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

**Founding Period**
- 1992: Company founded
- 1993: Launched Internet connectivity services
- 1994: Registered as Special Type II Carrier with (then) Posts and Telecommunications Ministry
- 1995: Began operation of Asian regional Internet backbone (A-Bone)
- 1996: First Japanese ISP to launch ISP business in the USA
- 1998: Launched IP multicast distribution service
- 1999: Listed on US NASDAQ National Market
- 2000: Launched world’s first wide-area Ethernet service
- 2001: Launched IX service JPNAP
- 2002: Developed SMF, World’s first network service operating system
- 2003: Listed on Mother’s section of Tokyo Stock Exchange
- 2004: Launched anti-spam mail service
- 2005: Listed moving to TSE First Section
- 2006: Launching anti-spam mail service
- 2007: Established IJJ Innovation Institute Inc.
- 2008: Launched MVNO service IIJ Mobile
- 2009: Launched IXGIO service
- 2010: Established IIJ Global Solutions Inc.
- 2011: Matsue Data Center Park launched
- 2012: Established Stratosphere Inc.
- 2013: Established IIJ Europe Limited
- 2014: Acquired RYUKOSHA NETWARE Inc.
- 2015: Established PT, Biznet Gio Nusantara with Biznet Networks in Indonesia
- 2016: Established Leap Solutions Asia Co., Ltd. with TCCT in Thailand

**Development of IJJ management and services**
- 1992: Launched Internet connectivity services
- 1996: First Japanese ISP to launch ISP business in the USA
- 2001: Launched IPv6 commercial service
- 2003: Developed SMF, World’s first network service operating system
- 2005: Launched anti-spam mail service
- 2008: Launched MVNO service IIJ Mobile
- 2009: Launched IXGIO service
- 2011: Matsue Data Center Park launched
- 2012: Established Stratosphere Inc.
- 2013: Established IIJ Europe Limited
- 2014: Acquired RYUKOSHA NETWARE Inc.
- 2015: Established PT, Biznet Gio Nusantara with Biznet Networks in Indonesia
- 2016: Established Leap Solutions Asia Co., Ltd. with TCCT in Thailand

**Trends in Internet and Telecommunications Industry**
- 1991: Japan Internet Society founded
- 1992: Japan Mosaic Communications founded, Yahoo! launched
- 1993: Netscape Navigator 1.0 released
- 1994: Windows 95 goes on sale in Japan
- 1995: The word “Internet” selected as one of the trendy words of the year
- 1996: Yahoo Japan service launched
- 1997: NTT, DCN service launched
- 1998: KDD launches domestic telecom service in Japan
- 1999: CATV Internet connectivity begun
- 2000: iMode (NTT DoCoMo) launched, 2-channel launched
- 2001: NTT East and West launch ISDN flat-rate communications service
- 2002: All companies launch ADSL connectivity services
- 2003: Optical fiber service launched (NTT East/West), Yahoo! BB business service launched, FOMA service launched (NTT DoCoMo), METI implements regulations to prevent spam email
- 2004: BB Phone commercial service launched (Softbank), Basic Resident Register Network goes into operation
- 2005: Tokyo Metropolitan Police Department issues warning about phishing scam
- 2006: P2P telephone Skype 1.0 launched
- 2007: Wireless broadband broadcast Gyao launched by USEN
- 2008: Government information leaks via Winny spark concern, Google purchases YouTube in a stock swap, NGN field trials begun (NTT Group)
- 2009: Apple Corporation releases Phone
- 2010: MIAC establishes New Generation Network Promotion Forum after NGN field trials begun (NTT Group)
- 2011: Cloud computing becomes hot topic
- 2012: Twitter usage expands
- 2013: World IPv6 Day established as IPv4 addresses start to run out
- 2014: With the Internet of Things (IoT) gathering momentum, Google has developed the "Google Glass" wearable device
- 2015: Mobile carriers are obliged to remove SIM locks from handsets.
Expanding into new business domains

From an ISP to a full ranged network solution provider

(Sales value: millions of yen)

<table>
<thead>
<tr>
<th>FY</th>
<th>Internet connectivity services</th>
<th>Outsourcing services</th>
<th>Systems construction (including equipment sales)</th>
<th>Systems operation and maintenance</th>
<th>WAN services</th>
</tr>
</thead>
<tbody>
<tr>
<td>94</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>98</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Expansion Period
Providing Total Network Solutions leads to growth

Foundation/Growth Period
Establishing a solid business foundation and diversifying into new business domains

Transformation Period
From an ISP to a Total Network Solution Provider, and diversifying into new business domains

Listing on US NASDAQ

Pricing and expansion of cloud services

Rapid growth in number of ISPs
ISP price competition
Shakeout of Corporate ISP industry
Rise of outsourcing and cloud computing
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

Many incident reports, everyday
More sophisticated, organized attacks
Constantly invented new attack methods

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

Automate 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

Anomaly detection alerts Social information

Guess type of attack and range

Auto detect

Assist

Security Operator

Past incident response Minimize the damage

Find similar incident response flow from past history

Propose counter method

Quick decision making and counter action
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)
  - 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, starting 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 0000 0000 0001 0000 0001
0x0050: 1a4b 6569 6963 6869 2773 0000 0000 0001
0x0060: 0000 0000 0001 0000 0001 1a4b 6569 6963
0x0070: 6869 2773 0000 0000 0001 0000 0001 1a4b
0x0080: 6569 6963 6869 2773 0000 0000 0001 0000
0x0090: 0001 0000 0001 1a4b 6569 6963 6869 2773
0x00a0: 0000 0000 0001 0000 0001 1a4b 6569 6963
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 001 0000 0001 1a4b 6569
0x0040:  6963 6869 2773 204d 6163 426f 6f6b 2050
0x0050:  726f 2032 3031 370f 5f63 6f6e 6d69 6e6b
0x0060:  04 5f74 6370 056c 6f63 616c 0000 1080
0x0070:  0100 0011 9400 6b16 7270 4241 3d32 373a
0x0080:  373a 3745 3a36 443a 3743 3a36 393a 4332
0x0090:  1172 7041 443d 6461 3669 6635 6635 1172
0x00a0:  7048 493d 6130 6363 6f6d 7061 6e69 6f6e
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

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>FPR</th>
<th>FNR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bruteforce</strong></td>
<td>0.9793</td>
<td>0.98%</td>
<td>0.19%</td>
</tr>
<tr>
<td><strong>Web attacks</strong></td>
<td>0.9565</td>
<td>0.00%</td>
<td>9.41%</td>
</tr>
<tr>
<td><strong>Botnet ARES</strong></td>
<td>0.9558</td>
<td>0.01%</td>
<td>3.41%</td>
</tr>
</tbody>
</table>
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

Fig. 2: Example SYN pictures. Note that the SYN pictures are transposed due to limitations of the space.

In contrast, normal hosts did not send SYN packets at such regular intervals. From the second row, periodic cycles can be seen in transitions of source port numbers in the SYN pictures of malicious hosts. Such cycles do not appear in the SYN pictures of normal hosts. In addition, normal hosts use a few destination ports during 100 SYN packets in contrast with malicious hosts sending SYN packets to a single destination port. In this manner, the SYN pictures reveal features of behavior on sending SYN packets in the human and machine-readable way.

C. Neural Network Topology

To classify the SYN pictures, we used a convolutional neural network (CNN) [9]. CNN is a variation of neural networks for analyzing visual images. The neural network topology we used is depicted in Figure 3. The neural network is composed of six layers: a 2D convolution layer, a 2D pooling layer, a flattening layer, a 128 units dense layer, a dropout layer and a fully-connected layer for output. The first 2D convolution layer and the next 2D pooling layer extract features from SYN pictures using 32 \( \times \) 2 filters for convolution and a 2 \( \times \) 2 filter for max pooling. The flattening layer converts output vectors of the 2D pooling layer into linear arrays for the next layer. Through the next full connected layers including the dropout layer for controlling overfitting, the bottom two nodes output the classification results. The result is the probability that the input SYN picture is malicious or normal.

In this paper, we used ReLU for the activation function in the 2D convolution and 128 units dense layers. We also used Softmax for the last full connected layer. The dropout rate of the dropout layer is 0.5. To implement this neural network, we used Keras [11] that is a python library for deep learning. In addition, all parameters were found and tuned by our heuristics, so that there is still room for improvement.

IV . EXPERIMENT

In our preliminary evaluation, we trained the neural network with our dataset described in Section III-B, and we tested the trained model by inputting other SYN packet data captured in the other darknet. If the SYN pictures certainly represent features of malicious or normal hosts and our model is sufficiently trained, the source IP addresses of the SYN packets captured in the other darknet should be classified as malicious.

A. Training Result

As a result of training, our neural network achieves 98.39% accuracy with the test data after 20 epochs. The half of both types of the SYN pictures was used for training, and another half was used for testing. Figure 4 shows the learning curve of this training. The y-axis indicates the ratio of correct to incorrect for classifying SYN pictures into malicious or normal. The blue solid line indicates accuracy on training data in each epoch, and the orange dashed line indicates accuracy on the test data in each epoch. As shown in Figure 4, the accuracy of classifying the test data that is not used in training achieves over 98%. This means that the model will classify a host into malicious or normal using 100 SYN packets from the host with over 98% accuracy.

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

Percentage that a SYN picture was classified as malicious

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

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

1. Split characters
2. Convert the URL into HEX values
3. Extract 8-bits values by shifting 4 bits in the HEX values

www.iij.ad.jp/index.html

77,77,77,77,77,72,2E, E6,69,96,69,96,6A,A2, E6,6A,A7,70

3F,F6,69,96,6E,E6,64,
E6,69,96,69,96,6A,A2,
46,65,57,78,82,2E,E6,
2E,E6,61,16,64,42,2E,
E6,6A,A7,70

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?

<table>
<thead>
<tr>
<th>256 dimensional sparse vector</th>
<th>256 dimensional sparse vector</th>
<th>512 dimensional sparse vector</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://www.iij.ad.jp">www.iij.ad.jp</a></td>
<td>index.html</td>
<td></td>
</tr>
<tr>
<td>16 → 1  2E → 3</td>
<td>2E → 1  46 → 1</td>
<td></td>
</tr>
<tr>
<td>42 → 1  61 → 1</td>
<td>57 → 1  65 → 1</td>
<td></td>
</tr>
<tr>
<td>64 → 1  69 → 2</td>
<td>68 → 1  6C → 1</td>
<td></td>
</tr>
<tr>
<td>6A → 2  70 → 1</td>
<td>6D → 1  74 → 1</td>
<td></td>
</tr>
<tr>
<td>72 → 1  77 → 5</td>
<td>78 → 1  82 → 1</td>
<td></td>
</tr>
<tr>
<td>96 → 2  A2 → 1</td>
<td>87 → 1  D6 → 1</td>
<td></td>
</tr>
<tr>
<td>A7 → 1  E6 → 3</td>
<td>E6 → 1</td>
<td></td>
</tr>
</tbody>
</table>
Neural Network Topology

URL String

A vector of host part (256 dims)

512 dims

(Linear)

256 dims

(Linear)

256 dims

(Linear)

A vector of path part (256 dims)

Dropout 0.75

Dropout 0.75
Making Datasets

Blacklist 1
26,722 URLs
(before 2017-04-25)

Blacklist 2
68,172 URLs
(before 2017-10-03)

Graylist
142,749,999 URLs
(on 2017-04-25)

Exclude

Sample

Blacklist
26,722 URLs

Whitelist
26,722 URLs
## Datasets

### TABLE I. URL DATASETS FOR TRAINING

<table>
<thead>
<tr>
<th>Type</th>
<th>Content</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blacklist 1</td>
<td>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.</td>
<td>26,722</td>
</tr>
<tr>
<td>Blacklist 2</td>
<td>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.</td>
<td>68,172</td>
</tr>
<tr>
<td>Whitelist 1</td>
<td>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.</td>
<td>26,722</td>
</tr>
</tbody>
</table>

Keiichi Shima, Daisuke Miyamoto, Hiroshi Abe, Tomohiro Ishihara, Kazuya Okada, Yuji Sekiya, Hirochika 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

<table>
<thead>
<tr>
<th></th>
<th>Optimizer</th>
<th>Accuracy (%)</th>
<th>Training time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our method</td>
<td>Adam</td>
<td>94.18</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>AdaDelta</td>
<td>93.54</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>SGD</td>
<td>88.29</td>
<td>31</td>
</tr>
<tr>
<td>eXpose[6]</td>
<td>Adam</td>
<td>90.52</td>
<td>119</td>
</tr>
<tr>
<td></td>
<td>AdaDelta</td>
<td>91.31</td>
<td>119</td>
</tr>
<tr>
<td></td>
<td>SGD</td>
<td>77.99</td>
<td>116</td>
</tr>
</tbody>
</table>

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

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


- 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


Internship Program