Toward automation of cybersecurity operations using machine learning techniques

Takeshi Takahashi, Ph.D., CISSP, PMP

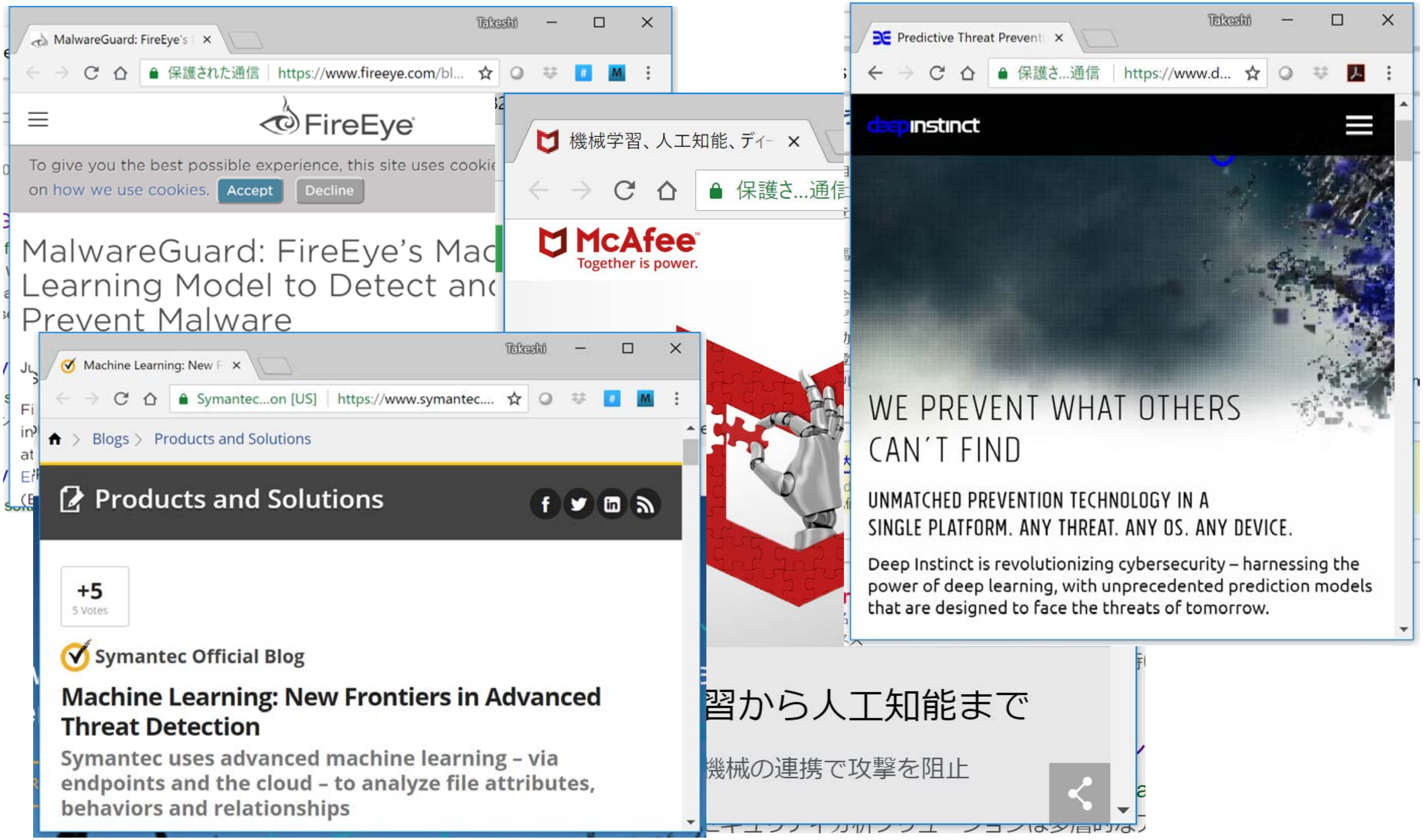
Research Manager
NICT

Agenda

1. Recent trend of AI-related researches in cybersecurity domain

2. Our research activities in a nutshell

AI techniques are already indispensable



To give you the best possible experience, this site uses cookie

on how we use cookies. Accept Decline

MalwareGuard: FireEye's Mac Learning Model to Detect and Prevent Malware

習から人工知能まで

機械の連携で攻撃を阻止

WE PREVENT WHAT OTHERS CAN'T FIND

UNMATCHED PREVENTION TECHNOLOGY IN A SINGLE PLATFORM. ANY THREAT. ANY OS. ANY DEVICE.

Deep Instinct is revolutionizing cybersecurity – harnessing the power of deep learning, with unprecedented prediction models that are designed to face the threats of tomorrow.

Machine Learning: New F
Takeshi
Symantec...on [US] | https://www.symantec....
Blogs > Products and Solutions

Products and Solutions

+5
5 Votes

Symantec Official Blog

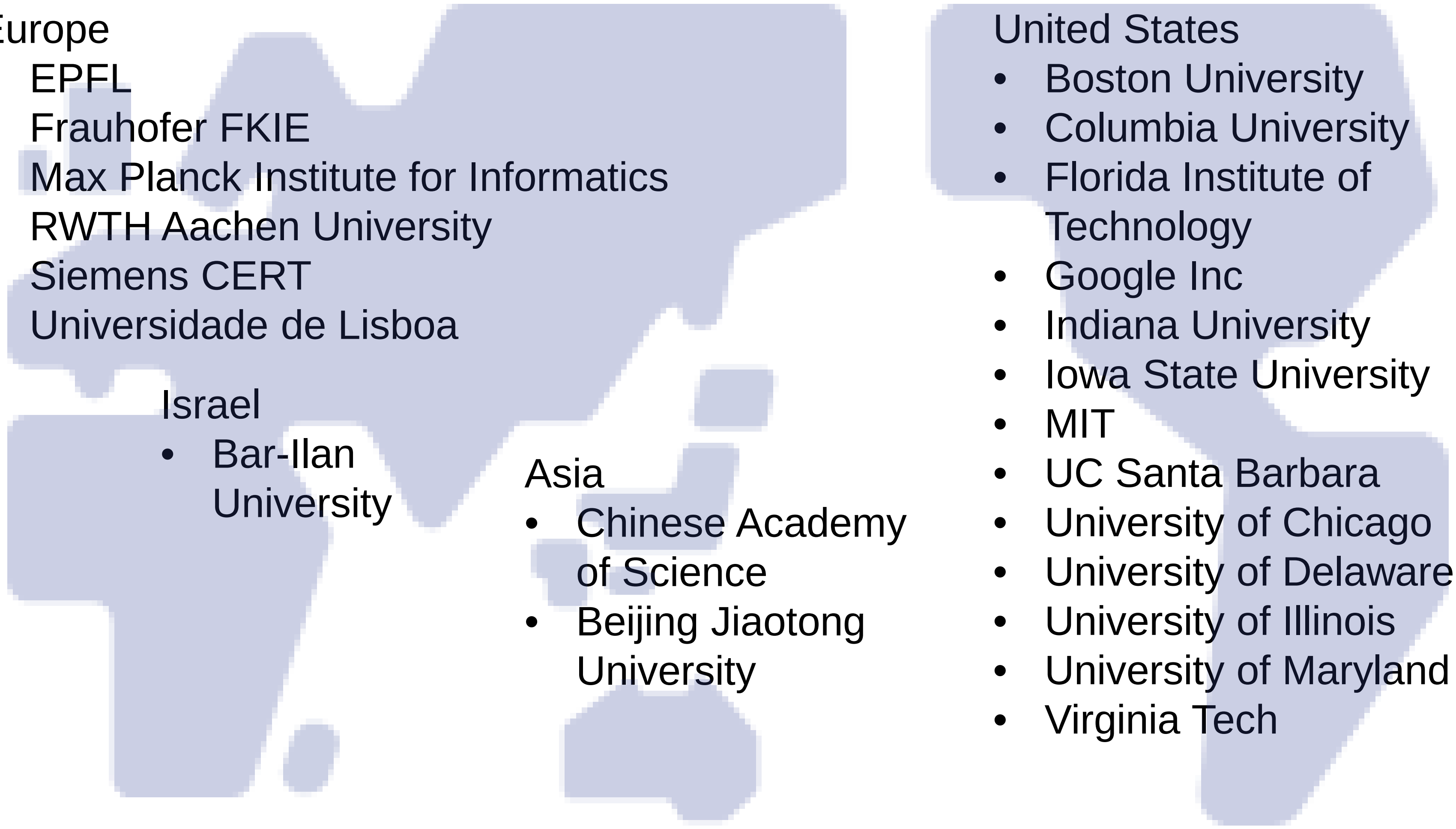
Machine Learning: New Frontiers in Advanced Threat Detection

Symantec uses advanced machine learning – via endpoints and the cloud – to analyze file attributes, behaviors and relationships

Anti-virus vendors claim that they use deep learning techniques, though the details were not usually disclosed.

AI-related issues have been actively studied

Authors of AI-related papers in USENIX Security 2018



Europe

- EPFL
- Frauhofer FKIE
- Max Planck Institute for Informatics
- RWTH Aachen University
- Siemens CERT
- Universidade de Lisboa

Israel

- Bar-Ilan University

Asia

- Chinese Academy of Science
- Beijing Jiaotong University

United States

- Boston University
- Columbia University
- Florida Institute of Technology
- Google Inc
- Indiana University
- Iowa State University
- MIT
- UC Santa Barbara
- University of Chicago
- University of Delaware
- University of Illinois
- University of Maryland
- Virginia Tech

AI-related issues have been actively studied

Authors of AI-related papers in CSS 2018

Europe

- Lancaster University
- University College London

Asia

- Inha University
- Peking University
- Zhejiang University
- The Hong Kong Polytechnic University
- Chinese Academy of Sciences
- Hanyang University
- National University of Singapore

United States

- University of Central Florida
- Florida International University
- Northwest University
- Lehigh University
- The Pennsylvania State University
- Virginia Tech
- University of Pennsylvania
- Symantec
- UC Riverside
- UC Berkeley
- University of Illinois at Urbana-Champaign
- University of Massachusetts

More AI-related topics have been explored

A few example topics on ML researches

Traffic anomaly detection & malware detection (long standing area)

- Explainable system
- Performance improvements /real-time operations

Attacks on computing systems

- Solving captcha
- Malfunctioning voice recognition systems

Deanonymization (attacks against privacy)

- Code Authorship Identification
- Document author attribute classification
- Identification of account pertaining review comments

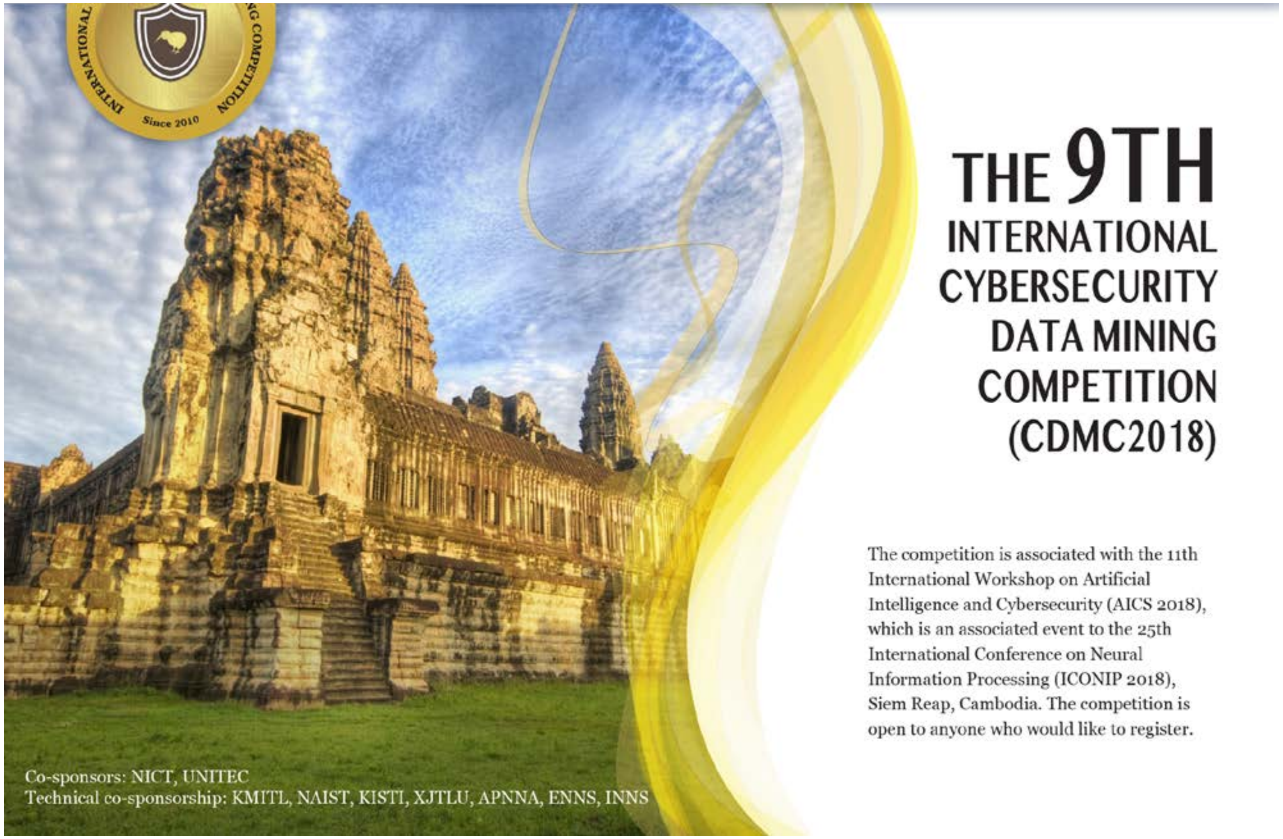
Proactive defense techniques

- Program debloating (minimize vulnerabilities)
- Watermarking DNN
- Event prediction

Vulnerabilities of ML

- Poisoning attacks
- Vulnerabilities of transfer learning
- Attribute inference attacks
- Model reuse attack

We worked on AI x cybersec. for more than a decade


INTERNATIONAL
G COMPETITION
Since 2010

THE 9TH
INTERNATIONAL
CYBERSECURITY
DATA MINING
COMPETITION
(CDMC2018)

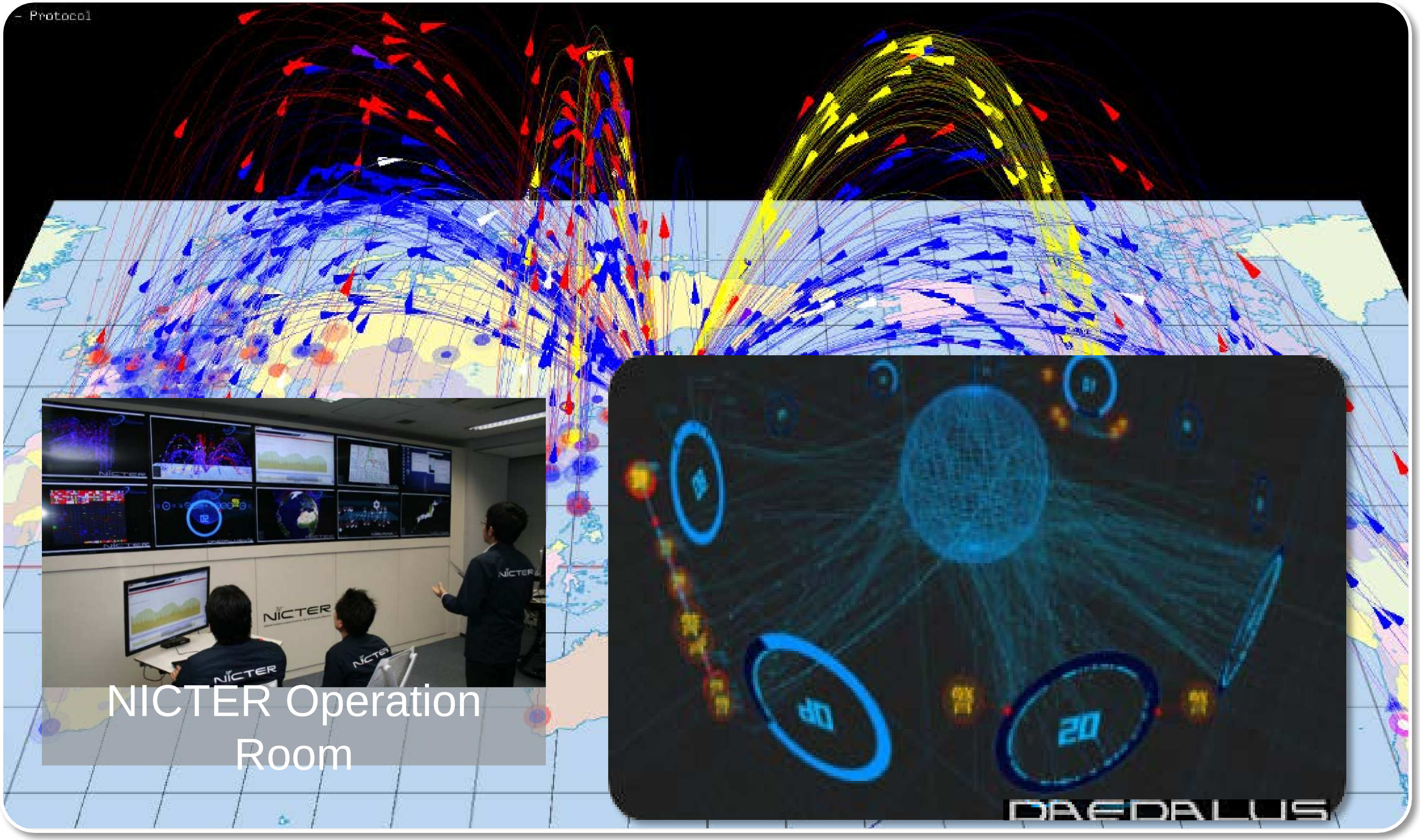
The competition is associated with the 11th International Workshop on Artificial Intelligence and Cybersecurity (AICS 2018), which is an associated event to the 25th International Conference on Neural Information Processing (ICONIP 2018), Siem Reap, Cambodia. The competition is open to anyone who would like to register.

Co-sponsors: NICT, UNITEC
Technical co-sponsorship: KMITL, NAIST, KISTI, XJTLU, APNNA, ENNS, INNS

- 11th International Data Mining and Cybersecurity Workshop (DMC), 2018
- 9th International Cybersecurity Data Mining Competition (CDMC), 2018

Our network monitoring systems accumulates data

- ✓ We monitor large-scale darknet spaces
- ✓ We built and have been operating systems, e.g., NICTER and DAEDALUS



Our dataset

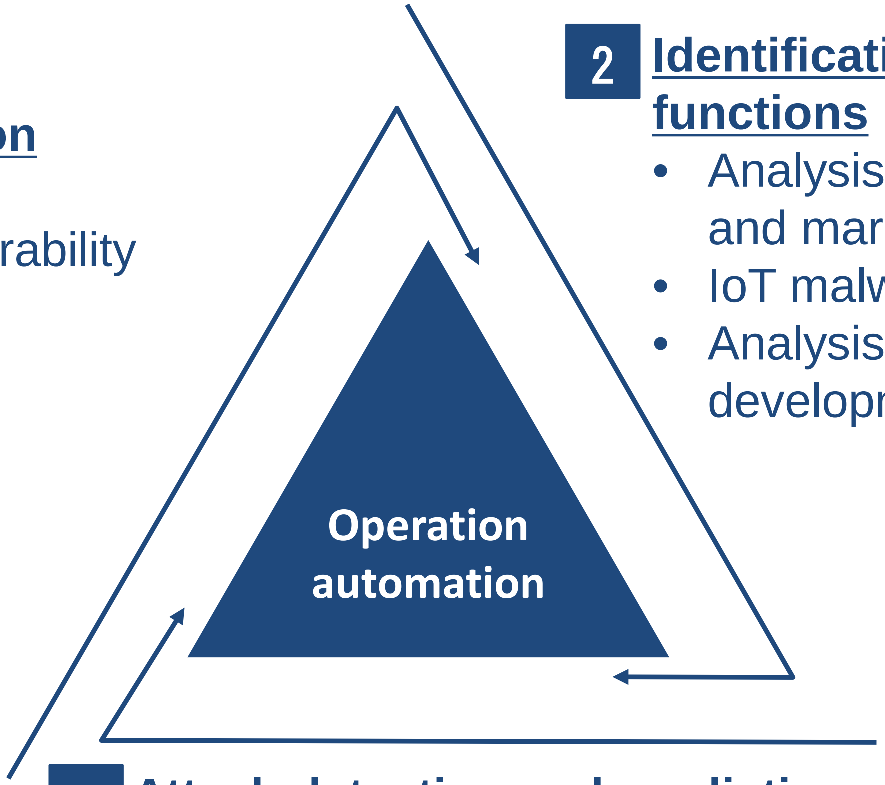
Category	Examples of accumulated data
Darknet related data	Data on the traffic sent to unused IP address spaces. This includes pcap files, statistical information, and malicious host information.
Livenet related data	Traffic data within NICT. This includes pcap files, flow data, security alerts generated by security appliances.
Malware related data	Malware samples, static and dynamic analysis results, etc.
Spam related data	Spam (double bounce) mail data, statistical information, etc.
Android related data	APK files and applications' metadata, e.g., category and description of applications
Blogs and articles	Tweets, security vendor blogs, etc.
Web crawler	URL list, Web contents, their evaluation results, etc.
Honeypot data	Data from High-interaction/low-interaction honey pots and high-interaction/low-interaction client honey pots
Commercial Intelligence data	Information on the sites hosting malware, bot, C&C server list, domain history, malware samples, threat reports, etc. purchased from VirusTotal, SecureWorks, Anubis, DomainTools, Malnet, Team 5, etc.

Agenda

1. Recent trend of AI-related researches in cybersecurity domain
2. Our research activities in a nutshell

Our research focus

We conduct R&D on AI techniques that analyze and understand security situation and automate security operations within an organization.



1

Priority determination

- Alert screening
- Evaluation of vulnerability severity

2 **Identification of malware functions**

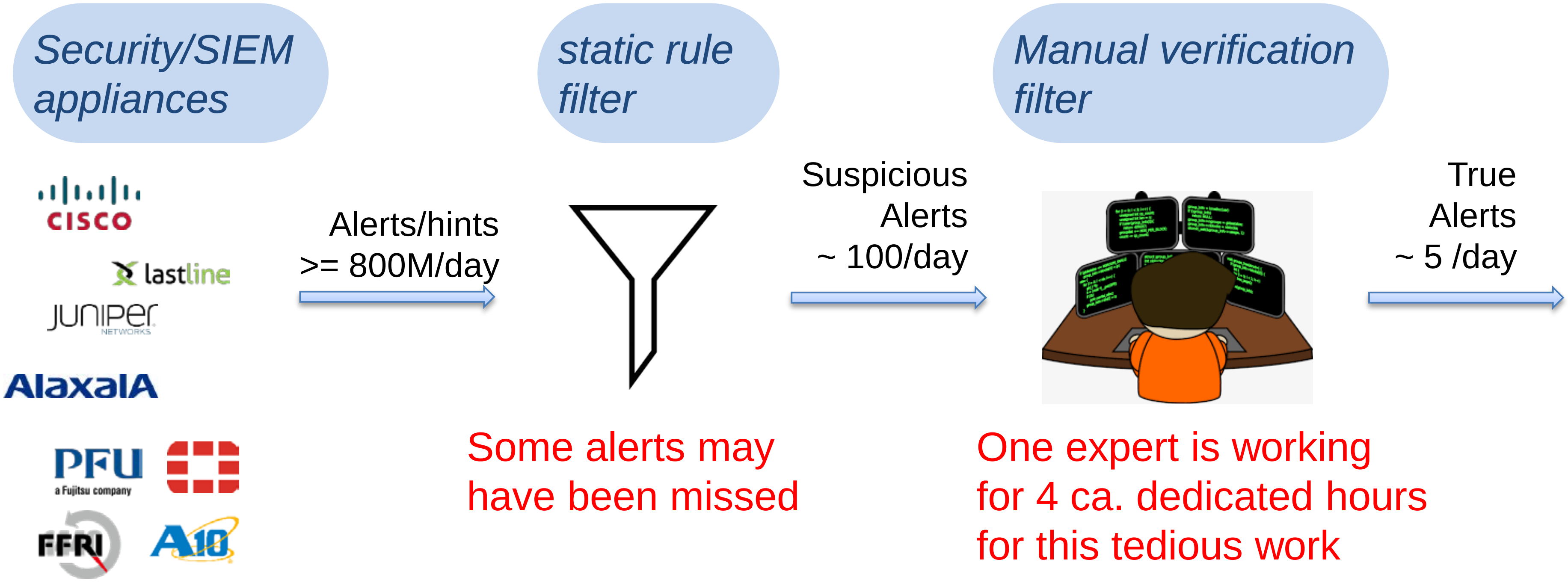
- Analysis of Android apps and markets
- IoT malware analysis
- Analysis automation tool development

3 **Attack detection and prediction**

- Darknet analysis
- Threat estimation and prediction
- Encrypted traffic analysis

Featured topic 1: alert screening and prioritization

Current process for identifying important security alert



Security/SIEM appliances

static rule filter

Manual verification filter

Alerts/hints
>= 800M/day

Suspicious
Alerts
~ 100/day

True
Alerts
~ 5 /day

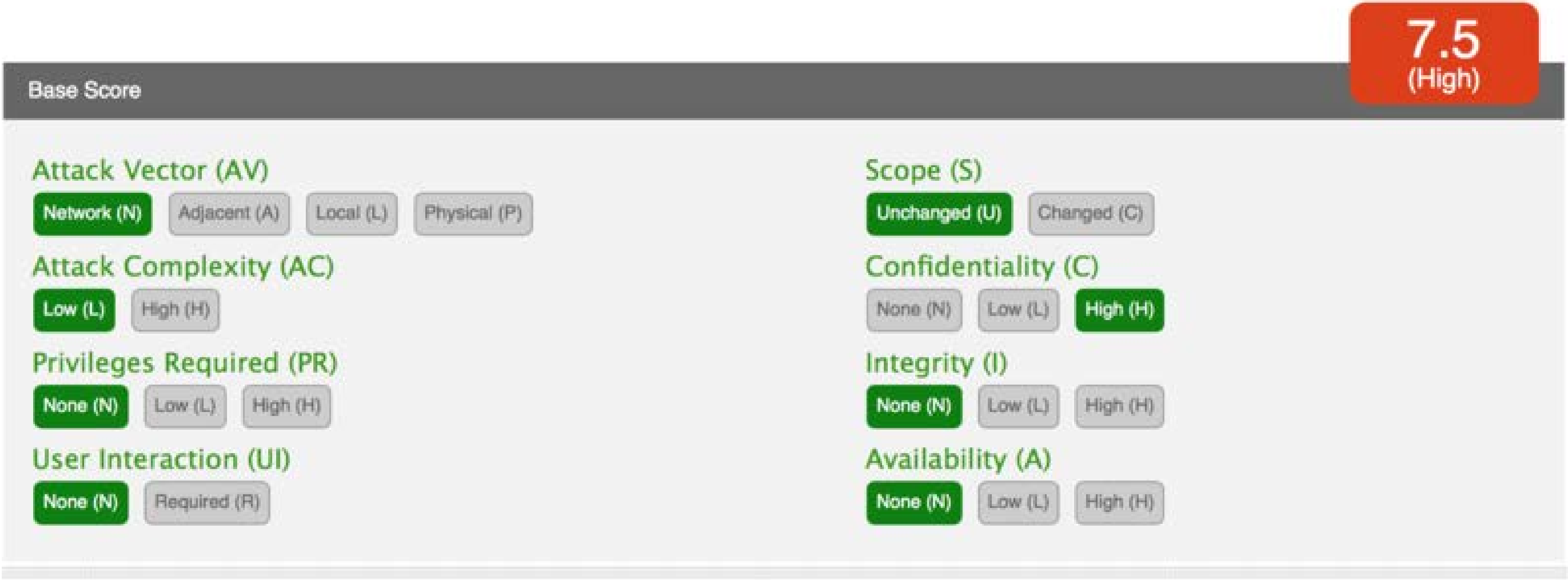
Some alerts may have been missed

One expert is working for 4 ca. dedicated hours for this tedious work

We ***replace and streamline*** the above 2-stage filtering process (static rule + manual verification) ***with machine learning techniques.***

Featured topic 2: vulnerability severity evaluation

1. CVSS base score provides the technical severity of vulnerabilities based on the value of eight metrics.
2. Currently, a registrant of a vulnerability note selects one of predefined values for each of the metrics to derive the score.
3. We use supervised machine learning techniques to select the values based on several features, including vulnerability descriptions.



7.5 (High)

Base Score

Attack Vector (AV)
[✓ Network (N)] Adjacent (A) Local (L) Physical (P)

Attack Complexity (AC)
[✓ Low (L)] High (H)

Privileges Required (PR)
[✓ None (N)] Low (L) High (H)

User Interaction (UI)
[✓ None (N)] Required (R)

Scope (S)
[✓ Unchanged (U)] Changed (C)

Confidentiality (C)
None (N) Low (L) [✓ High (H)]

Integrity (I)
[✓ None (N)] Low (L) High (H)

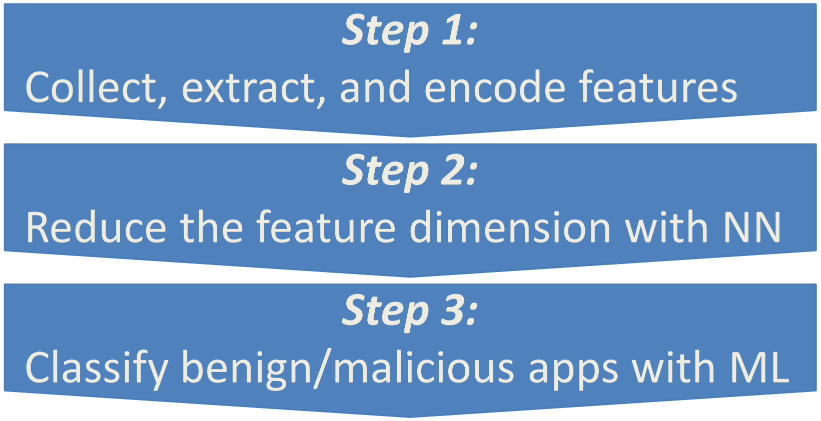
Availability (A)
[✓ None (N)] Low (L) High (H)

Source: T.Takahashi et al., "Toward Automated Vulnerability Handling," CARIS2, ISOC, 2019.

Featured topic 3: android application vetting

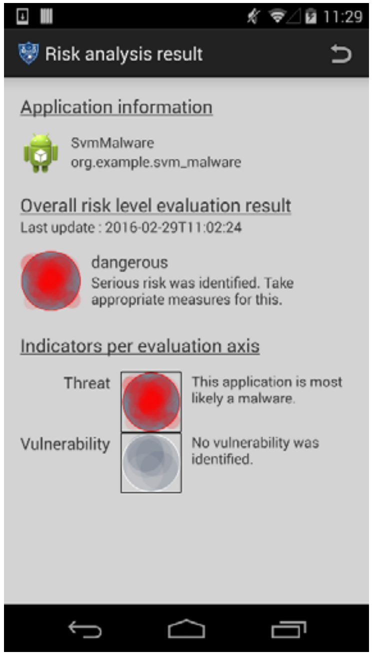
- *We detect malware using machine-learning (ML) and neural network (NN) techniques* (Accuracy ≒99.79%)
  - *Input features: permission requests, API calls, app categories, clusters(generated from app descriptions)*
  - *Step 2 drastically reduces the computational cost*

- Some analysis have been conducted
  - Performance without step 2 was around 94-95% by using SVM-RFE
  - *Influential features (analyzed by SVM-RFE):* API calls, some permission requests and application categories


Step 1:
Collect, extract, and encode features

Step 2:
Reduce the feature dimension with NN

Step 3:
Classify benign/malicious apps with ML


11:29

Risk analysis result

Application information

SvmMalware
org.example.svm_malware

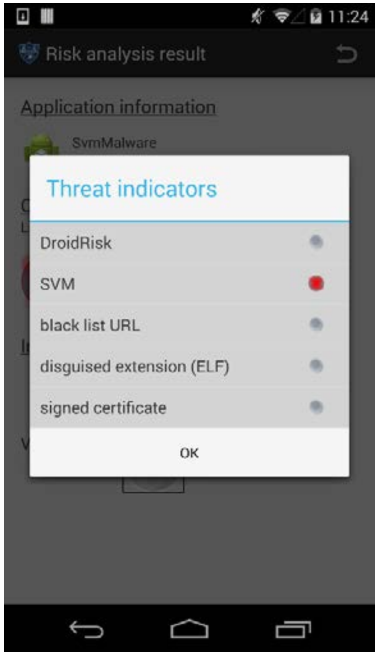
Overall risk level evaluation result
Last update : 2016-02-29T11:02:24

dangerous
Serious risk was identified. Take appropriate measures for this.

Indicators per evaluation axis

Threat — This application is most likely a malware.

Vulnerability — No vulnerability was identified.


11:24

Risk analysis result

Application information

SvmMalware

Threat indicators

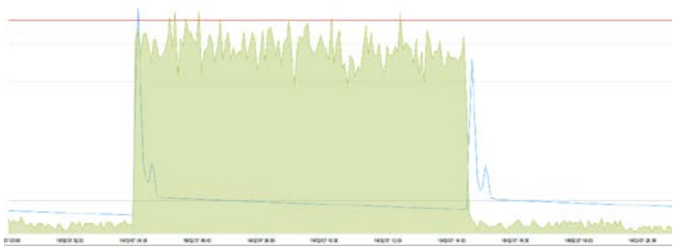
DroidRisk	●
SVM	●
black list URL	●
disguised extension (ELF)	●
signed certificate	●

OK

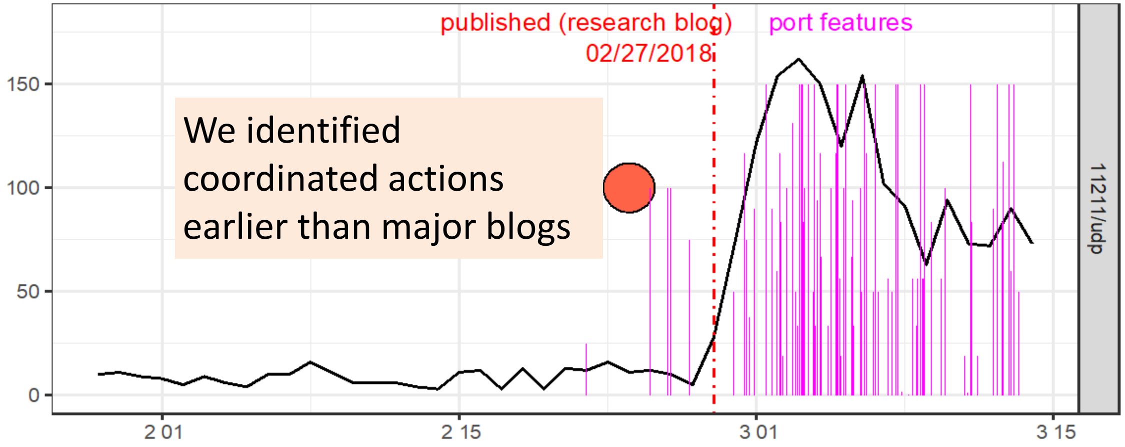
Sources: B.Sun et al., "A Scalable and Accurate Feature Representation Method for Identifying Malicious Mobile Applications," ACM SAC, 2019.
T.Takahashi et al., "Android Application Analysis using Machine Learning Techniques," Intelligent Systems Reference Library, 181 - 205, 2019.

Featured topic 4: detecting coordinated activities

Objective	We identify coordinated activities of hosts
Requirements	• Realtime detection • Minimizing false positive/negative
Approaches	• We analyze scans arriving at our darknet because bots are often coordinated by C2 server • We analyze darknet traffic with unsupervised learning techniques (glasso, NMF, and tensor decomposition) to identify coordinated scans • These techniques are tunes to run in real time



Coordinated scans
(x: date/time, y: number of sources)



We identified coordinated actions earlier than major blogs

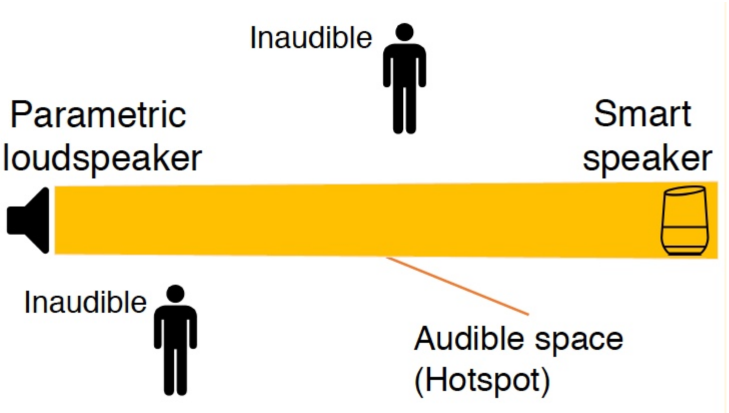
A sample case of a coordinated scan detection
(x: date/time, y: number of sources)

Source: H.Kanehara et al., "Real-Time Botnet Detection Using Nonnegative Tucker Decomposition," ACM SAC, 2019.

Featured topic 5: voice command security

Audio Hotspot Attack

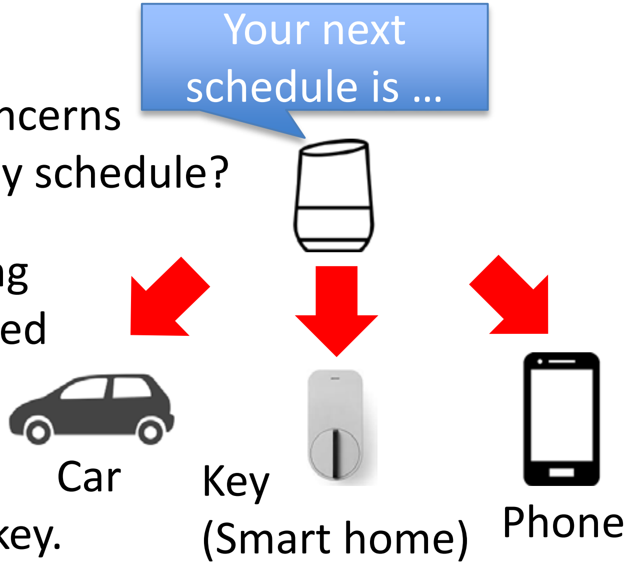
- A voice assistance system can be manipulated by illegitimate attacker without being noticed by anybody else
- We inject malicious voice commands using directional sound beams.
- Parametric loudspeaker can generate directional sound beams.



1. Privacy concerns
  ex) What’s my schedule?

2. manipulating other connected devices

ex) Open the key.
   Call to [someone]



Countermeasure

We made a new classifier that detects various voice attacks using 2D convolutional neural network (2DCNN).

Source: R.Iijima et al., "Audio Hotspot Attack: An Attack on Voice Assistance Systems Using Directional Sound Beams," ACM CCS poster, 2018.

Related publications in recent years

1. H.Kanehara, Y.Murakami, J.Shimamura, T.Takahashi, D.Inoue, N.Murata, "Real-Time Botnet Detection Using Nonnegative Tucker Decomposition," ACM SAC, 2019.
2. B.Sun, T.Ban, S.Chang, Y.Sun, T.Takahashi, D.Inoue, "A Scalable and Accurate Feature Representation Method for Identifying Malicious Mobile Applications," ACM SAC, 2019.
3. T.Takahashi, H.Kanehara, M.Kubo, N.Murata, D.Inoue, “Toward Automated Vulnerability Handling,“ CARIS2, 2019
4. T.Takahashi, T.Ban, "Android Application Analysis using Machine Learning Techniques," Intelligent Systems Reference Library, 181 - 205, 2019.
5. S.Chang, Y.Sun, W.Chuang, M.Chen, B.Sun, T.Takahashi, "ANTSdroid:Using RasMMA Algorithm to Generate Malware Behavior Characteristics of Android Malware Family," IEEE PRDC, 2018.
6. L.Zhu, T.Ban, T.Takahashi, D.Inoue, "Employ Decision Value for Binary Soft Classifier Evaluation with Crispy Reference," ICONIP, 2018.
7. R.Iijima, S.Minami, Z.Yunao, T.Takehisa, T.Takahashi, Y.Oikawa, T.Mori, "Audio Hotspot Attack: An Attack on Voice Assistance Systems Using Directional Sound Beams," ACM CCS poster, 2018.
8. T.Takahashi, B.Panta, Y.Kadobayashi, K.Nakao, "Web of cybersecurity: Linking, locating, and discovering structured cybersecurity information," Int J Commun Syst. 2017.