Data Driven Cybersecurity Research in NICT

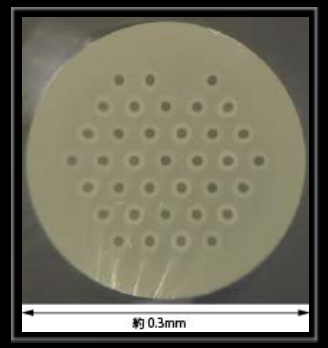
Daisuke INOUE
Cybersecurity Laboratory
Cybersecurity Research Institute
National Institute of Information and Communications Technology (NICT)

NICT - the sole national research institute in the field of ICT in Japan -

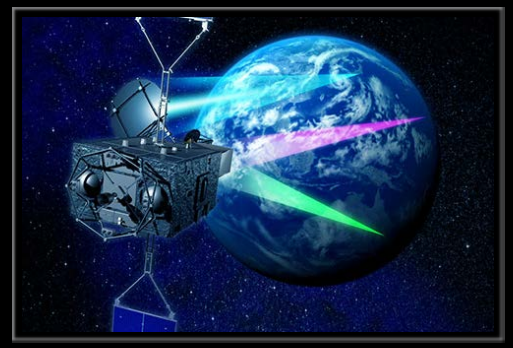
- ICT for sustainable world and human happiness
- Promoting its own research and development
- Cooperating with and supporting industry and academia

Research Topics in NICT

Japan Standard Time (JST)
（Leap second on Jan 1, 2017）



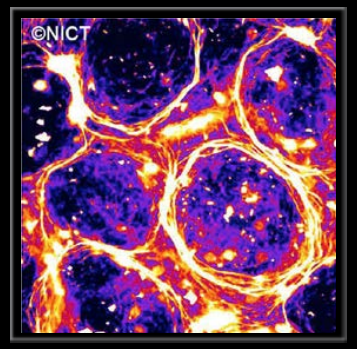
Optical Communication
(Peta bps class multi-core fiber)



Satellite Communication
（Internet Satellite WINDS）

Science Cloud
（Reai-time Web of Himawari-8）

Remote Sensing
（Pi-SAR2 image after 3.11）



Bio/Nano ICT
(Self-organizing bio molecule)

Brain ICT
(Brain-machine Interface)

Multi-lingual Machine Translation
(VoiceTra)

Ultra Realistic Communication
(Electronic Holography)

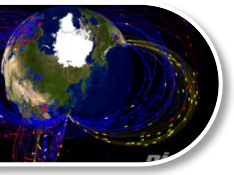
Cybersecurity
(DAEDALUS)

Cybersecurity Research is ...

Data Driven Research

- ✓real-time data collection
- ✓real-time data analytics
- ✓no data no cybersecurity research

Research Map of Cybersecurity Laboratory in NICT



Network Incident analysis Center for Tactical Emergency Response

NICTER

Direct Alert Environment for Darknet And Livenet Unified Security

DAEDALUS

Passive

NICTER Real-network Visual ANAlyzer KAI

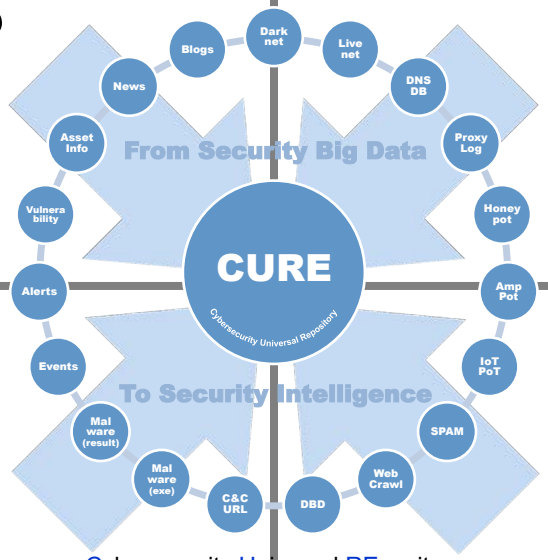
NIRVANA 改

NICTER Real-network Visual ANAlyzer KAI-II

NIRVANA 改弐

Global (indiscriminate attack)

(targeted attack) Local



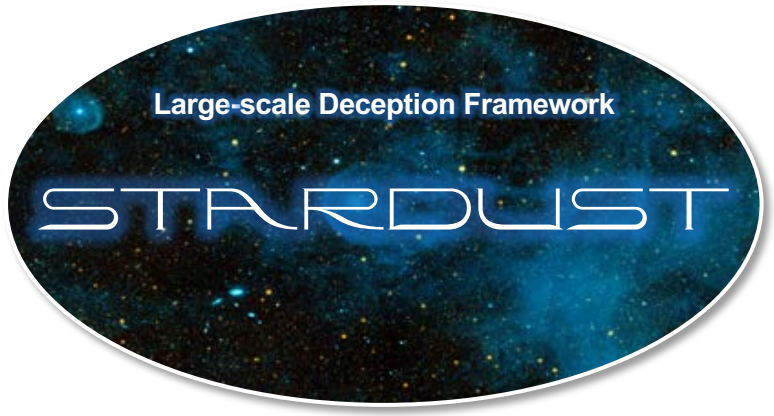
Cybersecurity Universal REpository

Honeypot for Amplification Attack

AmpPOT#

Honeypot for IoT Malware

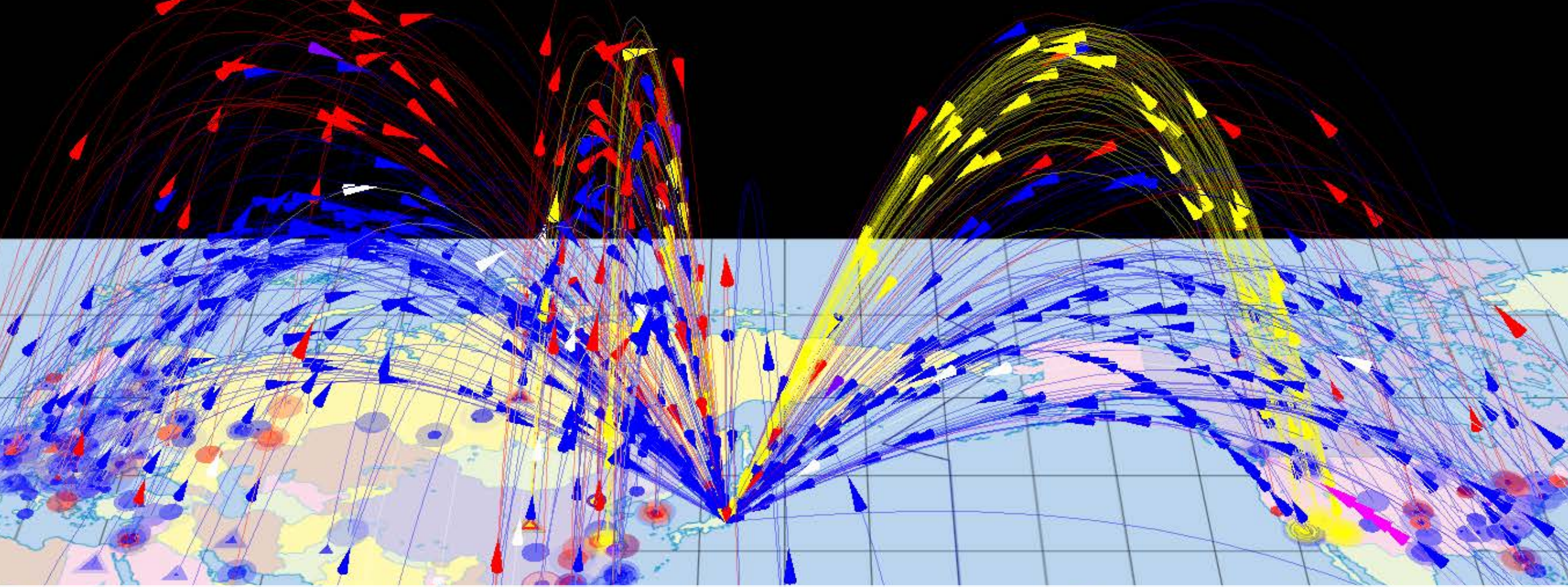
IoTPOT#



Active

Collaborative research with Yokohama National Univ. and Saarland Univ.

Network Incident analysis Center
for Tactical Emergency Response

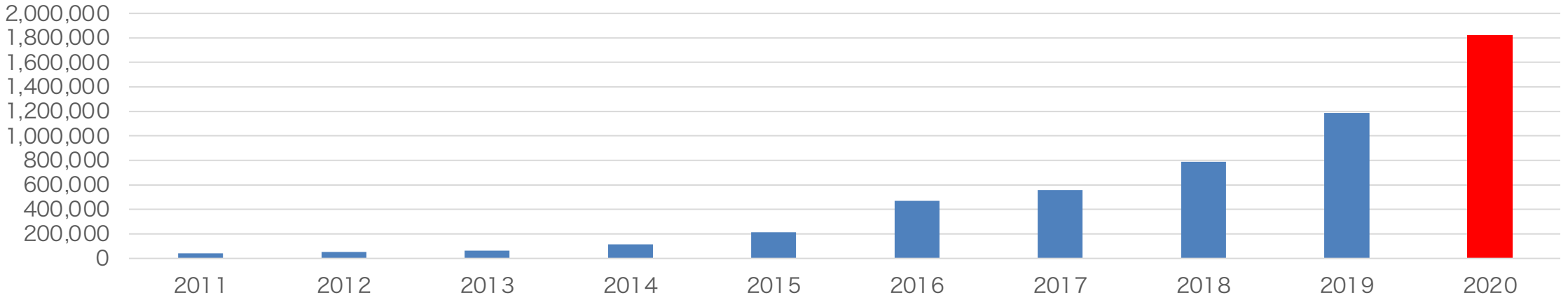


NICTER

- is an **integrated security system** for countering indiscriminate cyberattacks
- based on a large-scale **darknet monitoring**, an automated **malware analysis** and their **correlation**

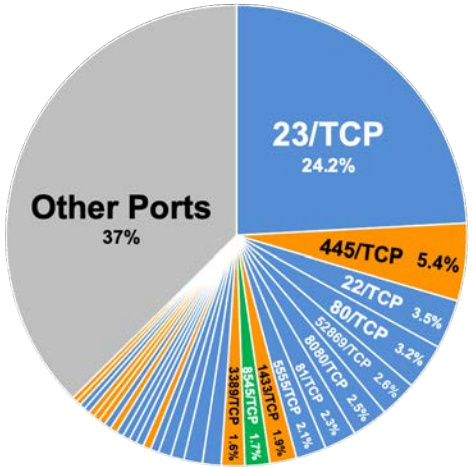
Yearly Stats of Darknet Traffic (Last 10 Years)

Year	Number of packets par year	Number of IP address for darknet	Number of packets par 1 IP address per year
2011	4.54 billion	120 thousands	40,654
2012	7.79 billion	190 thousands	53,085
2013	12.9 billion	210 thousands	63,655
2014	25.7 billion	240 thousands	115,323
2015	54.5 billion	280 thousands	213,523
2016	128.1 billion	300 thousands	469,104
2017	150.4 billion	300 thousands	559,125
2018	212.1 billion	300 thousands	789,876
2019	322.0 billion	300 thousands	1,187,935
2020	**500.1 billion**	**300 thousands**	**1,820,722**

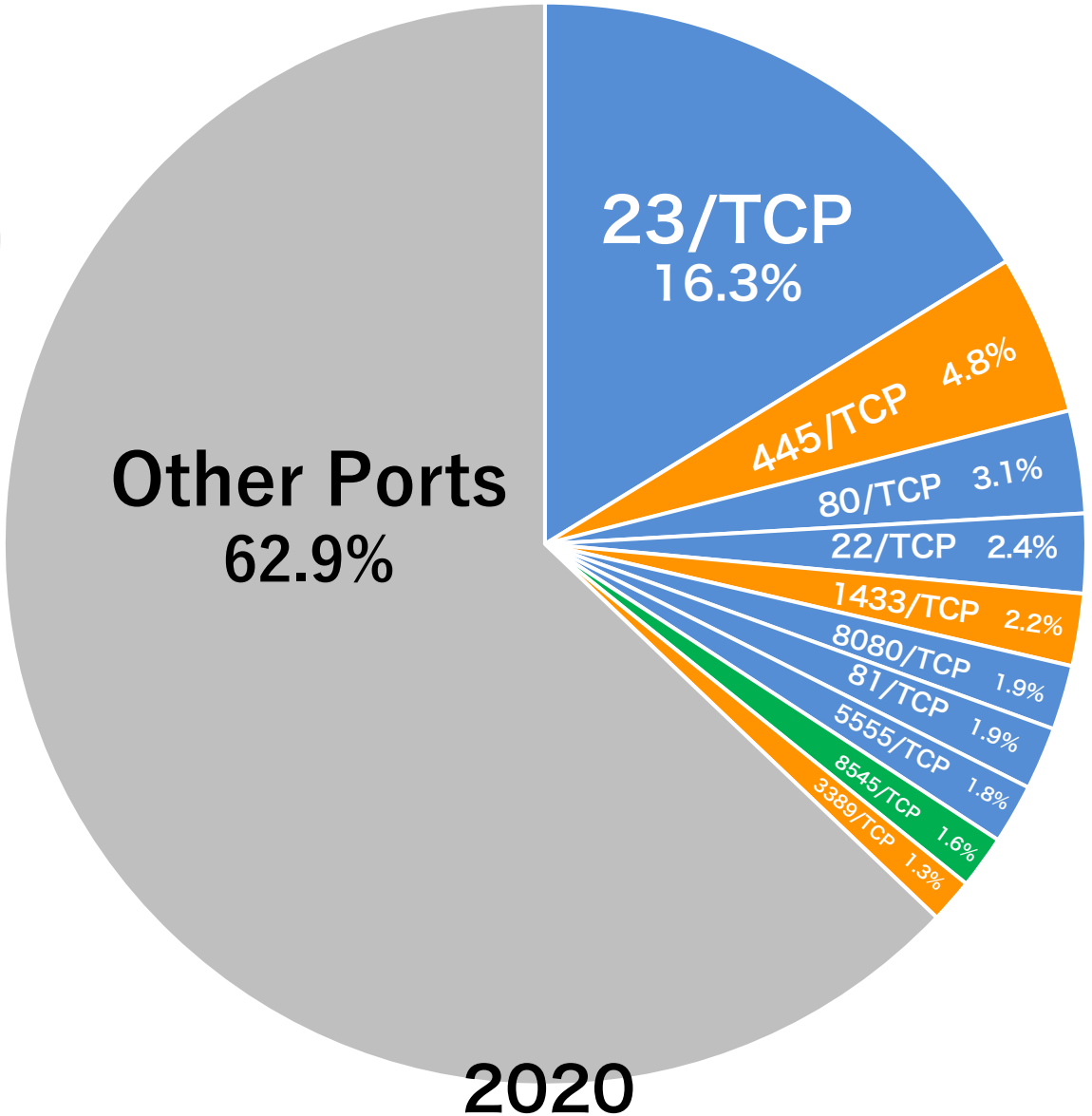


Number of packets per IP address per year

Top 10 Dst Ports observed by NICTER (2020)



2019



2020

Port Number	Target Service
23/TCP	IoT (Web Camera, etc.)
445/TCP	Windows (Server Service)
80/TCP	Web Server (HTTP) IoT (Web Interface)
22/TCP	IoT (Router, etc.)
1433/TCP	Windows (MS-SQL)
8080/TCP	IoT (Web Camera, etc.)
81/TCP	IoT (Home Router, etc.)
5555/TCP	Android (Set Top Box, etc.)
8545/TCP	Ethereum (Cryptocurrency)
3389/TCP	Windows (Remote Desktop)

Practical Use of Darknet Monitoring Results

- **SIGMON (Special Interest Group of Network Monitoring)**
  - ✓ Partners: JPCERT/CC, IPA, @Police, NICT, Universities
  - ✓ Sharing analysis results of darknet traffic (since 2004)

- **ICT-ISAC Japan**
  - ✓ ICT Information Sharing and Analysis Center
  - ✓ Sharing DDoS related information (since 2011)

- **Information Sharing for Tokyo 2020**
  - ✓ Preparation for Tokyo Olympic and Paralympic Games by NISC
  - ✓ Sharing DDoS related information (since 2015)

- **Information Sharing for General Public**
  - ✓ NICTERWEB (http://www.nicter.jp/)
  - ✓ NICTER **Report** (http://www.nict.go.jp/cyber/report.html)
  - ✓ NICTER **Blog** (http://blog.nicter.jp)

NICTERWEB
https://www.nicter.jp/en

National Institute of Information and Communications Technology

NICTER
Network Incident analysis Center for Tactical Emergency Response

NICTER Real-network Visual ANAlyzer KAI

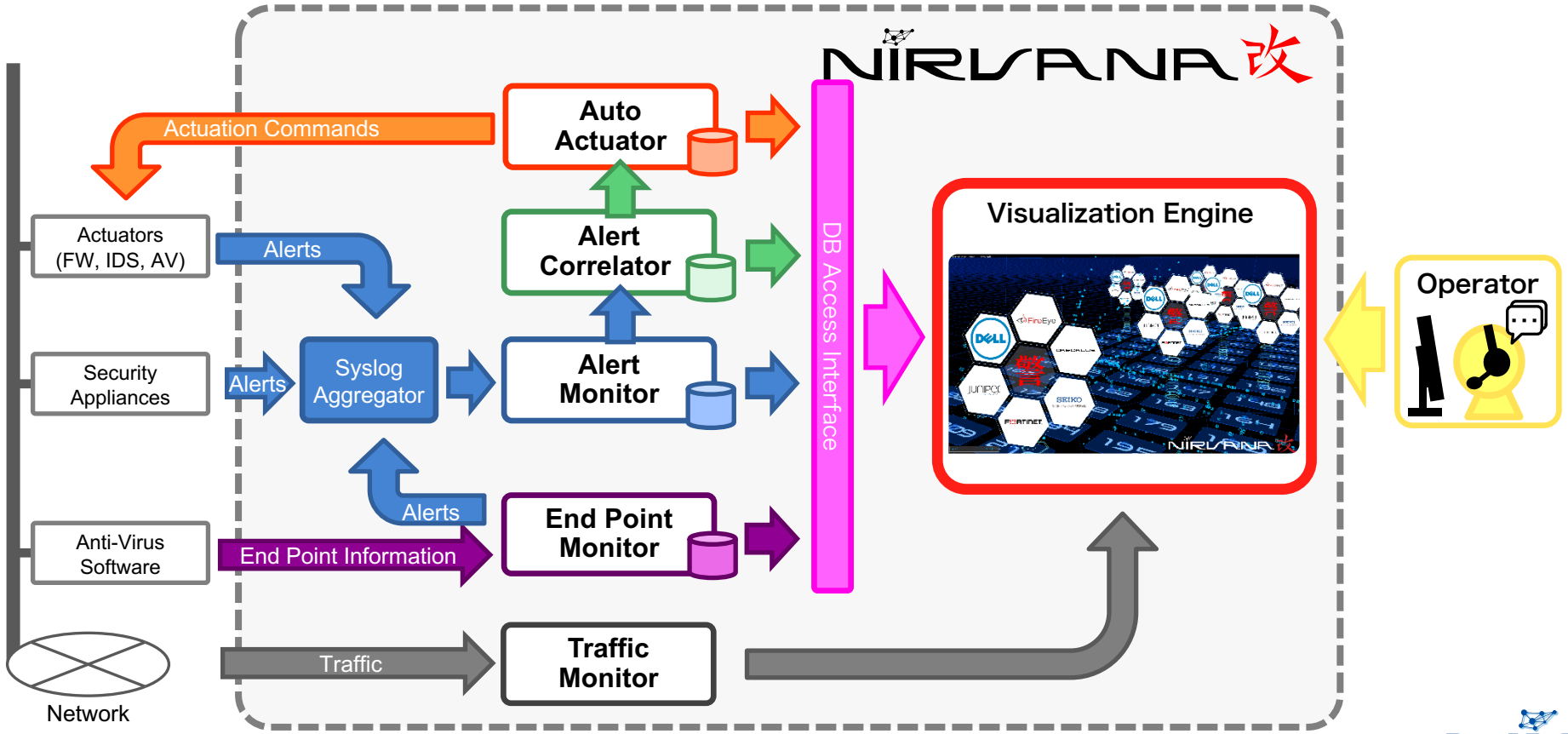
NIRVANA改

- is an **integrated security platform** against APT
- collects security alerts from many types of **security appliances** and **end point security software**

NIRVANA改 System Overview

NIRVANA改

= **Traffic Monitoring** + **Alert Aggregation** + **Auto Actuation** + **Visualization**



National Institute of Information and Communications Technology

Network Incident analysis Center for Tactical Emergency Response


SHOWNET
DIVE INTO THE NEXT
Interop Tokyo
13 – 15 JUNE 2018 | MAKUHARI MESSE | JAPAN
25th
HUAWEI
NTT
NEC
SCSK
YAMAHA
axsh
intel
FURUNO SYSTEMS
IoT+5G+V2X

NIRVANA改 2019

- Security Orchestration @Interop Tokyo 2019 -

- **Alert generators: 23 appliances (12 companies)**

Vendor Name	Product Name
NICT	DAEDALUS
	CURE Flow
Future	Vuls
FFRI	yarai
TrendMicro	TippingPoint TPS
	TippingPoint SMS
	Deep Discovery Inspector
	Deep Discovery Analyzer
Check Point	Security Appliances
	Smart-1 525
	SandBlast TE2000X
DAMBALLA	Network Insight

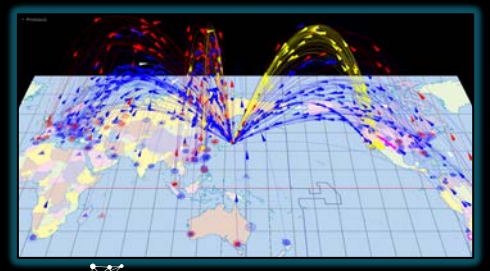
Vendor Name	Product Name
FireEye	NX5500
Fortinet	FortiGate 3601E
	FortiGate 601E
	FortiSandbox 3000E
	FortiDeceptor 1000F
Juniper Networks	JATP400
Lastline	Defender
Palo Alto Networks	PA-5280
	PA-5260
	M-600
A10 Networks	Thunder 3230 CFW

Web-based Attack Response with Practical and Deployable Research InitiatiVE

National Institute of Information and Communications Technology

NICTER
Network Incident analysis Center for Tactical Emergency Response

Indiscriminate Attack

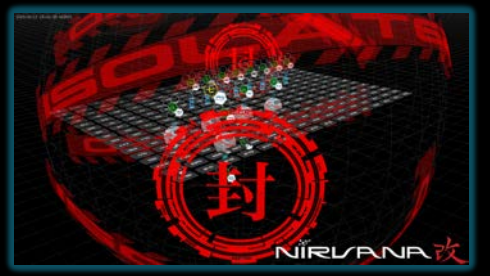


NICTER

DAEDALUS

Web-based Attack

WARPDRIVE



NIRVANA改

Targeted Attack

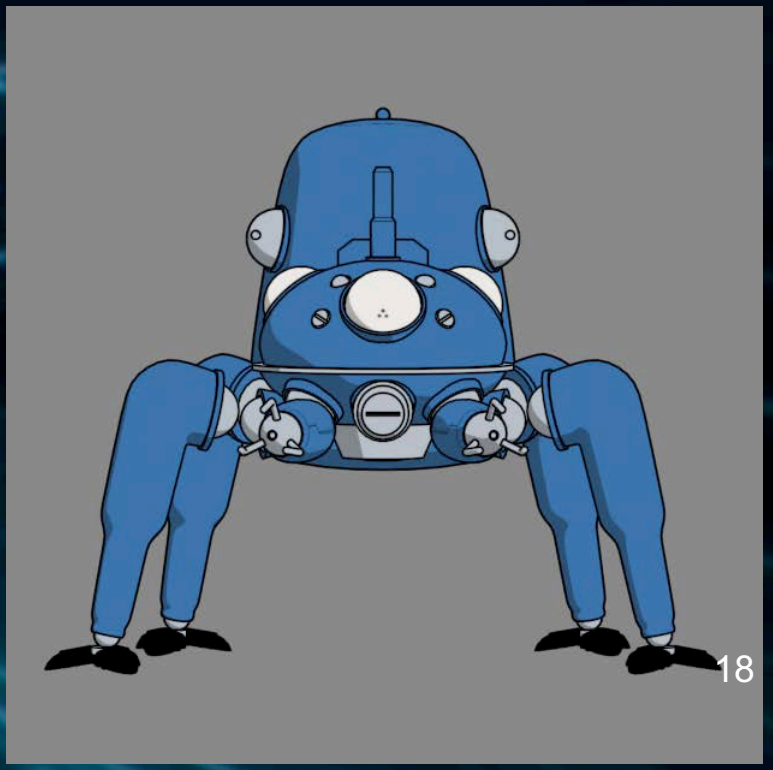
WarpDrive

Web-based **A**ttack **R**esponse with **P**ractical and **D**eployable **R**esearch **I**nitiati**VE**

A **Tachikoma** is a fictional walker with artificial intelligence (AI) from the Ghost in the Shell universe, appearing in the manga (created by Masamune Shirow) and in the Stand Alone Complex sub-universe. Nine of them are initially deployed to Section 9. They are spider-like, multi-legged combat vehicles, and are equipped with adaptive artificial intelligence. (Wikipedia, Jun 18, 2018)

WarpDrive project makes Tachikoma as...

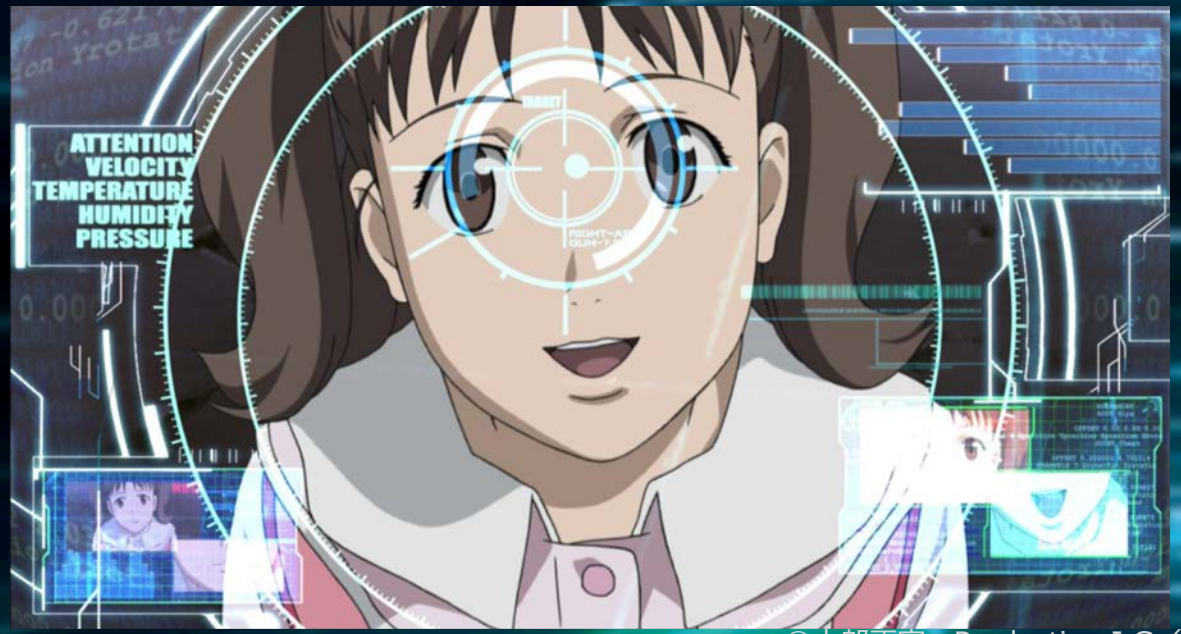
1. **Sensor** in the browser
2. **Actuator** to block web-based attacks
3. **Communicator** with users



WarpDrive

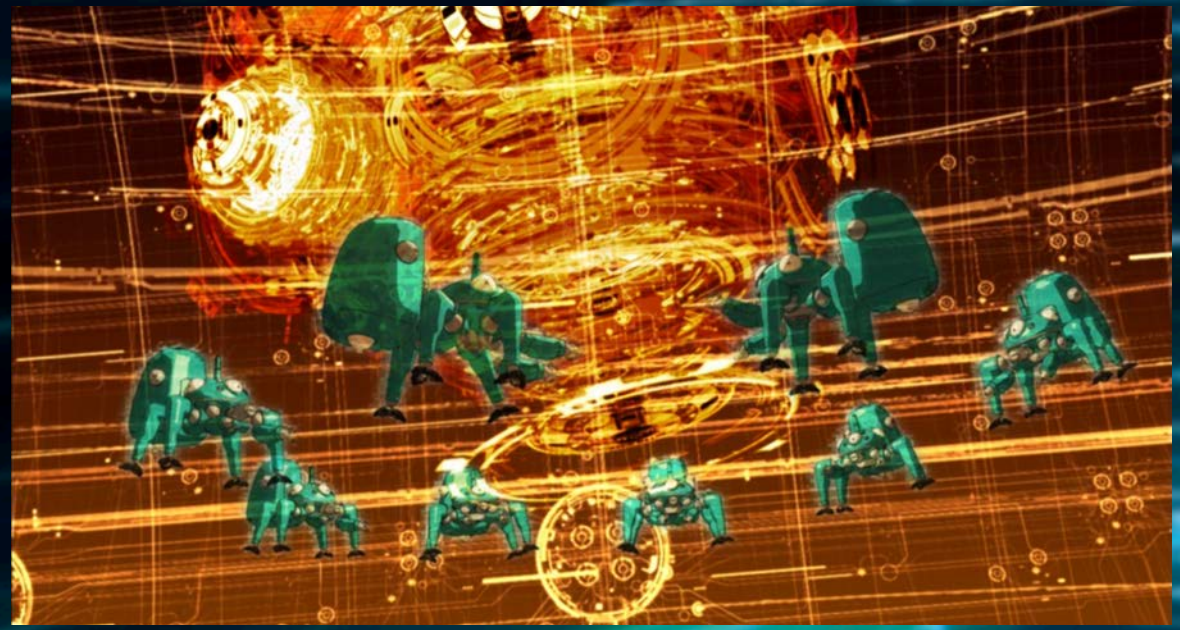
Web-based Attack Response with Practical and Deployable Research InitiatiVE

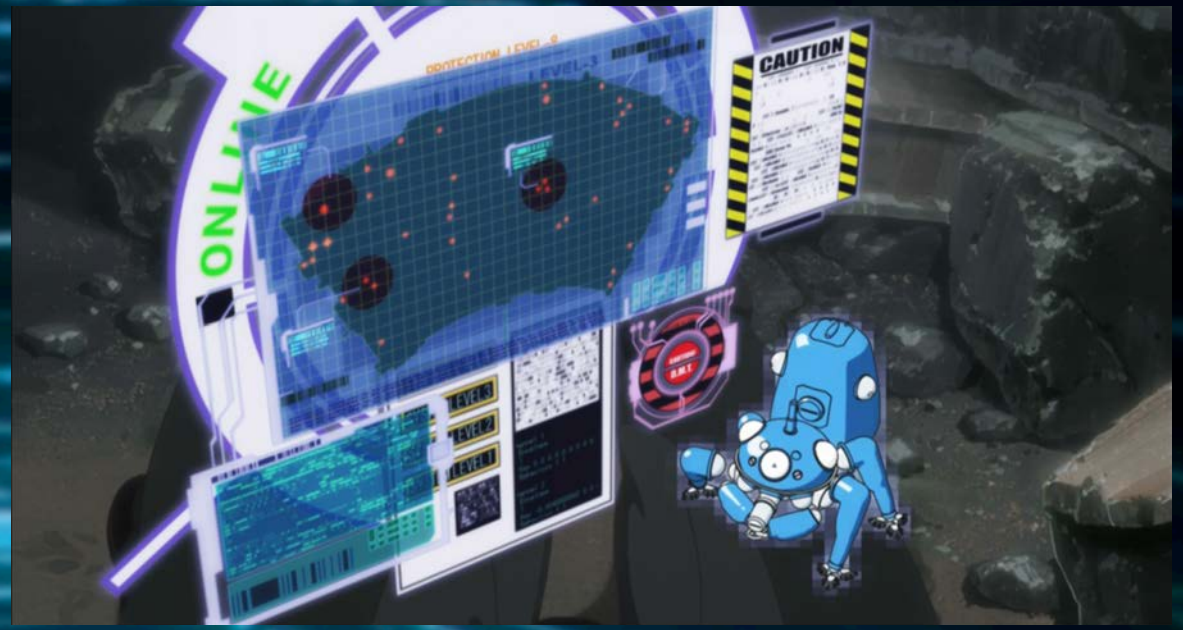
1. Install “Tachikoma Security Agent” into user’s browser (Chrome in Windows or Mac)



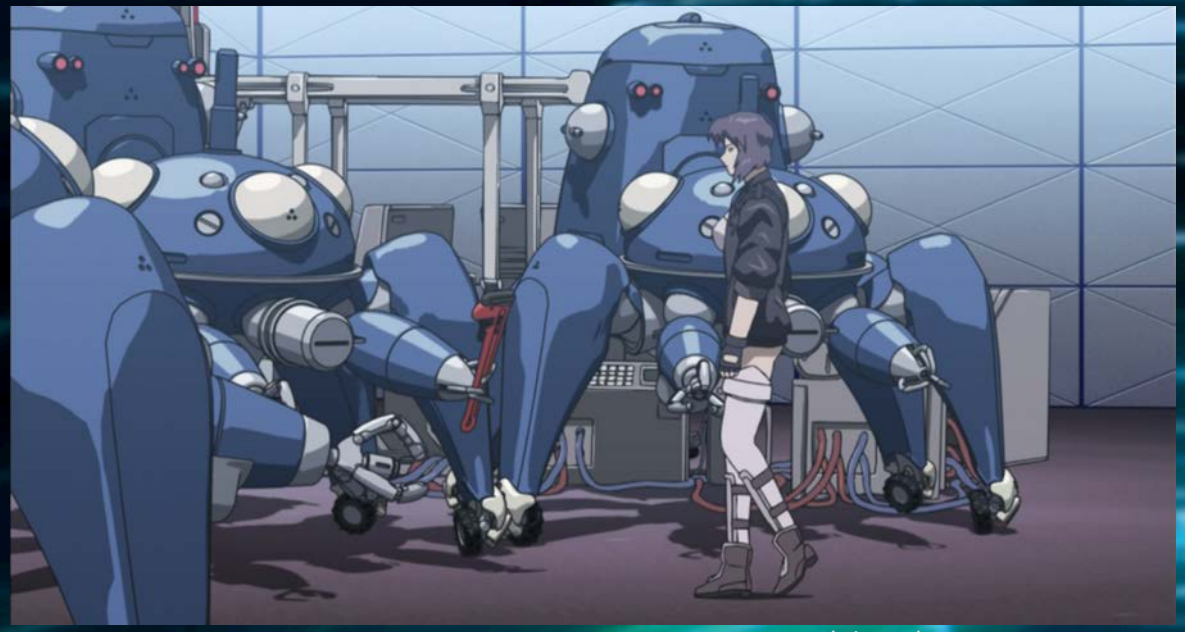
2. Tachikoma SA collects user's web accesses in real-time manner

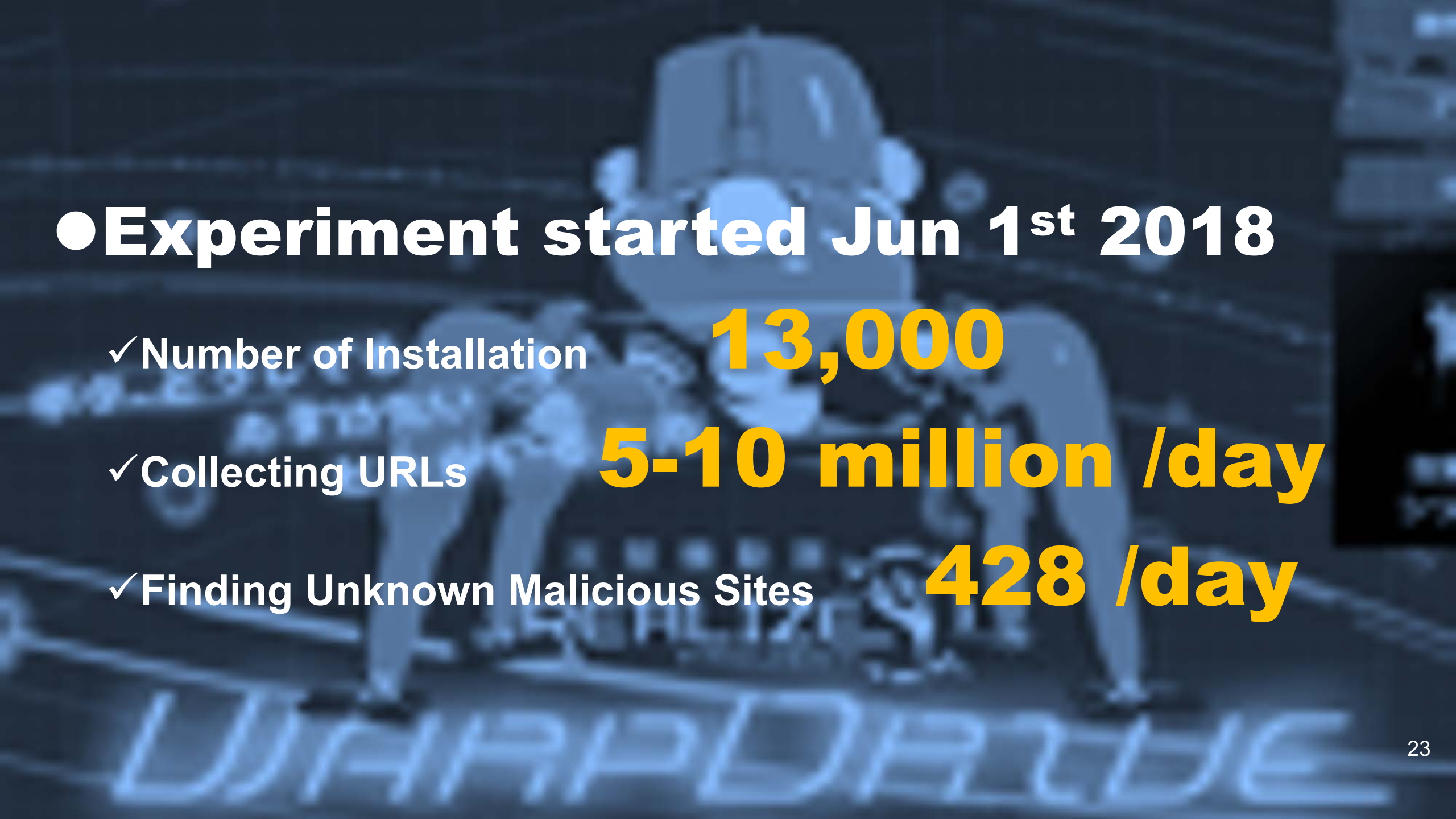
3. Tachikoma SA prevents and alerts user's access to malicious Web sites





4. Tachikoma SA has expanded the coverage to smart phones (Android)





- **Experiment started Jun 1st 2018**
  - ✓Number of Installation **13,000**
  - ✓Collecting URLs **5-10 million /day**
  - ✓Finding Unknown Malicious Sites **428 /day**

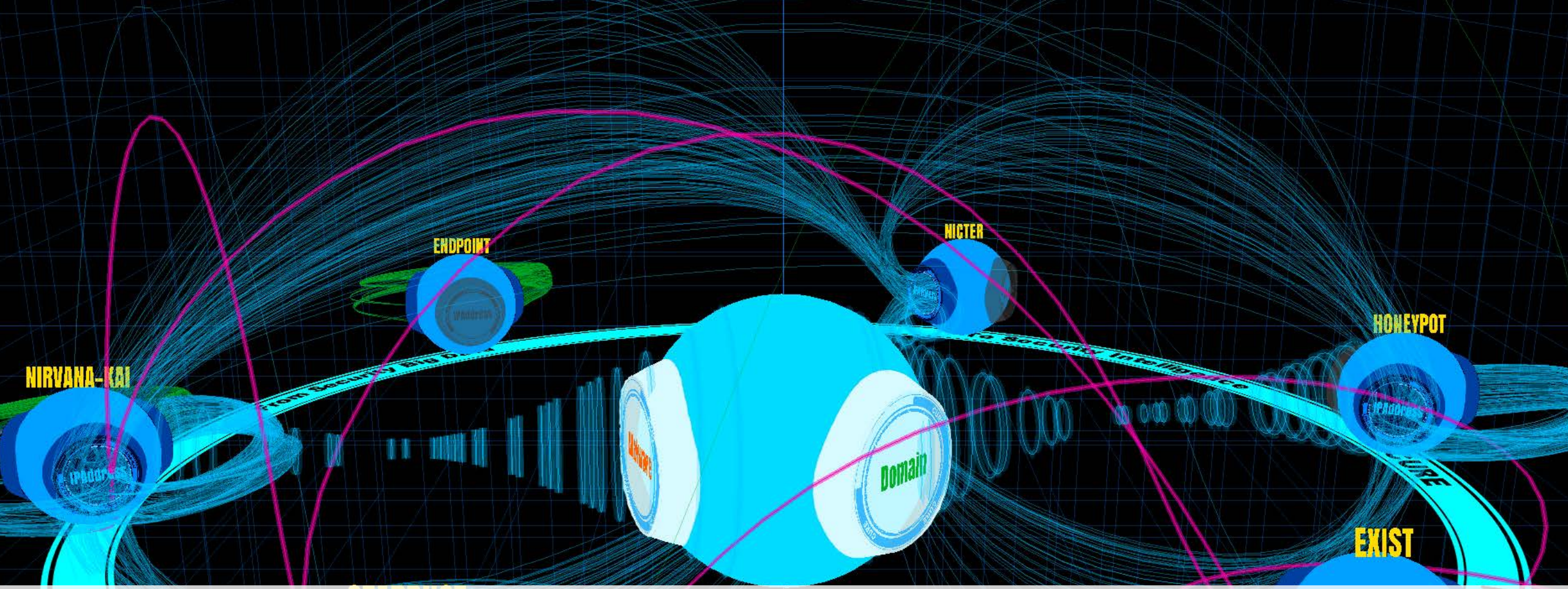
Cybersecurity **U**niversal **RE**epository

National Institute of Information and Communications Technology

Network Incident analysis Center for Tactical Emergency Response

Security Big Data in NICT

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 related data	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
Threat Intelligence	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.



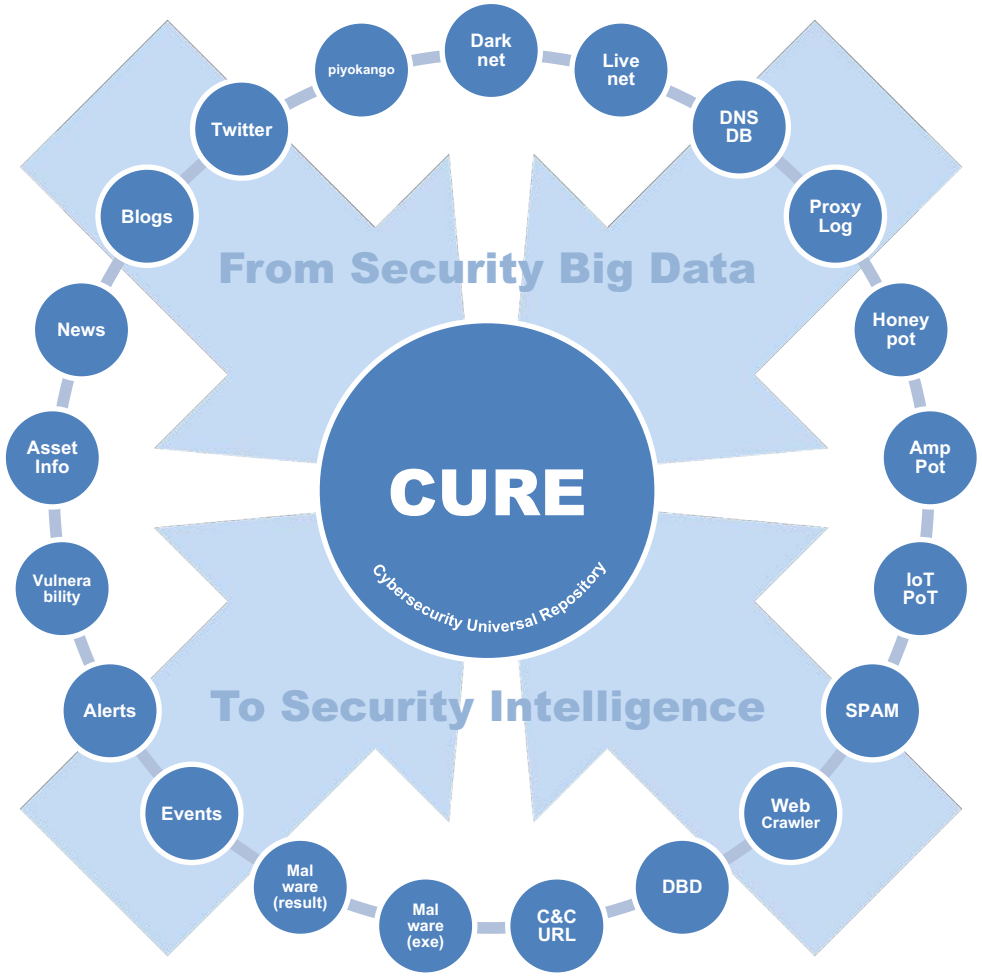
CURE

- is a platform for gathering, analyzing, and connecting **heterogeneous security big data**
- maps open source intelligence onto security alerts in an organization

CURE Concept and Implementation

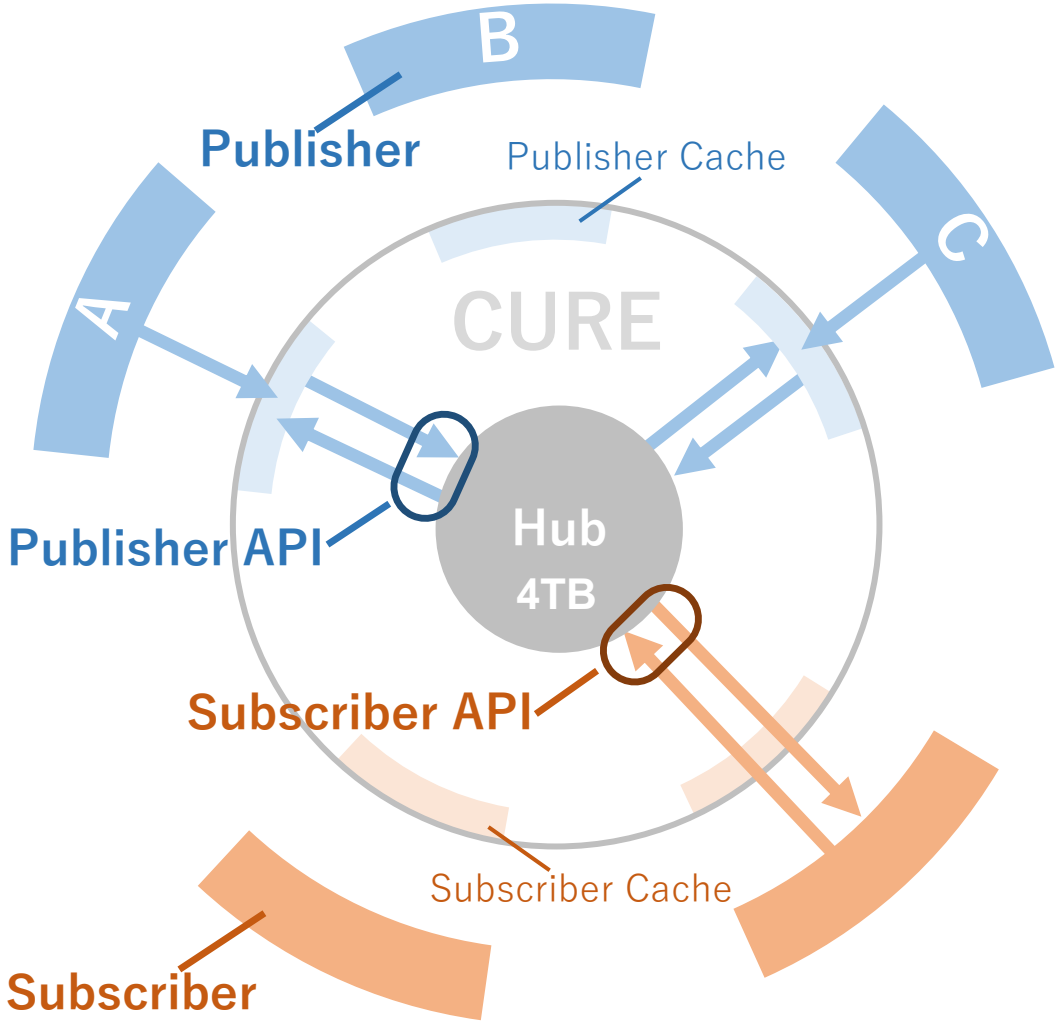
● CURE Concept

✓ Cybersecurity Universal Repository



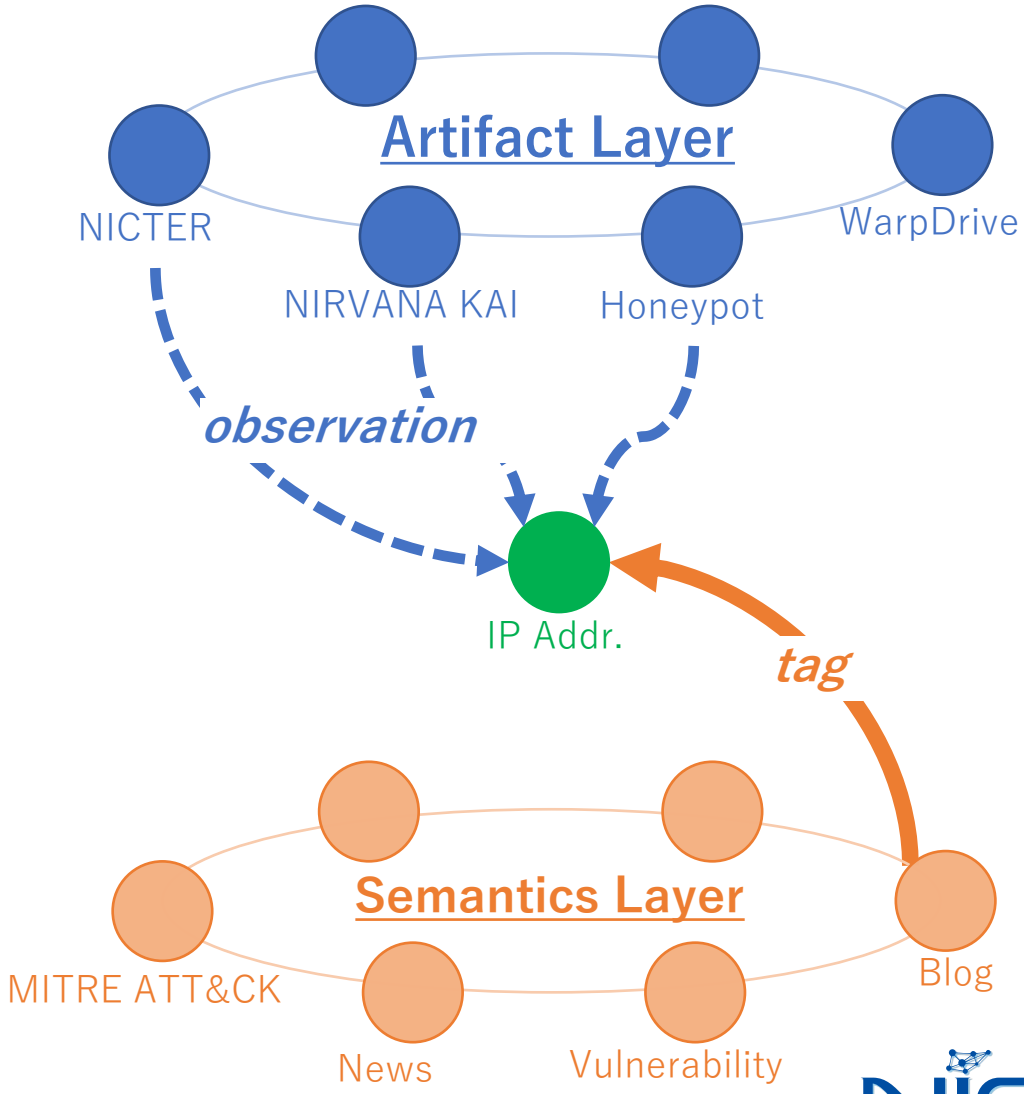
● CURE Implementation

✓ Pub/Sub Model + In Memory Database (Redis)



Artifact and Semantics in CURE

- **Artifact** (observed data)
  - ✓ IP address
  - ✓ URL
  - ✓ Malware hash
- **Semantics** (NL article)
  - ✓ Security reports
  - ✓ MITRE ATT&CK



National Institute of Information and Communications Technology

NICTER Network Incident analysis Center for Tactical Emergency Response

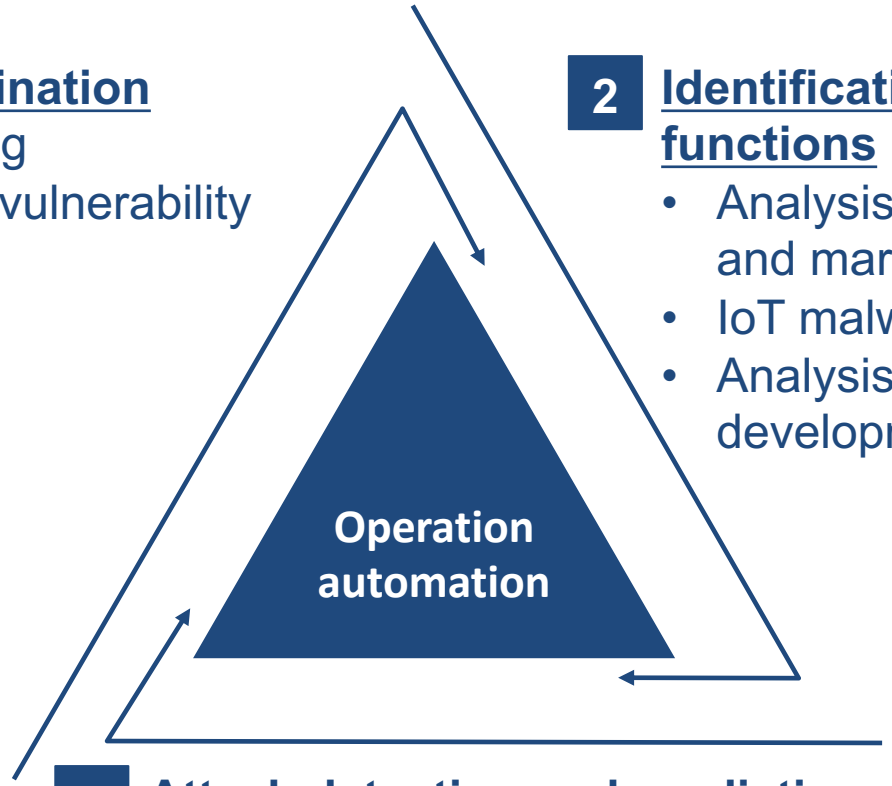
AI x Cybersecurity

Our Research Focus

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

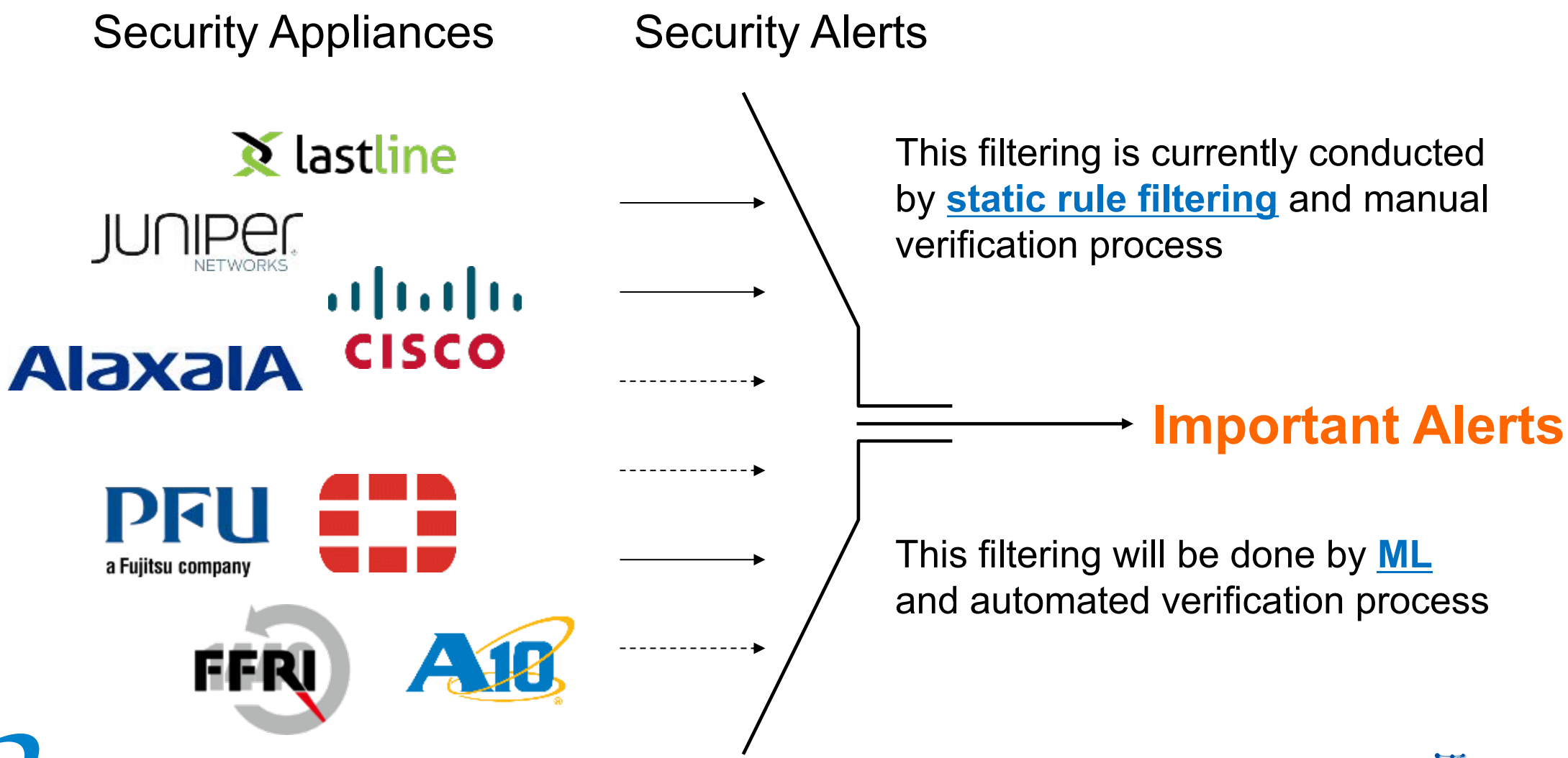
1 Priority determination
- Alert screening
- Evaluation of vulnerability severity

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



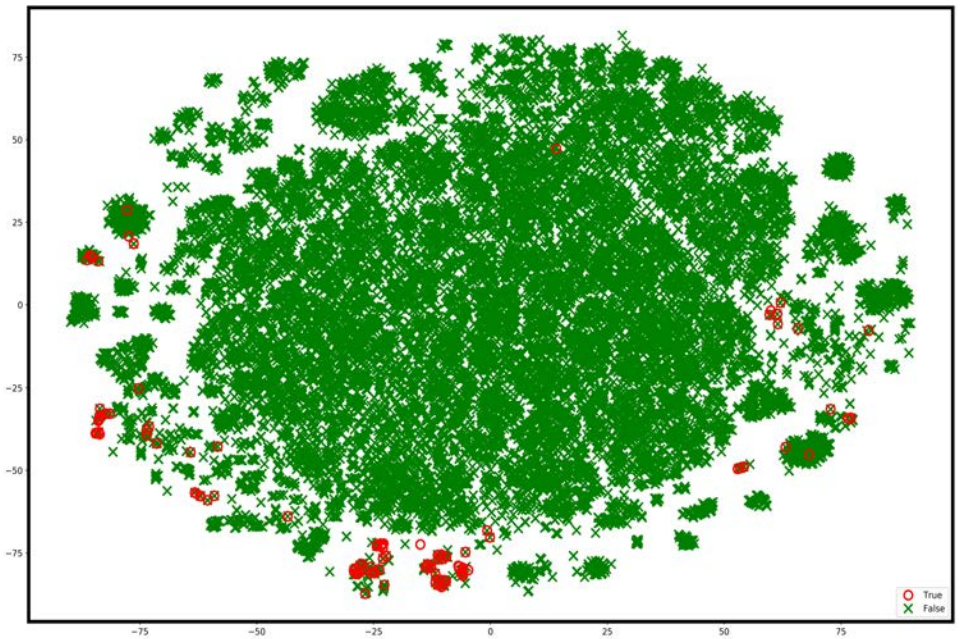
3 Attack detection and prediction
- Darknet analysis
- Threat estimation and prediction
- Encrypted traffic analysis

Alert Screening and Prioritization (1/2)

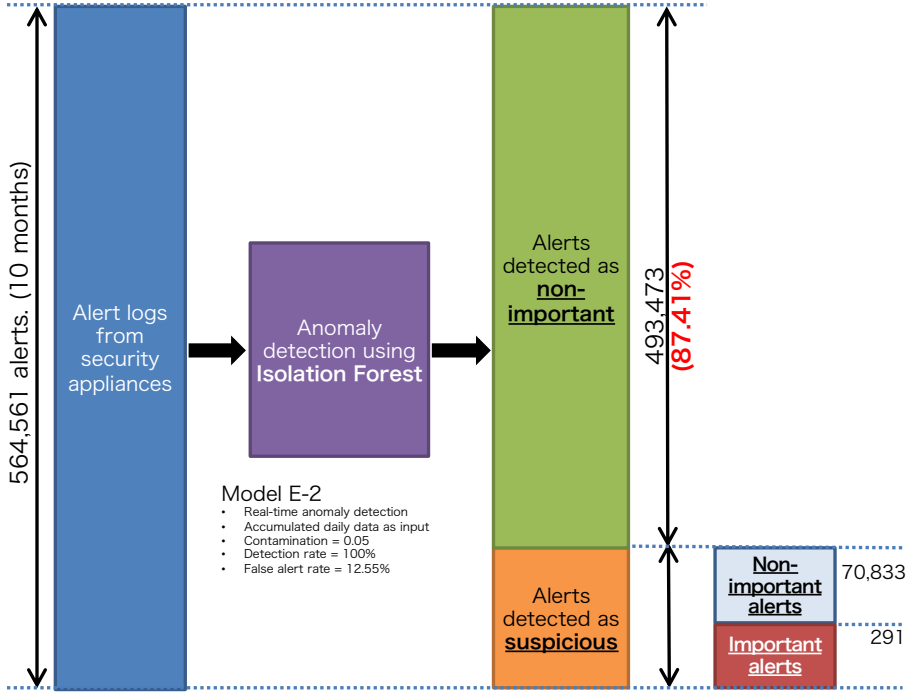


Alert Screening and Prioritization (2/2)

- **Research goal:** leverage AI to reduce the workload of security operator by screening out insignificant security alerts and recommend candidates of emerging threats on the rise.
- **Methodology:** apply an isolation forest to aggregate alerts from multiple security appliances.



Visualization of attack events using t-SNE

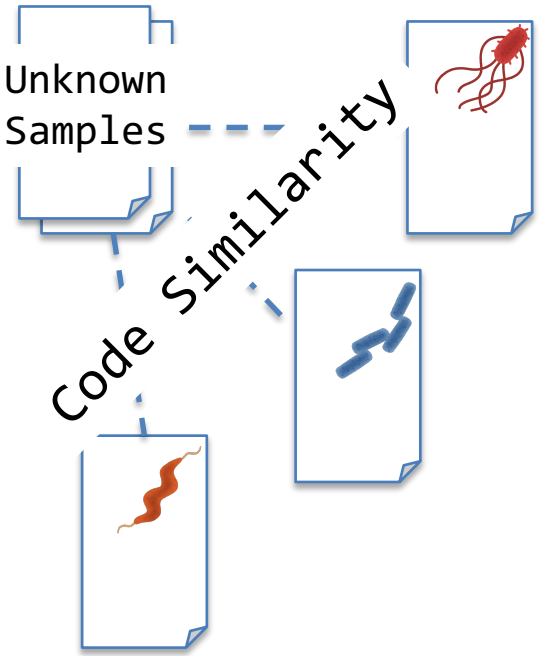


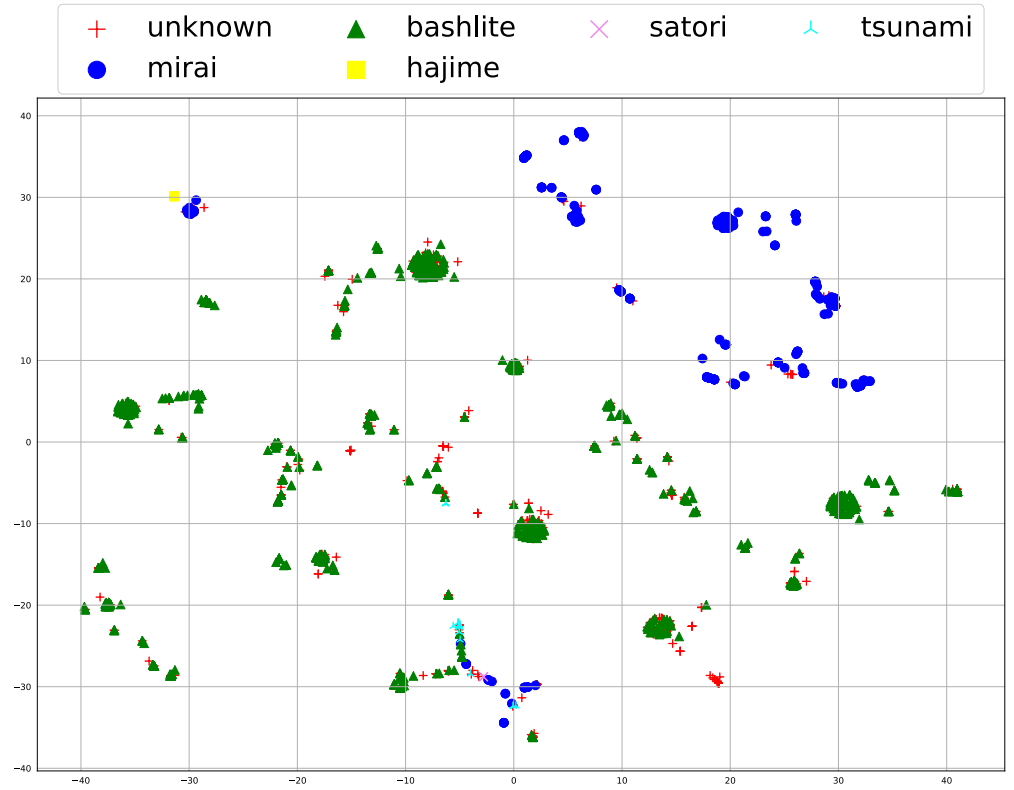
Reduce 87% insignificant alerts with no false negative[1]

[1] M. E. Aminanto et al., "Combating Threat-Alert Fatigue with Online Anomaly Detection using Isolation Forest," ICONIP'19, Dec 2019.

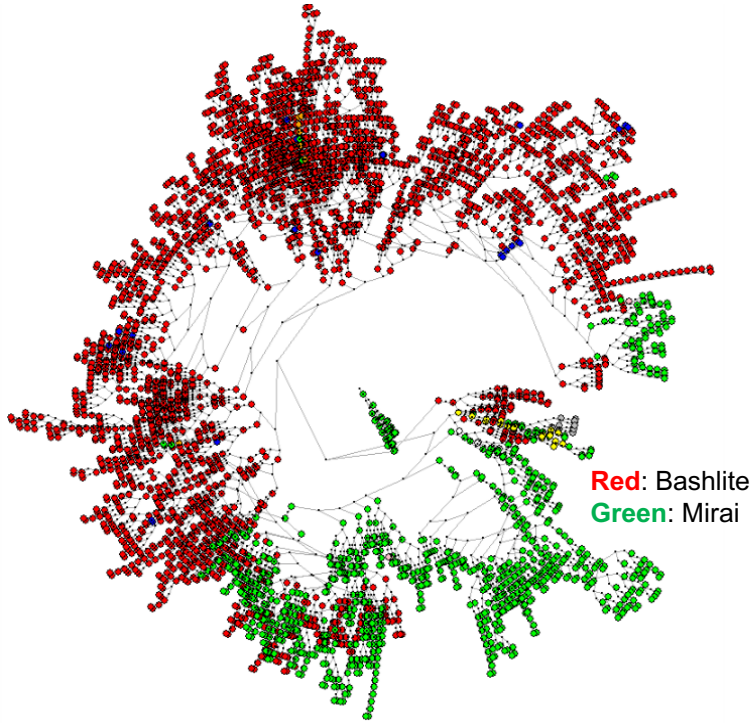
Clustering IoT Malware

- **Research goal:** identify unknown IoT malware based on their similarity
- **Methodology:** extract feature from disassembly code by n-gram and classified with SVM





IoT malware mapped with t-SNE[2][3]

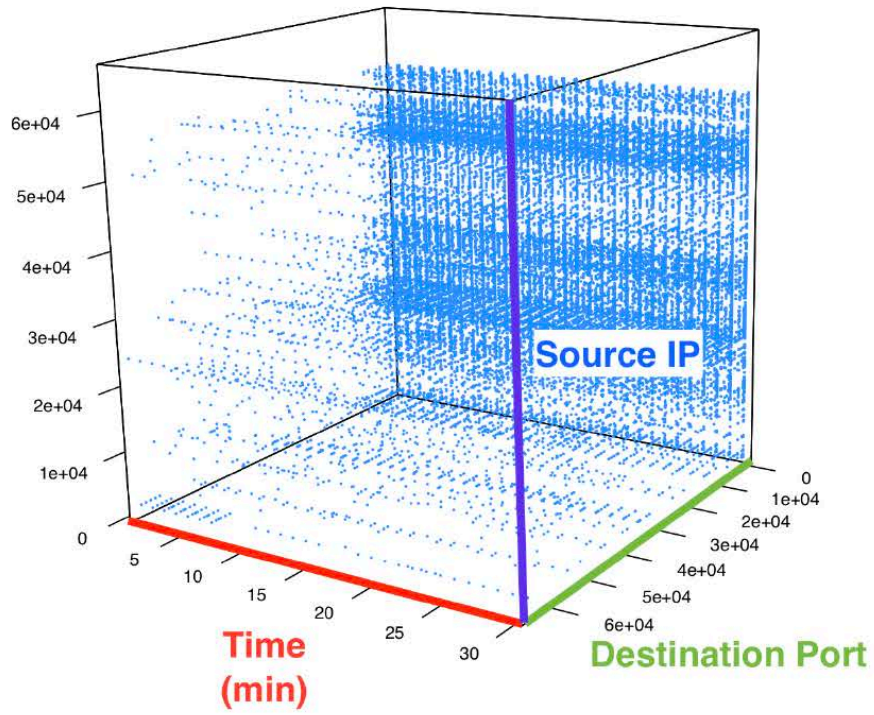


Phylogenetic trees of IoT malware[4]

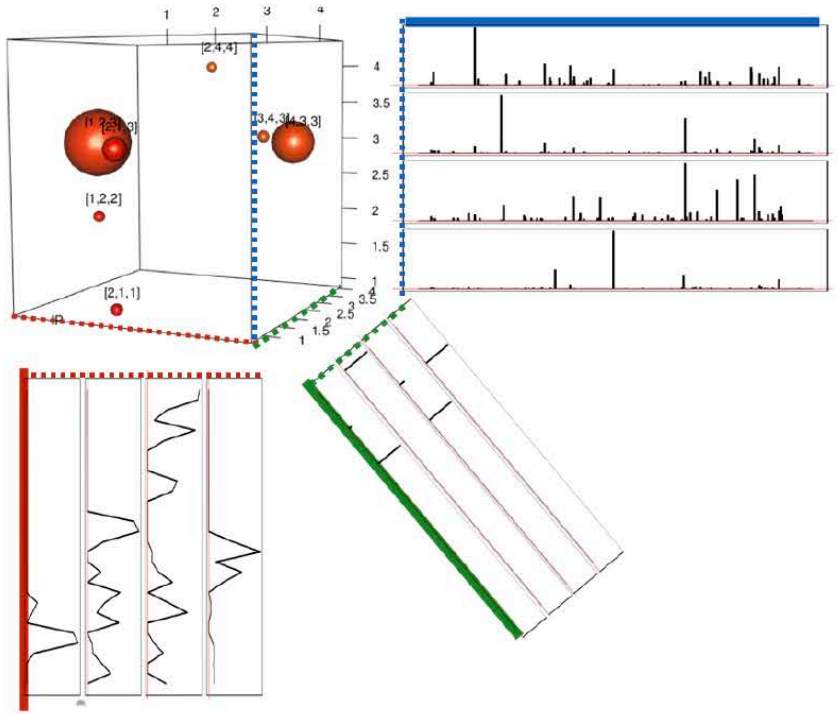
[2] R. Isawa et al., "Evaluating Disassembly-code based Similarity between IoT Malware Samples," AsiaJCIS 2018, Aug 2018.
[3] T. Ban et al., "A Cross-Platform Study on IoT Malware," ICMU2018, Oct 2018.
[4] T. He et al., "A Fast Algorithm for Constructing Phylogenetic Trees with Application to IoT Malware Clustering," ICONIP'19, Dec 2019.

Detecting Coordinated Activities (1/2)

- **Research goal:** identify coordinated activities of attacking hosts by analyzing darknet traffic with unsupervised learning techniques
- **Methodology:** tensor decomposition techniques to decompose darknet traffic into time characteristics, source IP address characteristics and destination port characteristics

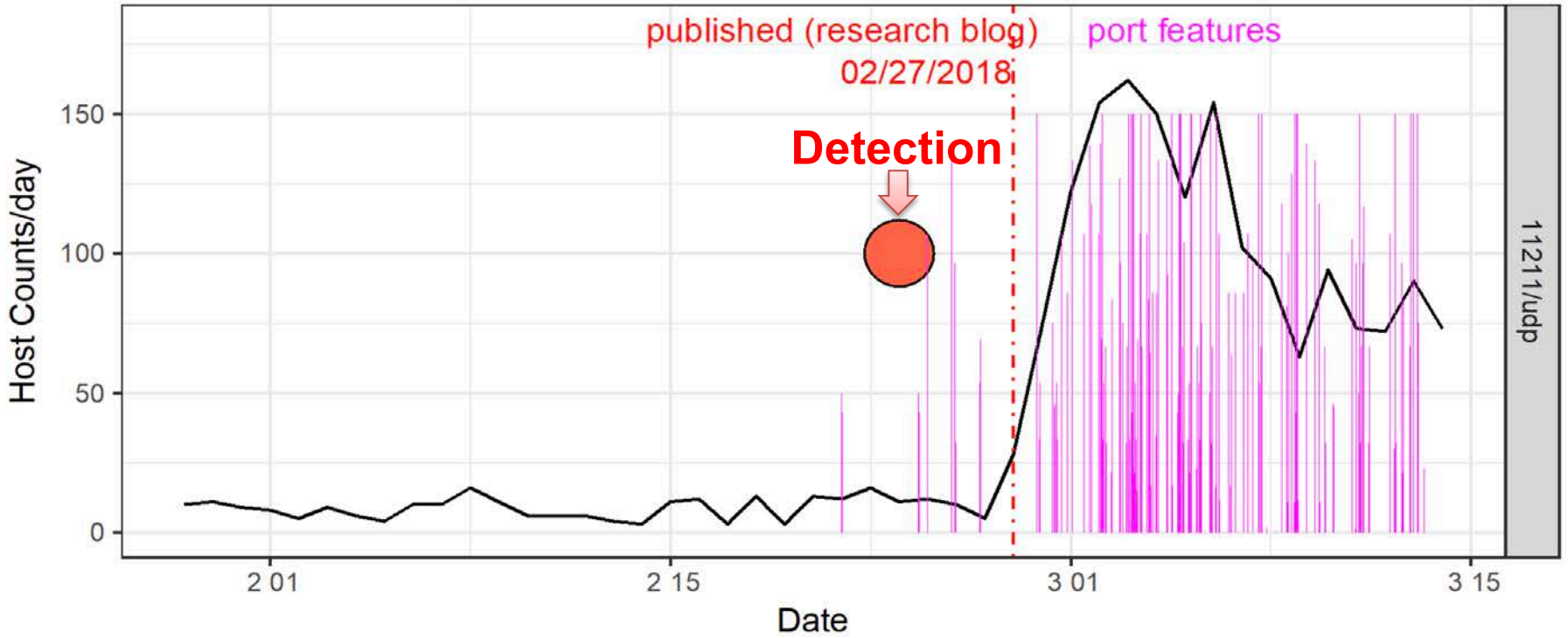


≈



Detecting Coordinated Activities (2/2)

- We were able to detect a coordinated activity toward port 11211/udp a few days prior to a well-known security blog. [5]
- It was coordinated scanning activity toward memcached servers for preparing Distributed Reflection Denial of Service (DRDoS).



[5] H. Kaneharaet al., "Real-Time Botnet Detection Using Nonnegative Tucker Decomposition," ACM SAC, 2019, Apr 2019.

Issues of AI x Cybersecurity

●Ground Truth

✓ how do we collect enough volume of labeled data?

●False Positive Reduction

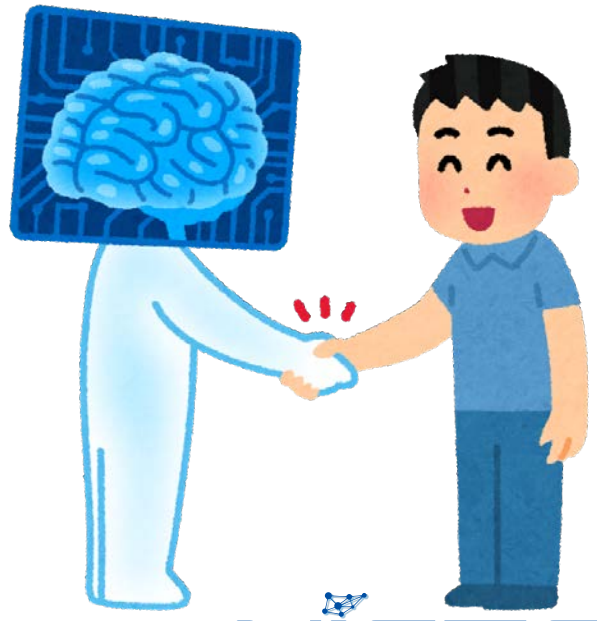
✓ true positive 99.9% → 100 thousand false positives in 100 million security alerts

●Explainable AI (XAI)

✓ explainability is the most important for real incident handling

●Real-time ML Engines

✓ security operation needs real-time and 24/7 ML engines



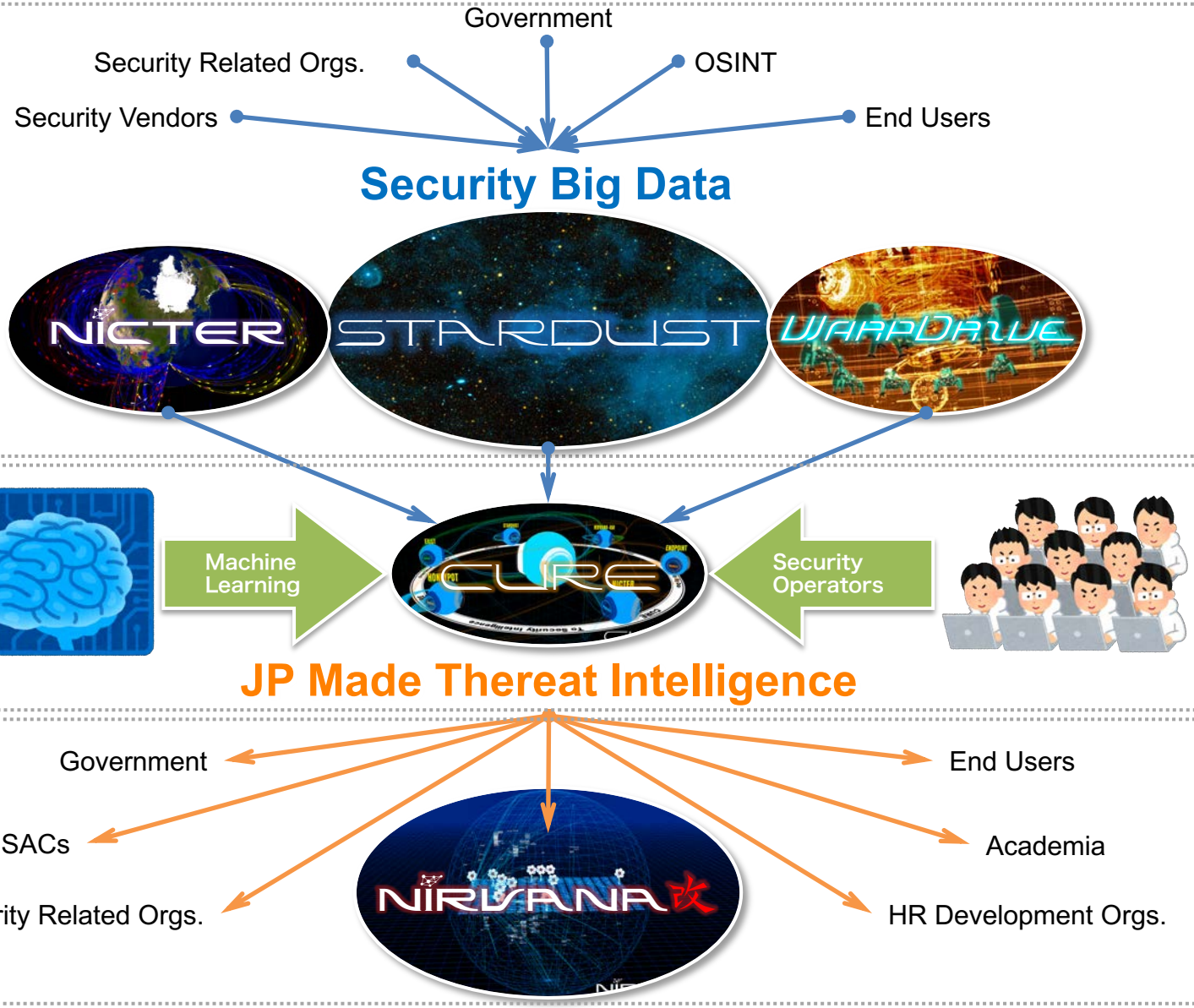
Future Works

Big Issue in Japan

- **Low self-sufficiency rate in Cybersecurity products**
  - ✓ Reported by Cyber Security Strategy Headquarters in NISC (May 2019)
- **Negative spiral of Cybersecurity data shortage**
  - ✓ No data → No R&D → No products → No data → …
- **What JP needs now is...**
  - ✓ Large-scale collection and accumulation of real data
  - ✓ Steady and systematic analysis of real data
  - ✓ Evaluation of domestic products with real data
  - ✓ Generation and share of Japan made threat intelligence

CYNEX

CYBERSECURITY NEXUS



Collection and accumulation

Steady and systematic analysis

Evaluation of domestic products

Japan made threat intelligence