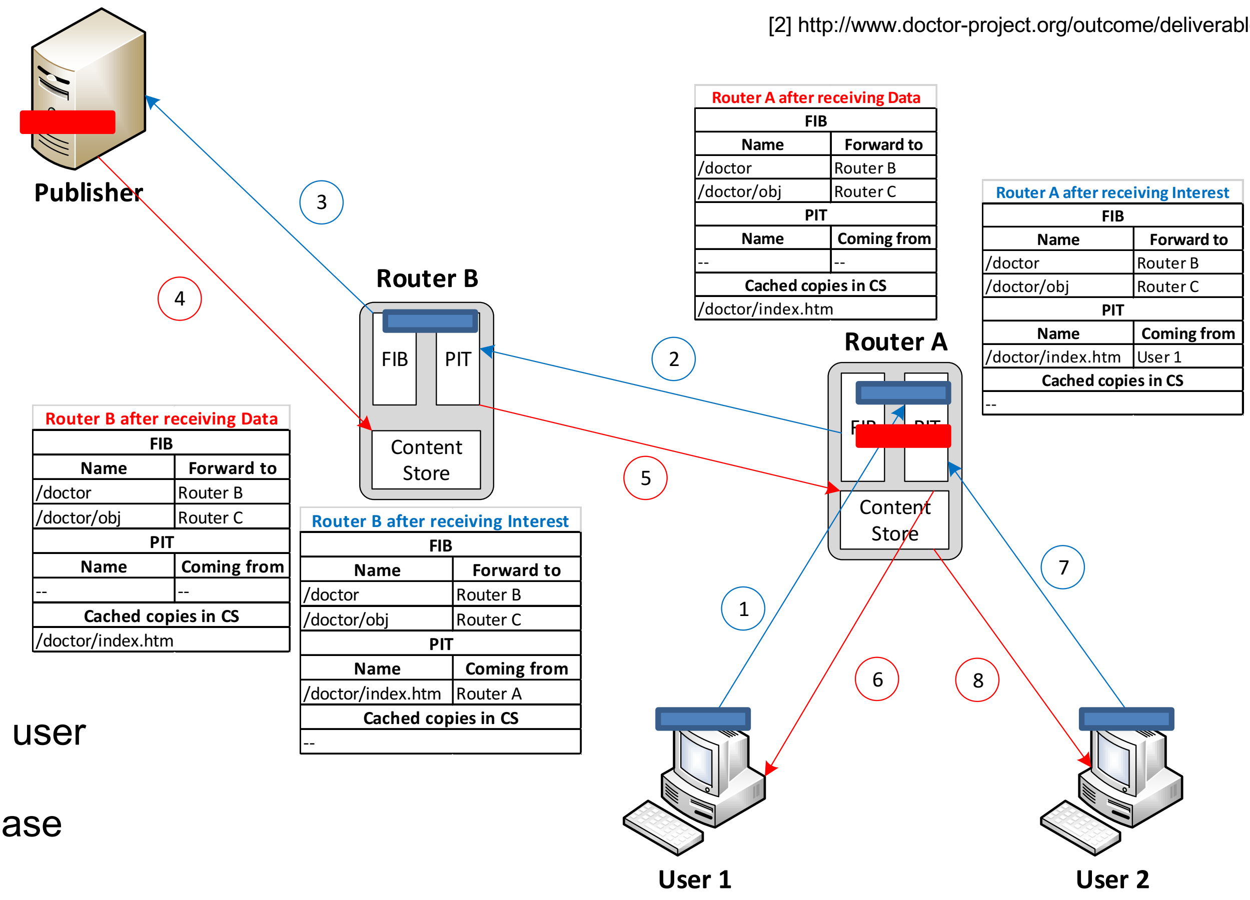
* Figures (n) refer to 2014 and 2019 traffic shares

NDN Overview

- Packet address *refers* to ***content name*** not location
  - Named-Data Networking
- Two primitives
  - *Interest,* user requests content by issuing an Interest message
  - *Data,* a node having the content answer with a Data message
- *In-Network* Caching
- Data at the *core* of the communication
- New *'Network Layer'* for Content Delivery

Overview of Named-Data Networking (NDN)

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



Router A after receiving Data	
FIB	
Name	Forward to
/doctor	Router B
/doctor/obj	Router C
PIT	
Name	Coming from
--	--
Cached copies in CS	
/doctor/index.htm	

Router A after receiving Interest	
FIB	
Name	Forward to
/doctor	Router B
/doctor/obj	Router C
PIT	
Name	Coming from
/doctor/index.htm	User 1
Cached copies in CS	
--	

Router B after receiving Data	
FIB	
Name	Forward to
/doctor	Router B
/doctor/obj	Router C
PIT	
Name	Coming from
--	--
Cached copies in CS	
/doctor/index.htm	

Router B after receiving Interest	
FIB	
Name	Forward to
/doctor	Router B
/doctor/obj	Router C
PIT	
Name	Coming from
/doctor/index.htm	Router A
Cached copies in CS	
--	

NDN/CCN packet

Interest: Request for content

Data/Content Object: Data to user

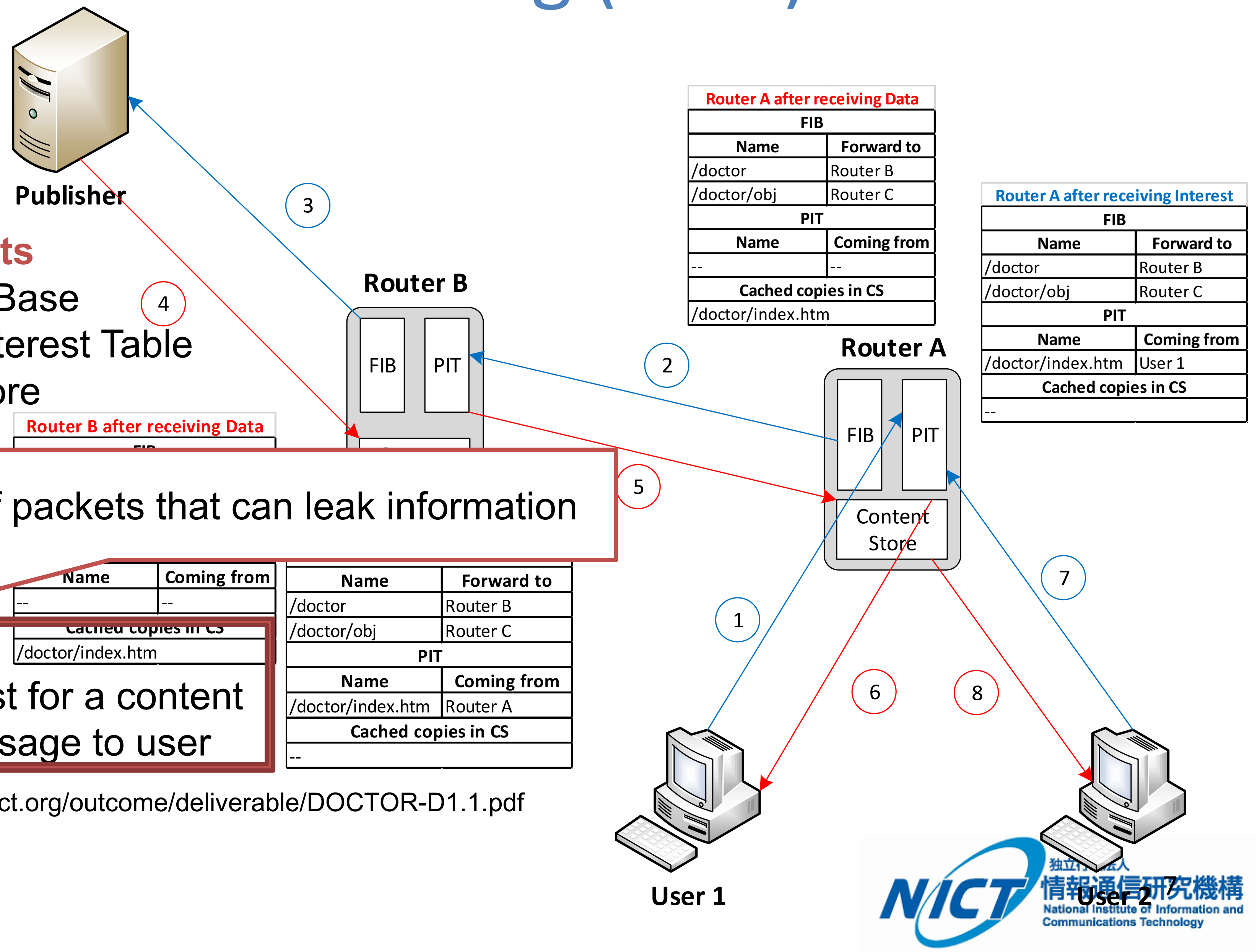
NDN/CCN component

FIB: Forwarding Information Base

PIT: Pending Interest Table

CS: Content Store

Overview of Named-Data Networking (NDN)



ICN components

FIB: Fwd. Info. Base

PIT: Pending Interest Table

CS: Content Store

Router A after receiving Data

FIB	
Name	**Forward to**
/doctor	Router B
/doctor/obj	Router C
PIT	
Name	**Coming from**
--	--
Cached copies in CS	
/doctor/index.htm	

Router A after receiving Interest

FIB	
Name	**Forward to**
/doctor	Router B
/doctor/obj	Router C
PIT	
Name	**Coming from**
/doctor/index.htm	User 1
Cached copies in CS	
--	

Router B after receiving Data

Name	**Coming from**
--	--
Cached copies in CS	
/doctor/index.htm	

Name	**Forward to**
/doctor	Router B
/doctor/obj	Router C
PIT	
Name	**Coming from**
/doctor/index.htm	Router A
Cached copies in CS	
--	

Two kinds of packets that can leak information

ICN messages

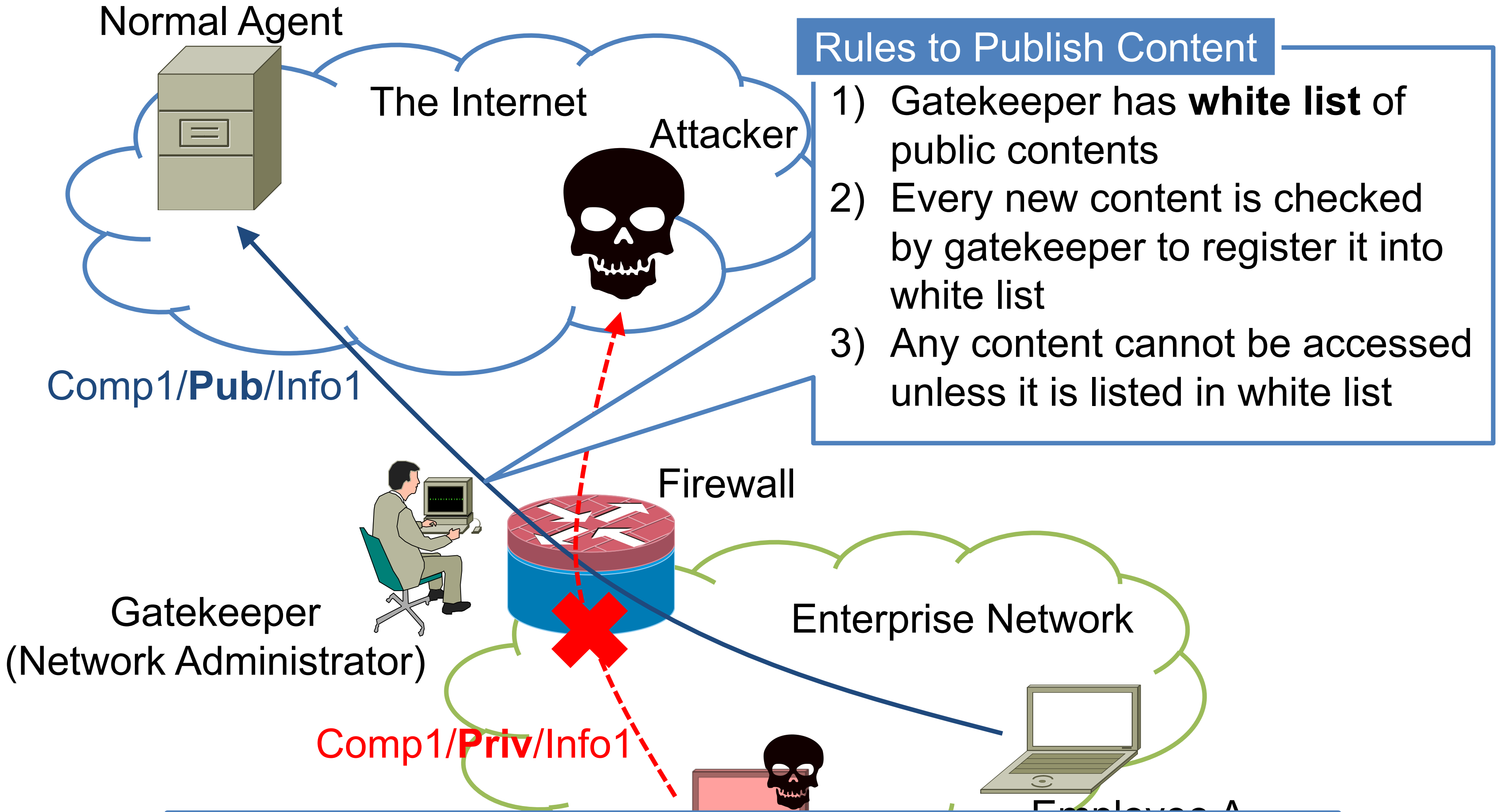
Interest: request for a content

Data: Data message to user

http://www.doctor-project.org/outcome/deliverable/DOCTOR-D1.1.pdf

独立行政法人 情報通信研究機構 National Institute of Information and Communications Technology

Information-leakage with *Data Packets*



Rules 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

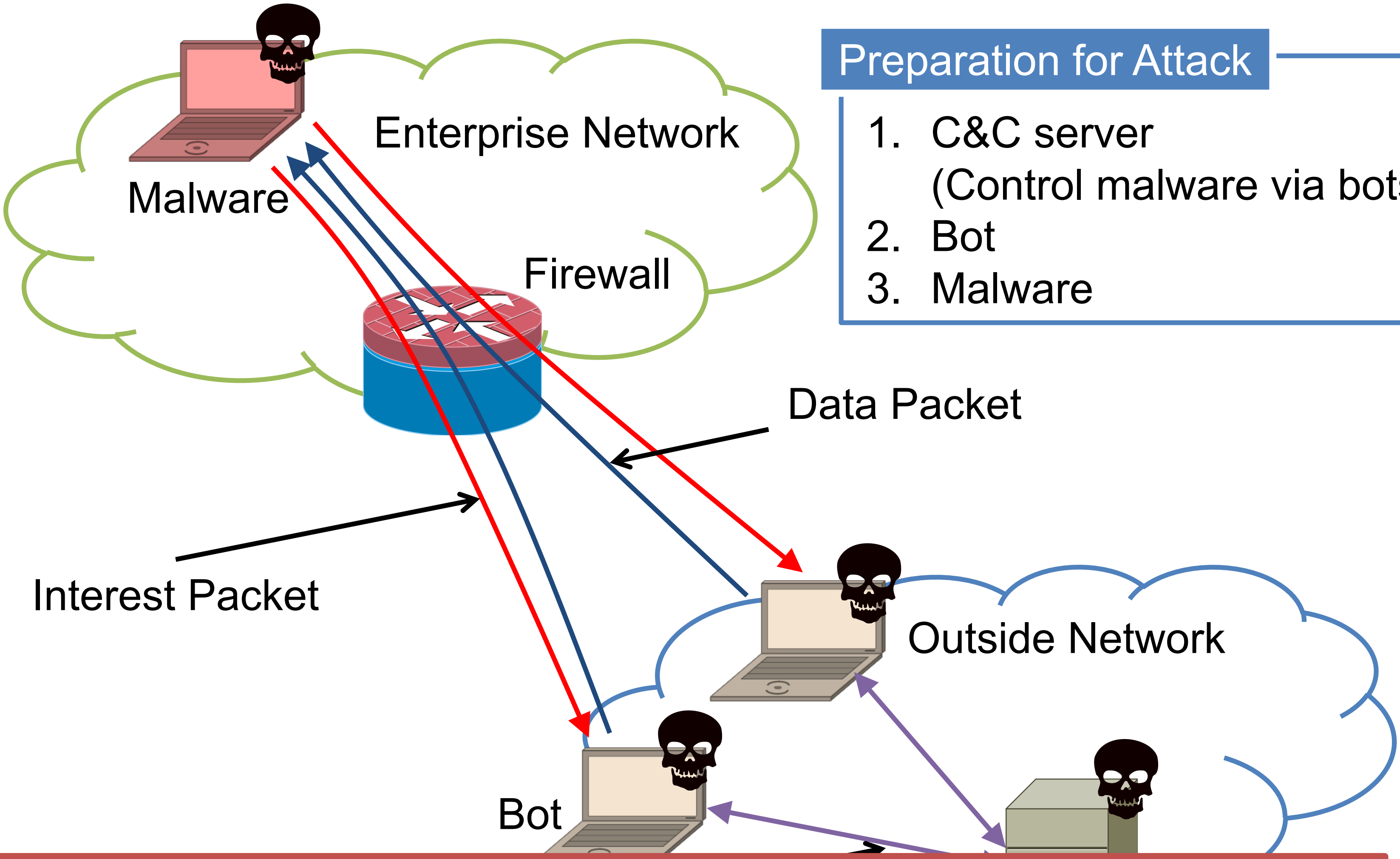
Gatekeeper can prevent information leakage through Data packet (reply messages)

Information-Leakage through Data Packet

- Data packet includes
  - Data, content name, etc.
- Characteristic of Data packet
  - Data packet cannot be sent if not a reply from Interest packet

Only Interest packets can leak information from network

Information-leakage with Interest



Preparation for Attack

1. C&C server
   (Control malware via bots)
2. Bot
3. Malware

Interest Name can be used to leak information through Targeted Attacks (request messages)

Summary : Information-leakage through NDN packets

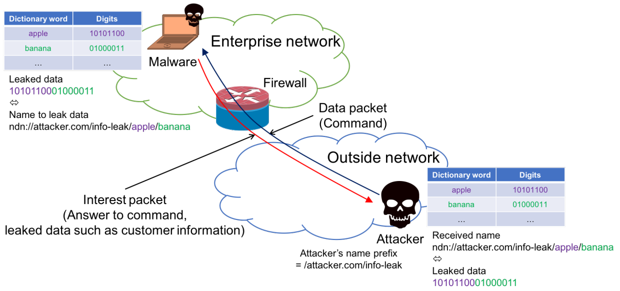
- *Interest/Data* packets are “Request/Reply”
  - Content name, etc.
- *Data* packets can be **filtered out** out by admin.
  - White/Black lists of (un)authorized content names
    - *CustomerList*, *BankingInfo*, etc.
- Interest packets are sent out the network to external publishers as requests (“free” names)
  - Malwares can use *Interest* to leak Information through Targeted Attacks (steganography-embedded)

Risk Analysis of Information-Leakage through Interest Packets in NDN

- Performing information-leakage with names in NDN Interest packets
- Prevent information-leakage in NDN (*Interest*)
  - Major threat in the Internet
  - Named-Data Networking: architecture for Future Internet
- Proposal
  - Interest (Packet) filtering based on anomalous names
    - firewall
- Methodology
  - Study Names in the Internet with URLs
- Assumption
  - NDN Names will be based on URLs
    - Easy to translate current URL Names into NDN names

Attack Model and Countermeasure

- Attack model
  - Malware builds anomalous names to leak information
    - steganography-embedded



Dictionary word	Digits
apple	10101100
banana	01000011
...	...

Leaked data
1010110001000011
⇔
Name to leak data
ndn://attacker.com/info-leak/apple/banana

Dictionary word	Digits
apple	10101100
banana	01000011
...	...

Received name
ndn://attacker.com/info-leak/apple/banana
⇔
Leaked data
1010110001000011

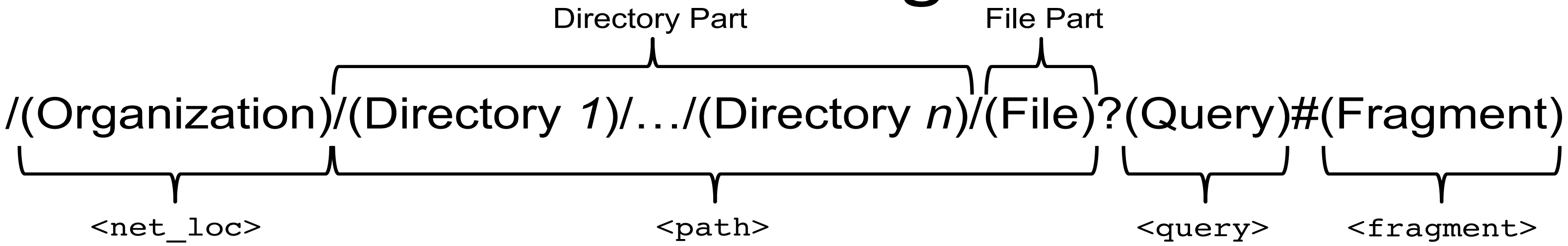
- Countermeasures

  1. Name-based filters using Name statistics
  2. Name-based filter using one-class SVM

- Assumption
  - NDN names will be extension of URLs in the current Internet

URLs Dataset

- Web Crawling of 7 main organizations
  - Amazon, Ask, Stackoverflow, BBC, CNN, Google, Yahoo
- 1 million URLs for each organization



Directory Part

File Part

/(Organization)/(Directory *1*)/…/(Directory *n*)/(File)?(Query)#(Fragment)

<net_loc>

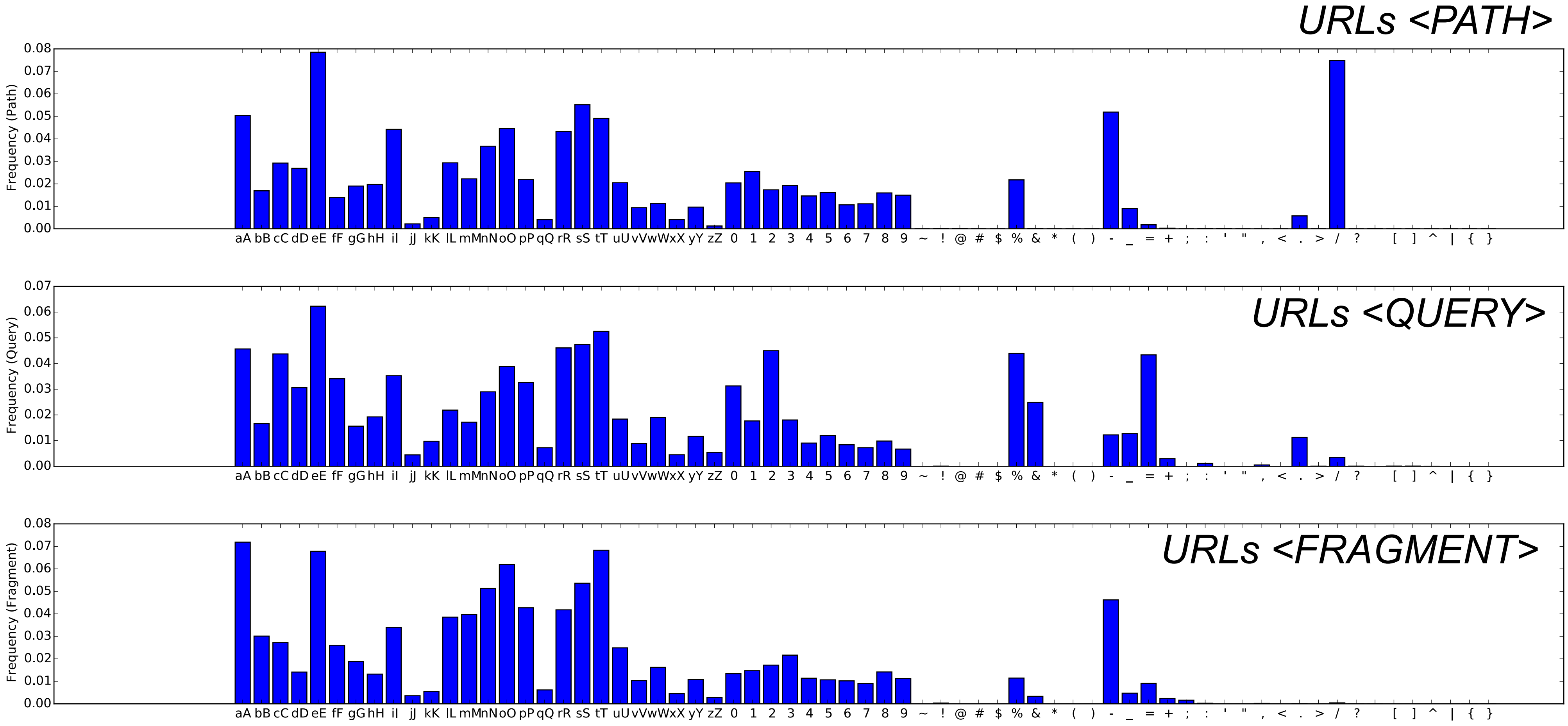
<path>

<query>

<fragment>

URLs Parameters (RFC 1808)	
Length of <PATH>	Number of '/' in <path>
Length of <QUERY>	Similarity of characters in <PATH>
Length of <FRAGMENT>	Similarity of characters in <QUERY>
Length of Directory	Similarity of characters in <FRAGMENT>
Length of File	

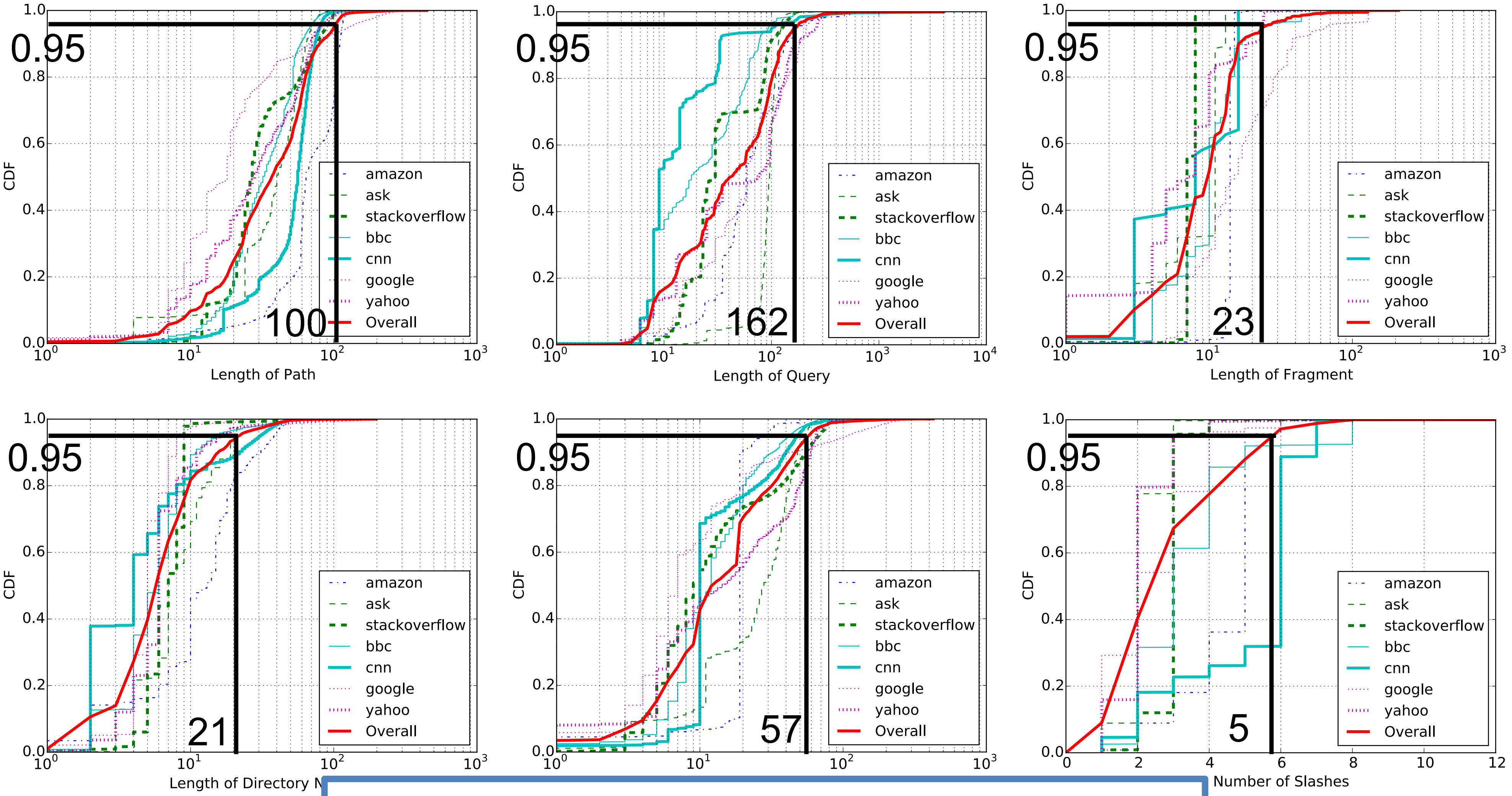
Character Frequencies in URLs



19/5/16

独立行政法人
情報通信研究機構
National Institute of Information and Communications Technology

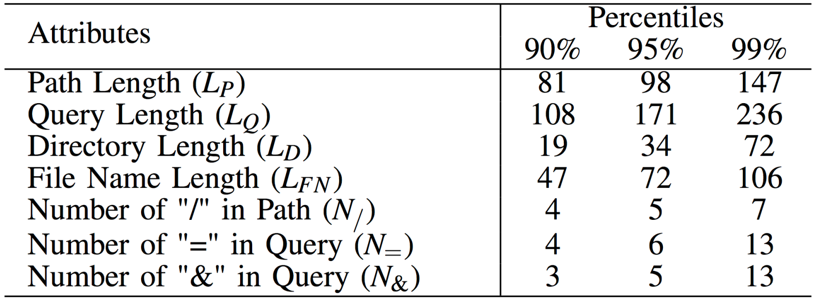
URLs Statistics



Legitimate names: 95th percentile

URLs Statistics

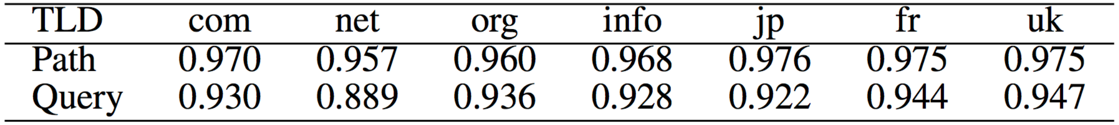
- URL attributes and computed percentiles



Attributes	Percentiles		
	90%	95%	99%
Path Length (L_P)	81	98	147
Query Length (L_Q)	108	171	236
Directory Length (L_D)	19	34	72
File Name Length (L_{FN})	47	72	106
Number of "/" in Path ($N_/$)	4	5	7
Number of "=" in Query ($N_=$)	4	6	13
Number of "&" in Query ($N_\&$)	3	5	13

- similarity of averaged frequencies of alphabets in Path and Query compared to typical English text [6]

[6] Frequency analysis, https://en.wikipedia.org/wiki/Frequency_analysis

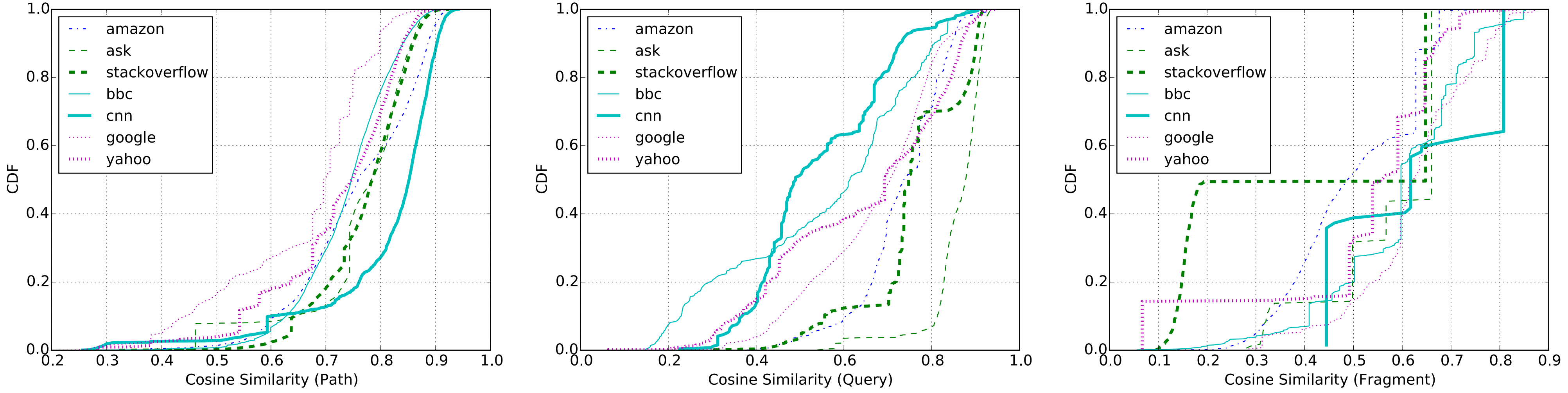


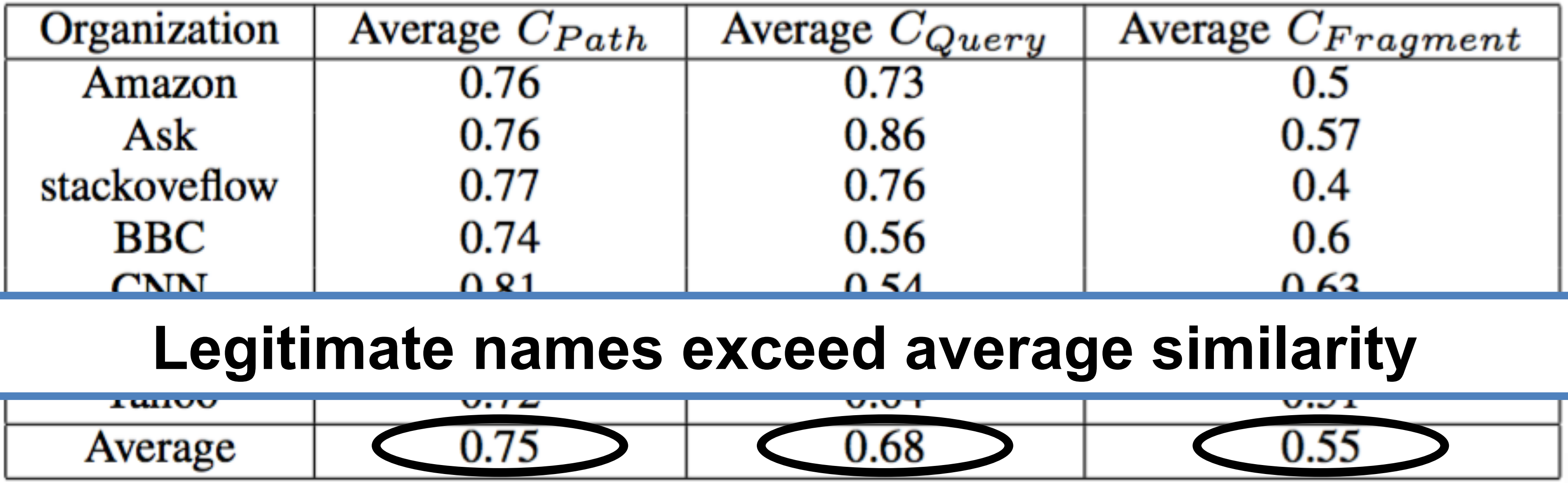
TLD	com	net	org	info	jp	fr	uk
Path	0.970	0.957	0.960	0.968	0.976	0.975	0.975
Query	0.930	0.889	0.936	0.928	0.922	0.944	0.947

➢ High similarity with typical English text => Using WordNet [7] for *steganography*

[7] G. A. Miller, "WordNet: A Lexical Database for English," *Commun. ACM*, vol. 38, no. 11, pp. 39–41, Nov. 1995.

URLs Similarity





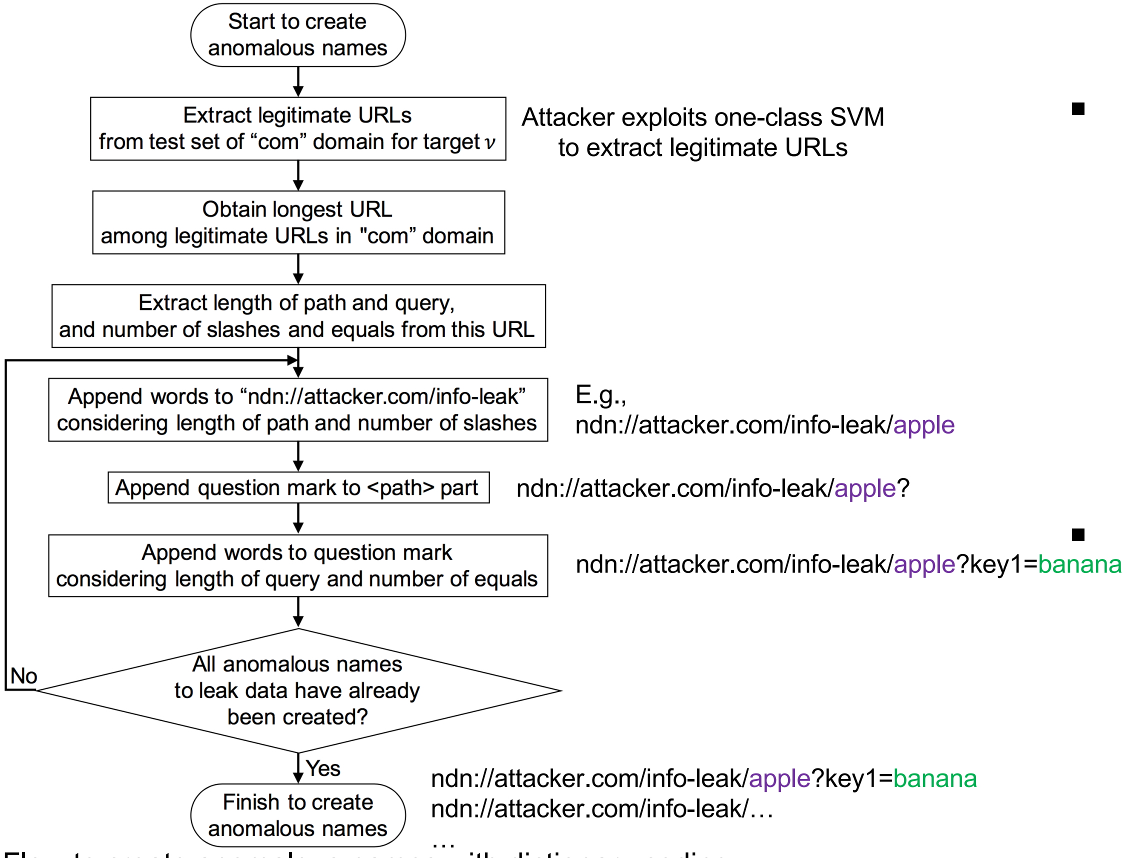
Organization	Average C_{Path}	Average C_{Query}	Average $C_{Fragment}$
Amazon	0.76	0.73	0.5
Ask	0.76	0.86	0.57
stackoveflow	0.77	0.76	0.4
BBC	0.74	0.56	0.6
CNN	0.81	0.54	0.63
Yahoo	0.72	0.64	0.51
Average	0.75	0.68	0.55

Legitimate names exceed average similarity

Names Filtering Heuristics

- Filter based on measured URL parameters
  - Length (Path , Query, Fragment, Direction, File), #/
    - 95th percentile
  - 33% anomalous URLs (67% are legitimate names)
- Filter with Similarity measure
  - Previous extended filter
    - Character frequencies w.r.t. average frequencies in URLs dataset (Path, Query, Fragment)
  - 15% anomalous URLs (85% legitimate names)

Attacker



Start to create anomalous names

Extract legitimate URLs from test set of “com” domain for target ν

Attacker exploits one-class SVM to extract legitimate URLs

Obtain longest URL among legitimate URLs in "com" domain

Extract length of path and query, and number of slashes and equals from this URL

Append words to “ndn://attacker.com/info-leak” considering length of path and number of slashes

E.g., ndn://attacker.com/info-leak/apple

Append question mark to <path> part

ndn://attacker.com/info-leak/apple?

Append words to question mark considering length of query and number of equals

ndn://attacker.com/info-leak/apple?key1=banana

All anomalous names to leak data have already been created?

No

Yes

Finish to create anomalous names

ndn://attacker.com/info-leak/apple?key1=banana
ndn://attacker.com/info-leak/…
…

Flow to create anomalous names with dictionary coding (i.e., *steganography*) in “com” domain

[9] ITU-T, http://www.itu.int/en/ITU-T/publications/Pages/latest.aspx

- Leaked data
  - 3.4 MB Zip file compressing 3 Pdf files from latest ITU-T recommendations [9]
- Threshold for each attribute in anomalous

	$\nu = 0.4$			
	L_P	L_Q	$N_/$	$N_=$
com	15	112	2	4
net	20	76	2	3
org	17	98	2	4
info	16	80	2	1
jp	190	0	4	0
fr	26	178	3	14
uk	12	132	2	7

- Dictionary coding with 65,536 dictionary words from WordNet [7]
  - Table with each dictionary word and 4 hexadecimal digits to each word (one word is equal to 2 Bytes)

Name-Based Filter Using One-Class SVM

- One-class SVM [4] is unsupervised method to perform anomaly detection
  - ➢Adapted if not many samples

[4] B. Schölkopf, et al., “Estimating the Support of a High-Dimensional Distribution,” *Neural Comput.*, vol. 13, no. 7, pp. 1443–1471, Jul. 2001.

- Regarding NDN architecture, there are currently not anomalous traffic nor names available
  - ➢Extracting URL properties as characteristics of legitimate names and applying them for one-class SVM filter

Filter using one-class SVM inspects names dropped by filter using search engine information

Performance Evaluation

- Performance metric
  - Per-packet throughput of information-leakage (Bytes/Interest_packet)
- Each TLD dataset is separated into two sets to create name-based filter using one-class SVM
  - Training set for each TLD: 800,000 URLs
  - Testing set for each TLD: 200,000 URLs
- Assumption
  - Defending knows attack method (i.e., *steganography-embedded* Interest packets) but not its parameters
  - Attacker knows countermeasure but not its parameters
    - This case is of benefit to attacker

Performance Evaluation

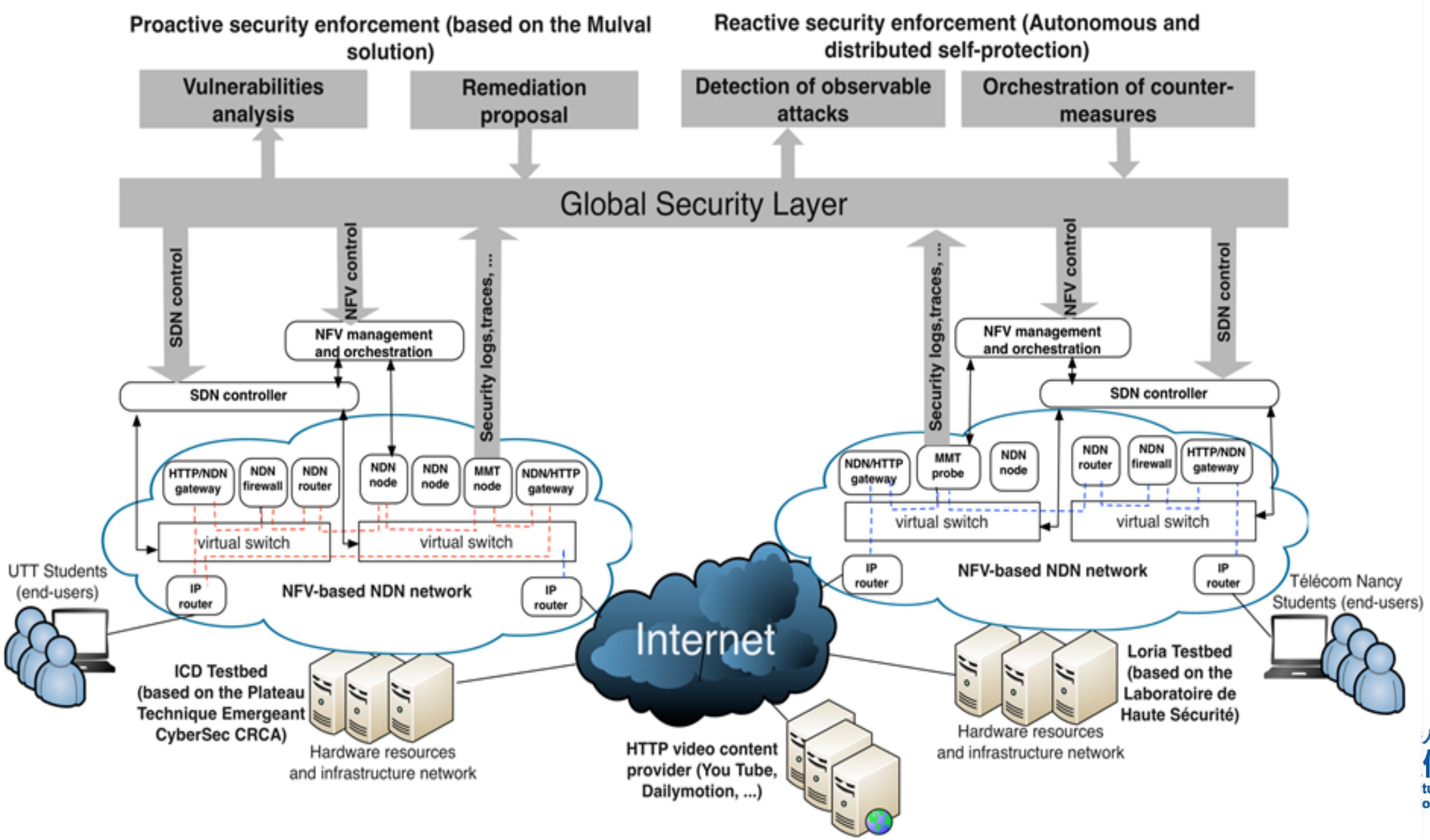
- Without SVM filter
  - Attacker builds names and leak information (steganography)
  - 2.06 Kbytes/Interest_packets

- With SVM filter (tuned parameters)
  - 7.79Bytes/Interest_packets

By using filter, malware has to send 264 times (2.06 KB/ 7.79B) more Interest packets to the attacker than without using filter

Project ANR Doctor (2014-2017)
http://www.doctor-project.org/

- Deployment of new network functions and protocols (e.g.: NDN) in a virtualized networking environment (e.g.: NFV)
  - Monitoring, managing and securing (using SDN for reconfiguration)
- Partners: Orange, Thlaes, Montimage, UTT, LORIA/CNRS (900k€)


Proactive security enforcement (based on the Mulval solution)
Reactive security enforcement (Autonomous and distributed self-protection)
Vulnerabilities analysis
Remediation proposal
Detection of observable attacks
Orchestration of counter-measures
Global Security Layer
SDN control
NFV control
Security logs,traces, ...
NFV management and orchestration
SDN controller
HTTP/NDN gateway
NDN firewall
NDN router
NDN node
NDN node
MMT node
NDN/HTTP gateway
virtual switch
virtual switch
IP router
NFV-based NDN network
IP router
UTT Students (end-users)
ICD Testbed (based on the Plateau Technique Emergeant CyberSec CRCA)
Hardware resources and infrastructure network
Internet
HTTP video content provider (You Tube, Dailymotion, ...)
Security logs,traces, ...
NFV control
SDN control
NFV management and orchestration
SDN controller
NDN/HTTP gateway
MMT probe
NDN node
NDN router
NDN firewall
HTTP/NDN gateway
virtual switch
virtual switch
IP router
NFV-based NDN network
IP router
Télécom Nancy Students (end-users)
Loria Testbed (based on the Laboratoire de Haute Sécurité)
Hardware resources and infrastructure network

Conclusion

- Information-leakage is main Internet Security threat
  - Targeted Attacks
- NDN as Future Internet architecture
  - Prevent leakage information from names (Interest Packets)
    - Steganography-embedded attacks in Names
- NDN Names filtering heuristics
  - Based on URLs statistics
  - Up to 15% of anomalous URLs
  - Firewall for NDN
- SVM-based filtering heuristics
  - Choke throughput of information-leakage
  - Up to 264 more Interest packets to leak the same amount of information
- Designing Naming Scheme for Named-Data Networking (NDN)
  - Privacy in NDN

Thank You

- Questions ?

thomas@nict.go.jp