Challenges to use AI technology for cyber security

December 2018

Keiichi Shima <keiichi@iijlab.net>

Self Introduction

- Keiichi Shima
  - Deputy Director, IIJ Research Laboratory
  - Work area: Distributed Systems, IP version 6, Mobile IPv6, Network Mobility, Distributed Filesystems, Security Log analysis

What is IIJ?

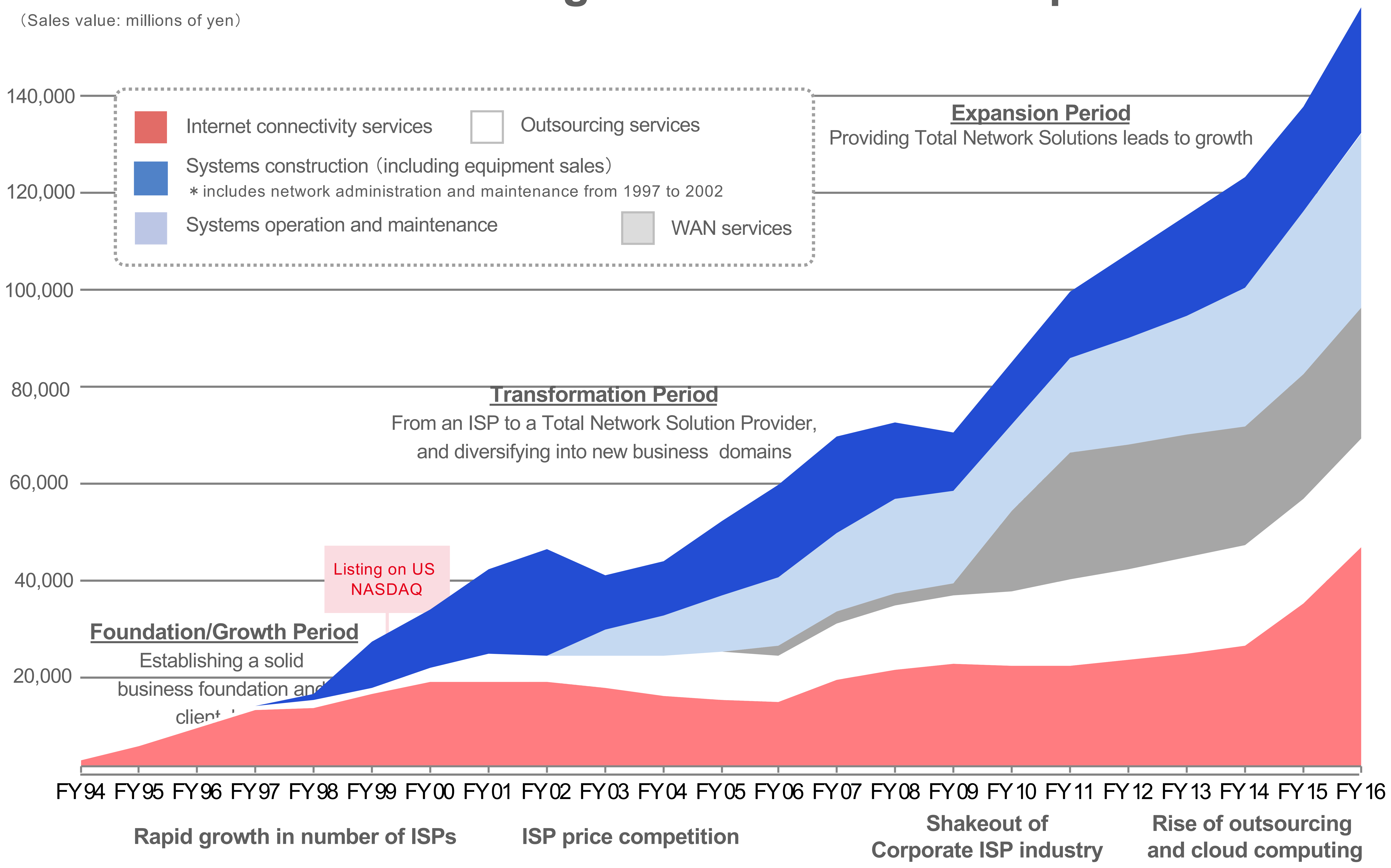
- IIJ is one of the major Internet service providers in Japan
- Providing Internet connectivity, Internet services, System Integration solutions
- Main customers are companies and government

Corporate history

Period	Year	Development of IIJ management and services	Trends in Internet and Telecommunications Industry
Founding Period Spread of Internet connectivity services	1992	Company founded	US Internet Society founded
	1993	Launched Internet connectivity services Japan's first	JPNIC founded Japan Internet Society founded（now Internet Association Japan）
	1994	Registered as Special Type II Carrier with（then）Posts and Telecommunications Ministry Launched dial-up IP service Japan's first Launched firewall service Japan's first	US Mosaic Communications founded, Yahoo! launched Netscape Navigator1.0 released
	1995		Windows95 goes on sale in Japan The word "Internet" selected as one of the trendy words of the year
	1996	Began operation of Asian regional Internet backbone（A-Bone） First Japanese ISP to launch ISP business in the USA	Yahoo! Japan service launched NTT, OCN service launched
	1997		KDD launches domestic telecom service in Japan MPT allows International Public-Private-Public Connection
Expansion Period Development of products based on Internet technology Development of outsourcing needs	1998	Launched IP multicast distribution service Japan's first Established Netcare,Inc. Development and sale of SEIL advanced router Japan's first	CATV Internet connectivity begun
	1999	Listed on US NASDAQ National Market Introduced Service Level Agreement（SLA）Japan's first Launched IPv6 commercial service Japan's first	i-Mode（NTT DoCoMo）launched, 2-channel launched NTT East and West launch ISDN flat-rate communications service
	2000		All companies launch ADSL connectivity services
	2001	Launched world's first wide-area Ethernet service world's first	Optical fiber service launched（NTT East /West）, Yahoo! BB business service launched, FOMA service launched（NTT DoCoMo）, METI implements regulations to prevent spam email
	2002	Launched IX service JPNAP Launched Japan's largest CDN platform service	BB Phone commercial service launched（Softbank） Basic Resident Register Network goes into operation
	2003	Developed SMF, World's first network service operating system Japan's first	
Evolution of social infrastructure Emerging demand for solutions	2004		Tokyo Metropolitan Police Department issues warning about phishing scam P2P telephone Skype 1.0 launched
	2005	Listed on Mother's section of Tokyo Stock Exchange（TSE）	Wireless broadband broadcast Gyao launched by USEN
	2006	Listing moved to TSE First Section Launched anti-spam mail service Japan's first Patents issued for SMF（3774433）and SFM-LAN（3996922）	Government information leaks via Winny spark concern, Google purchases YouTube in a stock swap, NGN field trials begun（NTT Group）
	2007	Established IIJ Innovation Institute Inc.	Apple Corporation releases iPhone MIAC establishes New Generation Network Promotion Forum after NGN
	2008	Launched MVNO service IIJ Mobile Launched IIJ Direct Access Japan's first	NGN service FLETS HIKARI NEXT launched
	2009	Launched IIJ GIO service Launched IIJ Secure Web gateway service	Cloud computing becomes hot topic
	2010	Established IIJ Global Solutions Inc.	Twitter usage expands
	2011	Matsue Data Center Park launched	World IPv6 Day established as IPv4 addresses start to run out
	2012	Established Stratosphere Inc.	
	2013	Established IIJ Europe Limited	With the Internet of Things（IoT）gathering momentum, Google has developed the"Google Glass"wearable device
	2014	Acquired RYUKOSHA NETWARE Inc.	
	2015	Established PT. Biznet Gio Nusantara with Biznet Networks in Indonesia	Mobile carrieres are obliged to remove SIM locks from handsets.
	2016	Established Leap Solutions Asia Co., Ltd. with TCCT in Thailand	

Expanding into new business domains

From an ISP to a full ranged network solution provider



IIJmio

日本語 English 中文（简体） 中文（繁體） 한국어 ภาษาไทย

JAPAN TRAVEL SIM フルMVNO版 1.5GB 30days 3GB 30days

2018.4/2 START!

詳しくはこちら ▶

Service overview	Recharge	APN settings Registration	Where to buy	FAQ	List of tested devices	Support Page	Customer support

1GB 30days

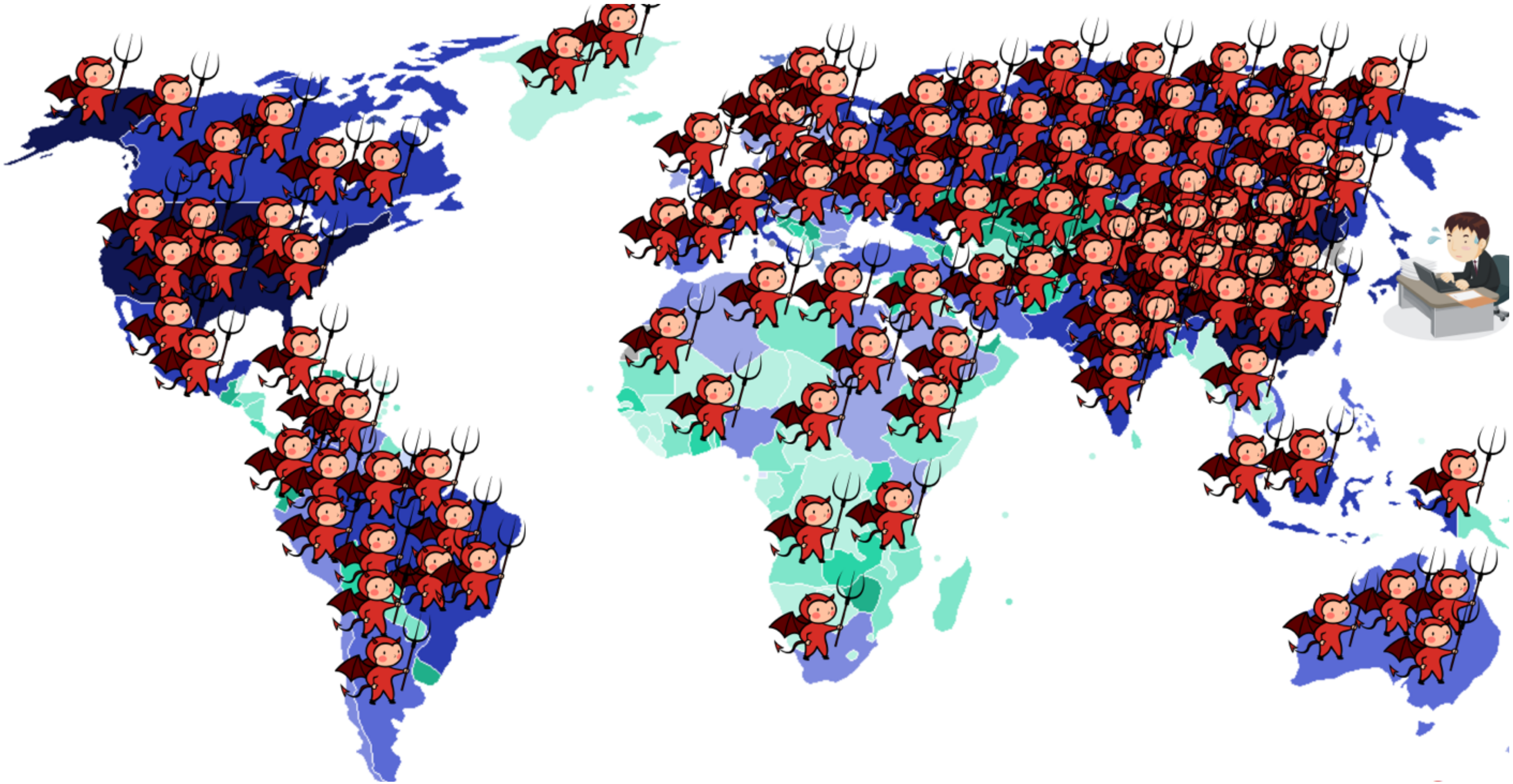
2GB 3months

Charge

Security Concerns

- More and more services depends on IT infrastructure
- (Bad) people found that security attacks make money
- New technologies are invented every day
- Easy to deploy a technology since Internet is designed to be so

Research Background



Research Background

<u>Many incident reports, everyday</u>

More sophisticated, organized attacks
Constantly invented new attack methods

<u>Depends on individual</u>

Incident handling depends on skill
Quality depends on experience
Not scalable operation

Automete incident type, affected range, and counter actions

Find “Symptom” of incident and guess type and range
Propose counter methods based on the past action history to operators

How AI will be Used

Our Objectives

1. Detection of symptom of attack or anomaly using big data and machine learning
   - Mitigation for zero-day attacks
   - Combined with existing IDS/IPS
2. Prediction and discovery of symptom of attack using social dataset
   - Finding relationship between social actions monitored on Web / SNS and cyber space activities
   - Prediction of attack using darknet information
3. Incident response assistance using machine learning
   - Assisting operator to pickup evidence of attack from large dataset
   - Suggesting first response action learned from past response history
4. Providing open dataset
   - Keeping individual privacy that may be included in the dataset
   - Try to provide wide variety of dataset for security research

This Project is

- Supported by the Japanese Government Funding
- 2.5 year long program started from Oct. 2017

Topics Today

- AI assisted data classification
  - Classify packets into normal or attack
  - Classify IP sources into normal or malicious
  - Classify URL strings into benign or phishing

AI is Great?

Why?

Is AI new idea?

- AI is not a new idea (depends on what is AI)
- Machine leaning (SVM: 1961, Random Forest: 2001)
  - Need to carefully define “Features”
  - Require deep knowledge of the target domain to find “effective” features
- Deep learning
  - The concept was published around 2000
  - But was not widely adopted for real use cases

Change

- The idea of deep leaning was great but how to train the network was difficult
- In 2012, Krizhevsky won the prize at ILSVRC (ImageNet Large Scale Visual Recognition Challenge) using neural network
  - 10% better accuracy than past
- After that, staring from image/voice recognition field, many classification fields, text recognition field, and computer Go game fields, the application area is keep spreading

What is Different?

- (Recent) Deep learning may help to solve difficulties to find good features
- Using a lot of existing data
  - Collecting and using huge amount of data becomes possible
  - Train the neural network to react the “features” of the data by giving that amount of data
  - Data processing speed becomes feasible thanks to GPU technology

Can we use DL for Network Data?

- DL achieved remarkable success in image recognition fields
- Ideally, we just want put “Log” data and let DL judge something
- Without deep domain-specific knowledge of the target data

Case 1: Classify Packet Data

Classify Packet Data

- Classify a packet into benign or malicious

Classify Packet Data

- In image recognition, we give the binary data of an image to the neural network to train it
- Can it be possible for network data?

Packet Data

```
0x0000:  6006 551d 00d5 11ff fe80 0000 0000 0000
0x0010:  14c5 786e cfa3 4b36 ff02 0000 0000 0000
0x0020:  0000 0000 0000 00fb 14e9 14e9 00d5 8e5e
0x0030:  0000 8400 0000 0001 0000 0001 1a4b 6569
0x0040:  6963 6869 2773 204d 6163 426f 6f6b 2050
0x0050:  726f 2032 3031 370f 5f63 6f6d 7061 6e69
0x0060:  6f6e 2d6c 696e 6b04 5f74 6370 056c 6f63
0x0070:  616c 0000 1080 0100 0011 9400 6b16 7270
0x0080:  4241 3d32 373a 3745 3a36 443a 3743 3a36
0x0090:  393a 4332 1172 7041 443d 6461 3663 3639
0x00a0:  3965 6635 6635 1172 7048 493d 6130 6361
0x00b0:  …
```

Packet Data

```
0x0000:   6006 551d 00d5 11ff fe80 0000 0000 0000
0x0010:   14c5 786e cfa3 4b36 ff02 0000 0000 0000
0x0020:   0000 0000 0000 00fb 14e9 14e9 00d5 8e5e
0x0030:   0000 8400 0000 0001 0000 0001 1a4b 6569
0x0040:   6963 6869 2773 204d 6163 426f 6f6b 2050
0x0050:   726f 2032 3031 370f 5f63 6f6d 7061 6e69
0x0060:   6f6e 2d6c 696e 6b04 5f74 6370 056c 6f63
0x0070:   616c 0000 1080 0100 0011 9400 6b16 7270
0x0080:   4241 3d32 373a 3745 3a36 443a 3743 3a36
0x0090:   393a 4332 1172 7041 443d 6461 3663 3639
0x00a0:   3965 6635 6635 1172 7048 493d 6130 6361
0x00b0:   …
```

Think Differently

- Can we treat the packet similar to the image data?

Count Them

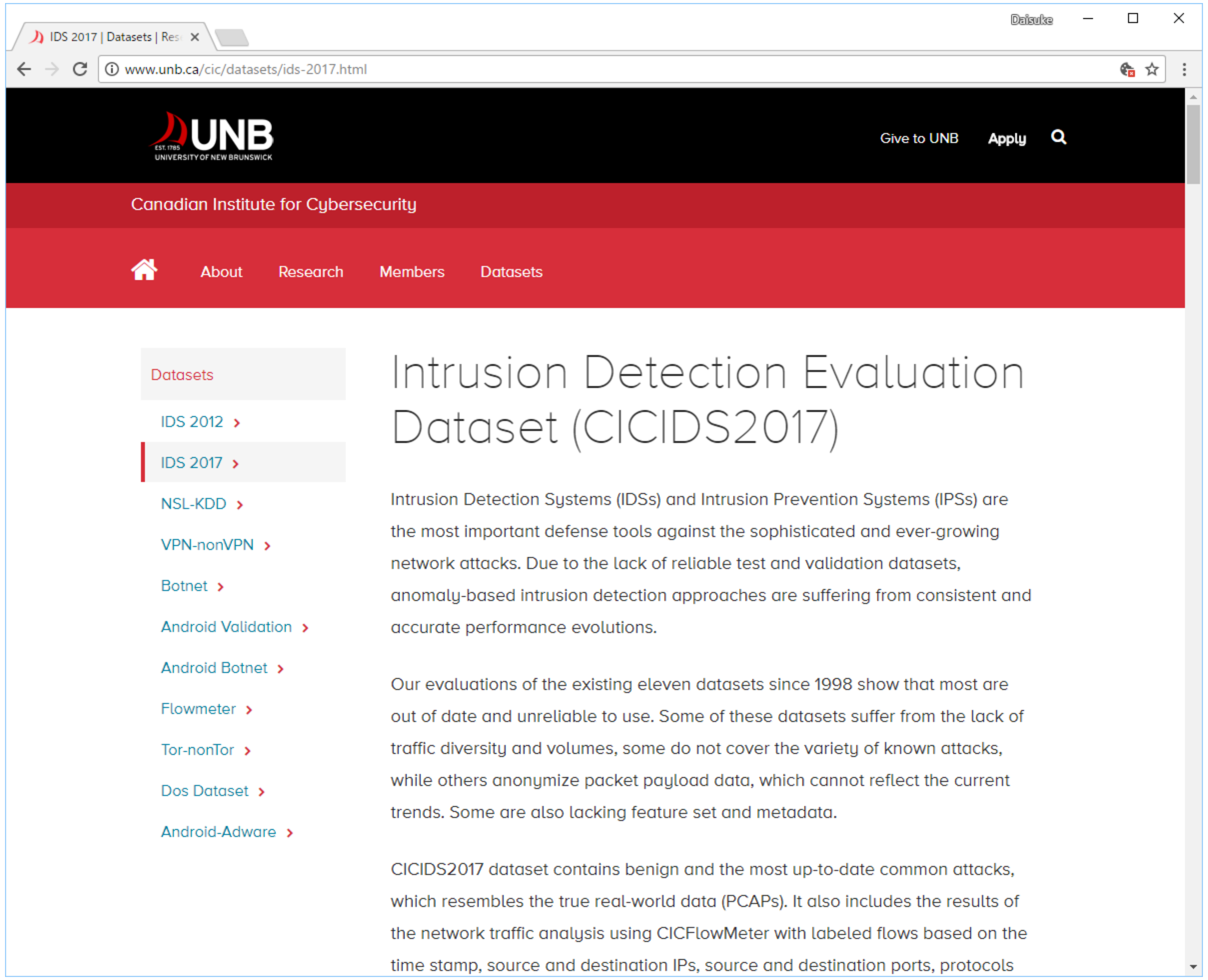
```
0x0000:  6006 551d 00d5 11ff fe80 0000 0000 0000
0x0010:  …
```

0x60 => 1, 0x00 => 13, 0x06 => 1, 0x65 => 1, …

256 dimension data

CIC-IDS Dataset

- Publicly available datasets provided by University of New Brunswick
- IDS2017 dataset contains
  - Monday: Normal data only
  - Tuesday: w/ Bruteforce
  - Wednesday: w/ DoS/DDoS
  - Thursday: w/ Web attacks
  - Friday: w/ Botnet ARES



Preliminary Results

	Accuracy	FPR	FNR
Bruteforce	0.9793	0.98%	0.19%
Web attacks	0.9565	0.00%	9.41%
Botnet ARES	0.9558	0.01%	3.41%

Case 2: Classify TCP Connections

Classify TCP Connections

- Can we distinguish “good” TCP connections and “bad” TCP connections based on their connection establishment patterns?

Basic Idea

Make an image of SYNs (Timestamp, Src port, Dst port, Seq #, Window size)

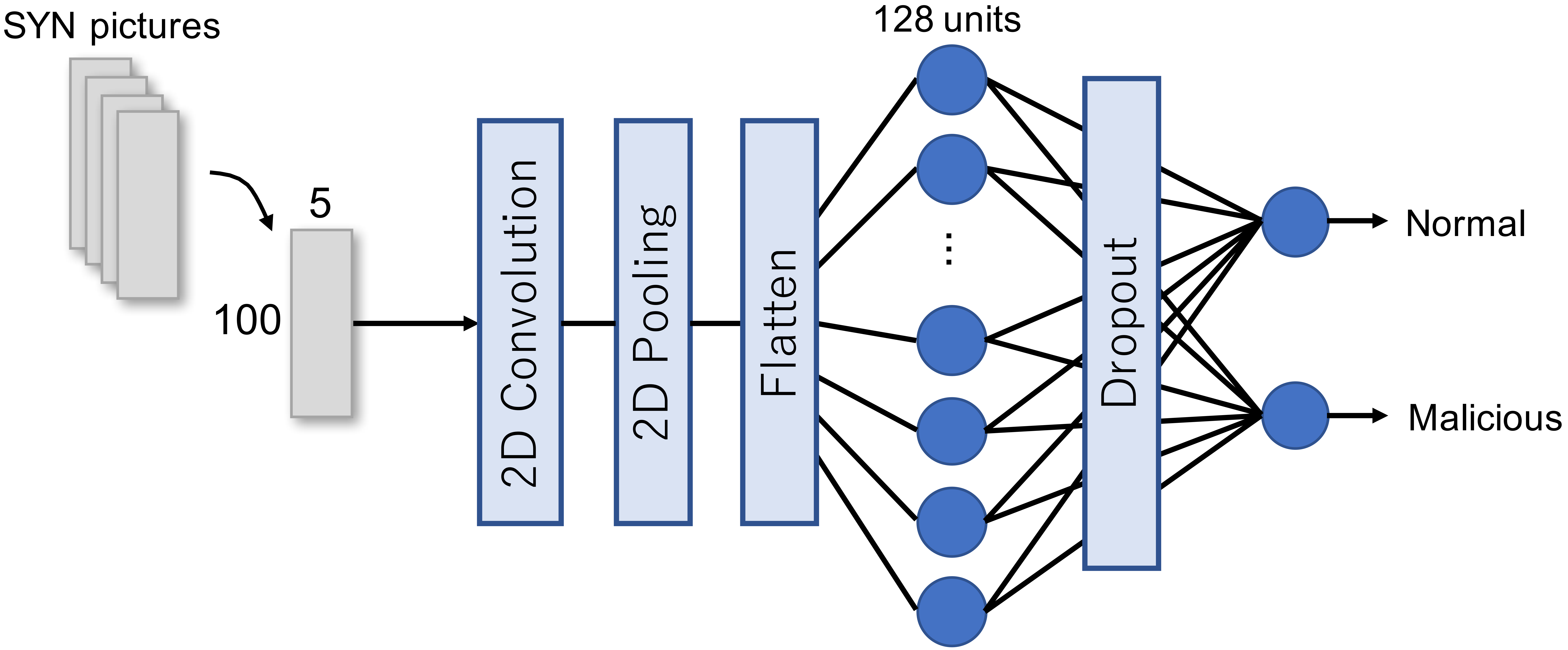
SYNs arrived at Honeypot → Bad

SYNs observed in a life segment → Good

Examples of “Bad” SYN packets

Examples of “Good” SYN packets

CNN Topology



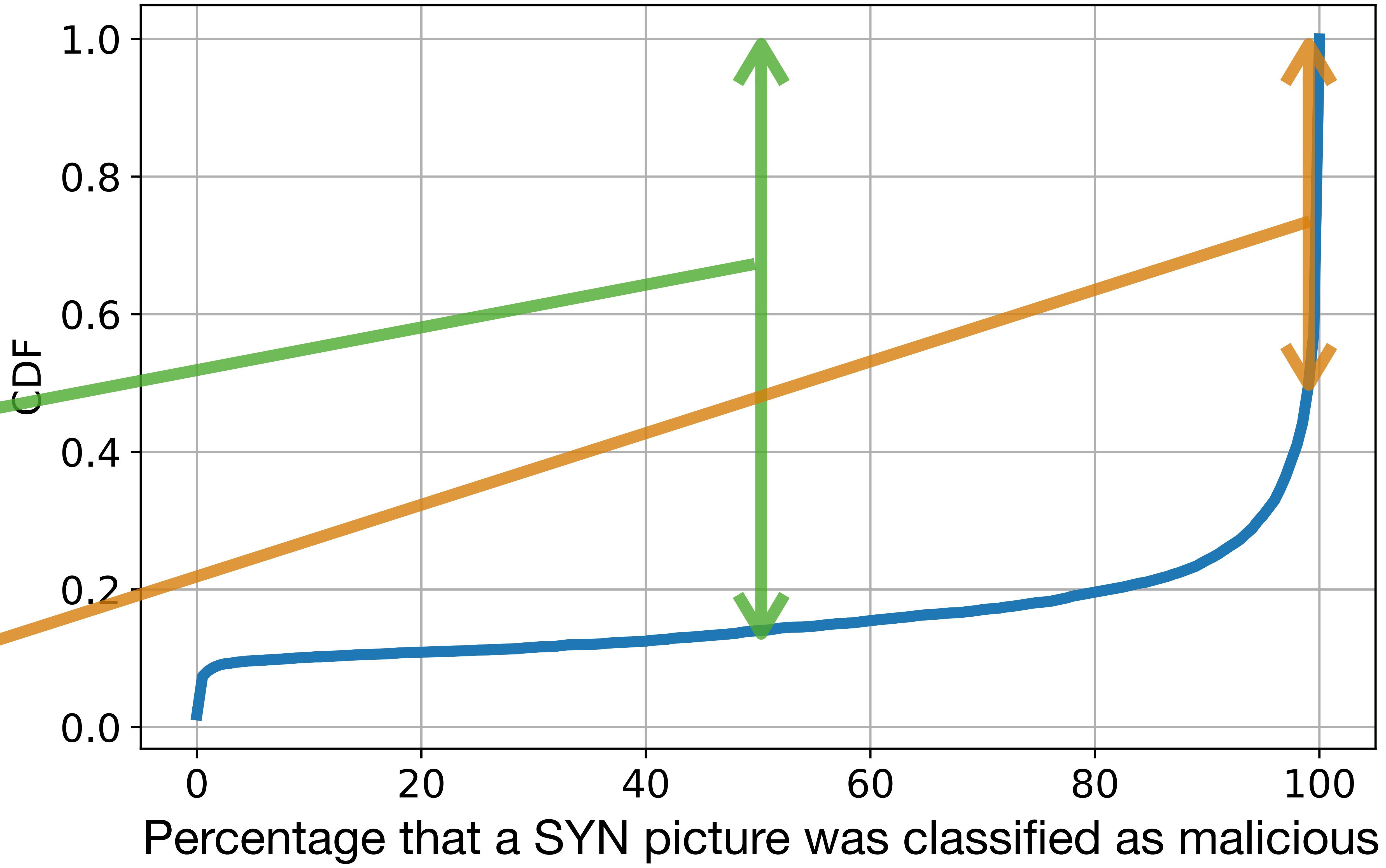
Ryo Nakamura, Yuji Sekiya, Daisuke Miyamoto, Kazuya Okada, Tomohiro Ishihara, "Malicious Host Detection by Imaging SYN Packets and A Neural Network", Proceedings of IEEE International Symposium on Networks, Computers and Commnications (ISNCC2018), Rome, Italy, June 2018.

Preliminary Results

Classify packets arrived at the Darknet
(Assuming that all of them are malicious)

86% packets are classified as malicious
with more than 50% accuracy

50% packets are classified as malicious
With more than 99% accuracy



Ryo Nakamura, Yuji Sekiya, Daisuke Miyamoto, Kazuya Okada, Tomohiro Ishihara, "Malicious Host Detection by Imaging SYN Packets and A Neural Network", Proceedings of IEEE International Symposium on Networks, Computers and Commnications (ISNCC2018), Rome, Italy, June 2018.

Case 3: Classify URL strings into benign or phishing

Phishing

- Phishing is one of the major techniques to steal personal information
  - 233,040 attacks were reported in 2Q 2018 (*1)
- There exists several services (products) to defend them
  - URL whitelisting
  - Contents investigation

(*1) Anti Phishing WG report: http://docs.apwg.org/reports/apwg_trends_report_q2_2018.pdf

URL Features?

- Challenges
  - Is there any hidden features in the URL strings used for phishing sites?
  - Is it possible to distinguish “white” URLs and “black” URLs by just looking at the URL strings?
- We try to vectorize URLs to use as input information of ML methods without any specific domain knowledge

Traditional Features

- The length of URL
- The number of dots and/or slashes
- Ratio of alphabets, numbers, and marks
- Site rank
- The time from when the domain was registered
- etc…

Think Differently Again

How to Vectorize

```
www.iij.ad.jp/index.html
```

Split characters

```
w w w . i i j . a d . j p / i n d e x . h t m l
```

Convert the URL into HEX values

```
7777772E69696A2E61642E6A703F696E6465782E68746D6C
```

Extract 8-bits values by shifting 4 bits in the HEX values

```
77,77,77,77,77,72,2E,
E6,69,96,69,96,6A,A2,
2E,E6,61,16,64,42,2E,
E6,6A,A7,70
```

```
3F,F6,69,96,6E,E6,64,
46,65,57,78,82,2E,E6,
68,87,74,46,6D,D6,6C
```

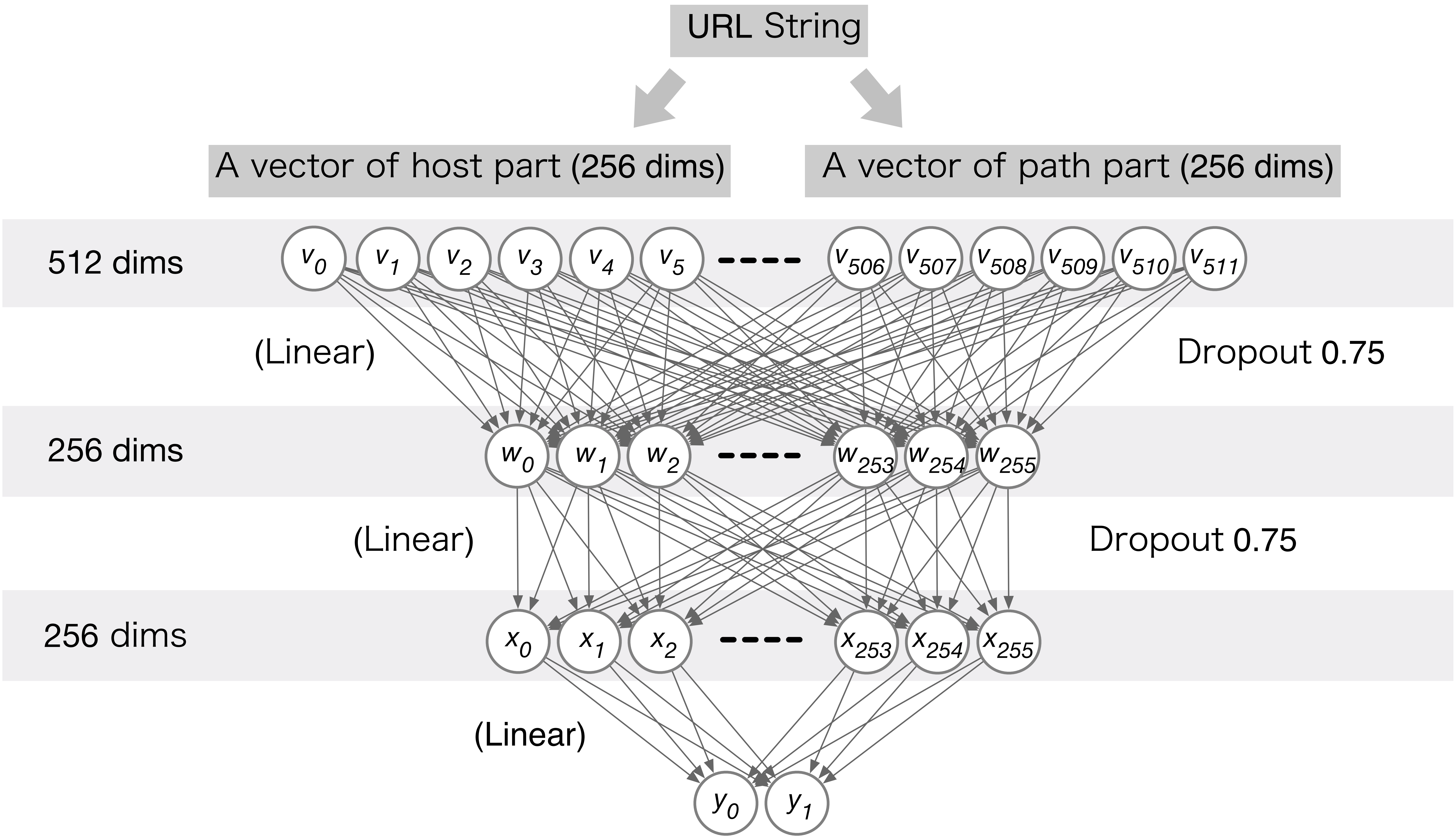
Count the number of unique values for the host part and the URL path part respectively (Bag of features)

How to Vectorize?

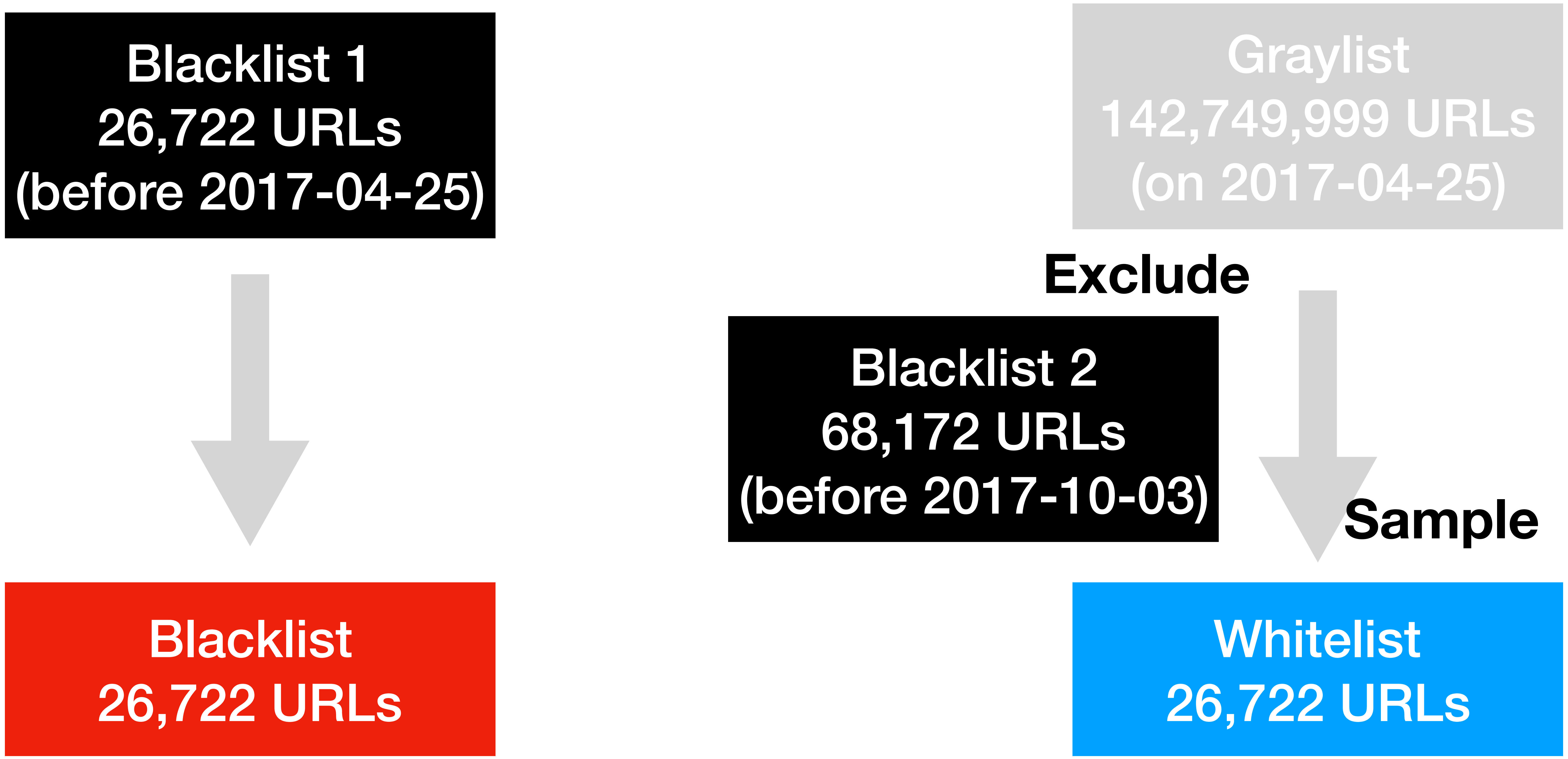
www.iij.ad.jp		index.html	
16 → 1	2E → 3	2E → 1	46 → 1
42 → 1	61 → 1	57 → 1	65 → 1
64 → 1	69 → 2	68 → 1	6C → 1
6A → 2	70 → 1	6D → 1	74 → 1
72 → 1	77 → 5	78 → 1	82 → 1
96 → 2	A2 → 1	87 → 1	D6 → 1
A7 → 1	E6 → 3	E6 → 1	
256 dimensional sparse vector		256 dimensional sparse vector	

512 dimensional sparse vector

Neural Network Topology



Making Datasets



Datasets

TABLE I. URL DATASETS FOR TRAINING

Type	Content	Count
Blacklist 1	Phishing site URLs reported at PhishTank.com before 2017-04-25. This list is used as a blacklist for learning and testing in conjunction with the Whitelist 1.	26,722
Blacklist 2	Phishing site URLs reported at PhishTank.com before 2017-10-03. This list is used to cleanse the target access log captured at the anonymous research organization X.	68,172
Whitelist 1	A sampled list of URL access log captured at the anonymous research organization X on 2017-04-25 excluding the entries listed in the Blacklist 2. This list is used for learning and testing in conjunction with the Blacklist 1.	26,722

Keiichi Shima, Daisuke Miyamoto, Hiroshi Abe, Tomohiro Ishihara, Kazuya Okada, Yuji Sekiya, Hirohchika Asai, Yusuke Doi, “Classification of URL bitstreams with Bag of Bytes”, First International Workshop on Network Intelligence (NI2018), 20-22 February 2018

Results

TABLE II. RESULTS OF ACCURACY AND TRAINING TIME USING WHITELIST 1 AND BLACKLIST 1 IN TABLE I

	Optimizer	Accuracy (%)	Training time (s)
Our method	Adam	94.18	32
–	AdaDelta	93.54	31
–	SGD	88.29	31
eXpose[6]	Adam	90.52	119
–	AdaDelta	91.31	119
–	SGD	77.99	116

- Our approach could achieve better accuracy compared to the eXpose(*1) work which uses similar approach using a more complex deep neural network

(*1) J. Saxe and K. Berlin, "eXpose: A character-level convolutional neural network with embeddings for detecting malicious URLs, file paths and registry keys," CoRR, vol. abs/1702.08568, February 2017.

Discussion

- Difficulties in making datasets
  - How to label network data
  - How to generalize the dataset
- Difficulties in comparison of results
  - How to compare our idea and past idea without using the same data

Summary

- The breakthrough of deep Learning technology affects many existing fields
- We are trying to utilize the technology for network data
- The goal is to provide better assistant mechanism without any domain specific knowledge of target data
- We propose stupidly simple vectorization mechanisms to handle network data to use for neural network
- So far we are seeing fairly good results (but not sure it is general results or not)

Related Work

- M. Antonakakis et al., “Understanding the mirai botnet,” in 26th USENIX Security Symposium (USENIX Security 17). Vancouver, BC: USENIX Association, 2017, pp. 1093–1110. [Online]. Avail- able: https://www.usenix.org/conference/usenixsecurity17/technical- sessions/presentation/antonakakis
- Y. Ohsita et al., “Detecting distributed denial-of-service attacks by analyzing tcp syn packets statistically,” in Global Telecommunications Conference, 2004. GLOBECOM’04. IEEE, vol. 4. IEEE, 2004, pp. 2043–2049.
- D. M. Divakaran et al., “Detection of syn flooding attacks using linear prediction analysis,” in 2006 14th IEEE International Conference on Networks, vol. 1, Sept 2006, pp. 1–6.
- S. H. A. Ali et al., “A neural network model for detecting ddos attacks using darknet traffic features,” in 2016 International Joint Conference on Neural Networks (IJCNN), July 2016, pp. 2979–2985.
- X.Yuanetal.,“Deepdefense:Identifyingddosattackviadeeplearning,” in 2017 IEEE International Conference on Smart Computing (SMART- COMP), May 2017, pp. 1–8.
- C. Fachkha and M. Debbabi, “Darknet as a source of cyber intelligence: Survey, taxonomy, and characterization,” IEEE Communications Surveys Tutorials, vol. 18, no. 2, pp. 1197–1227, Secondquarter 2016.
- S. Panjwani et al., “An experimental evaluation to determine if port scans are precursors to an attack,” in 2005 International Conference on Dependable Systems and Networks (DSN’05), June 2005, pp. 602–611.

Related Work

- S. Garera, N. Provos, M. Chew, and A. D. Rubin, “A framework for detection and measurement of phishing attacks,” in Proceedings of the 2007 ACM Workshop on Recurring Malcode, ser. WORM ’07. New York, NY, USA: ACM, November 2007, pp. 1–8.
- J. Ma, L. K. Saul, S. Savage, and G. M. Voelker, “Beyond blacklists: Learning to detect malicious web sites from suspicious URLs,” in Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ser. KDD ’09. New York, NY, USA: ACM, June 2009, pp. 1245–1254.
- P. Prakash, M. Kumar, R. R. Kompella, and M. Gupta, “PhishNet: Predictive blacklisting to detect phishing attacks,” in 2010 Proceedings IEEE INFOCOM, ser. INFOCOM, 2010, pp. 1–5.
- B. Sun, M. Akiyama, T. Yagi, M. Hatada, and T. Mori, “AutoBLG: Automatic URL blacklist generator using search space expansion and filters,” in 2015 IEEE Symposium on Computers and Communication, ser. ISCC, July 2015, pp. 625–631.
- J. Saxe and K. Berlin, “eXpose: A character-level convolutional neural network with embeddings for detecting malicious URLs, file paths and registry keys,” CoRR, vol. abs/1702.08568, February 2017.

Internship Program

https://www.iij-ii.co.jp/en/career/internship.html