Distributed and Cloud-Based Network Defense System for NRENs (DCNDS)

Tat-Chee Wan

tcwan@usm.my

Deputy Director
National Advanced IPv6 Centre (NAv6)
Universiti Sains Malaysia
11800 USM, Penang, Malaysia

This publication has been produced with co-funding of the European Union for the Asi@Connect Project under Grant contract ACA 2016-376-562. The contents of this documents are the sole responsibility of the "Distributed and Cloud-Based Network Defense System for NRENs" Project and can under no circumstances be regarded as reflecting the position of the European Union

European Union

ASIACONNECT
DISTRIBUTED AND CLOUD-BASED NETWORK DEFENSE SYSTEM FOR NRENs

6 Dec. 2021

European Union

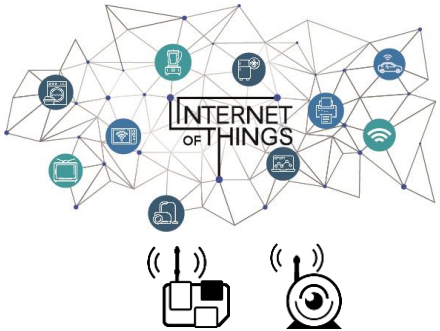
Asi@Connect

TEIN COOPERATION CENTER

Project Details

Title	*Distributed and Cloud based Network Defense System for NRENs [AsiaConnect-18-062]*
PI/ORG	*Wan Tat Chee / USM (MY), Md. Saiful Islam / BUET (BD)*
Budget/Period	*250K Euro / Jun '18 – May '20 (24 M)* ***(Extended till Feb 2021) [Project Completed]***
Beneficiary Countries	*Malaysia, Bangladesh, Indonesia, Philippines*
EU Partners	*FIWARE Foundation (France), University of Hamburg (Germany)*
Summary of Capacity Building Activities	- *Series 1 Workshop & Stakeholders Dialog on "Cloud-based Web Security Best Practices and System Configuration Overview": 113 registered participants in 4 countries trained* - *Series 2 Workshop & Stakeholders Dialog on "Botnet Mitigation Best Practices and System Evaluation": 128 registered participants in 4 countries trained.*

Background



- Increasing network threats to NRENs:
  - Web security compromises
  - Distributed botnets
- Lack of Trained NREN and Research personnel in these areas
  - Capacity Building workshops conducted in beneficiary countries
- Lack of services and tools for dealing with emerging threats
  - Deploy and evaluate commercial cloud-based security solutions for NREN services
  - Develop new tools to address distributed botnet issue
- Participating Countries:
  - Beneficiaries: Malaysia, Bangladesh, Philippines, Indonesia
  - EU Partners: Germany, France

Discovered	Botnet	Active?
2006	Nugache	No
2007	Storm	No
2008	Waledac	No
	Sality (P2P)	Yes
2009	Conficker	No
2010	Kelihos	No
2011	ZeroAccess	No
	Miner	No
	GameOver Zeus	No
2016	Hajime	Yes
	DDG	Yes
2018	Hide'N'Seek	Yes
2019	Roboto	Yes
	Mozi	Yes
	IPFS-Storm	Yes
2020	HEH	Yes

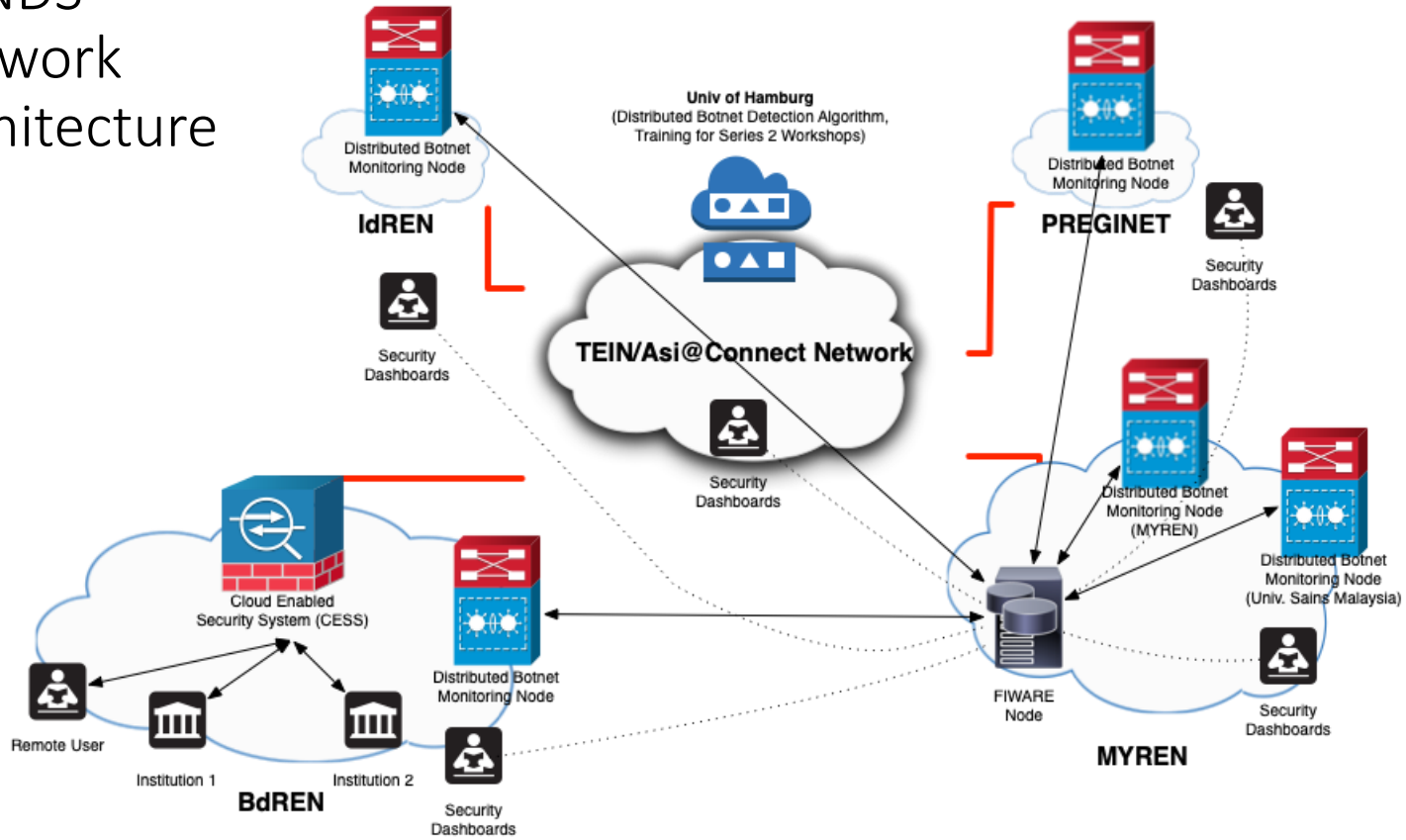
Overview of DCNDS Project

- Capacity Building in Cloud and Botnet related areas
  - 2 series of workshops, conducted in each Asian partner country (Malaysia, Bangladesh, Indonesia, Philippines)
    - *Cloud-based Web Security Best Practices and System Configuration Overview*
    - *Botnet Mitigation Best Practices and System Evaluation*
- Distributed and Cloud-Based Network Defense System
  - Cloud-enabled Security System (CESS) for BdREN
    - To manage web security for participating institutions in Bangladesh
  - DCNDS Botnet Detection System (DBDS) for project partners
    - Monitoring and data aggregation/anonimization at partner NREN
    - Data analytics and detection on FIWARE Cloud system (hosted at MYREN/USM)
    - Security Dashboard for NREN partners
    - Published Research Dataset for reference (hosted on Zenodo.org)

Capacity Building Workshop Statistics

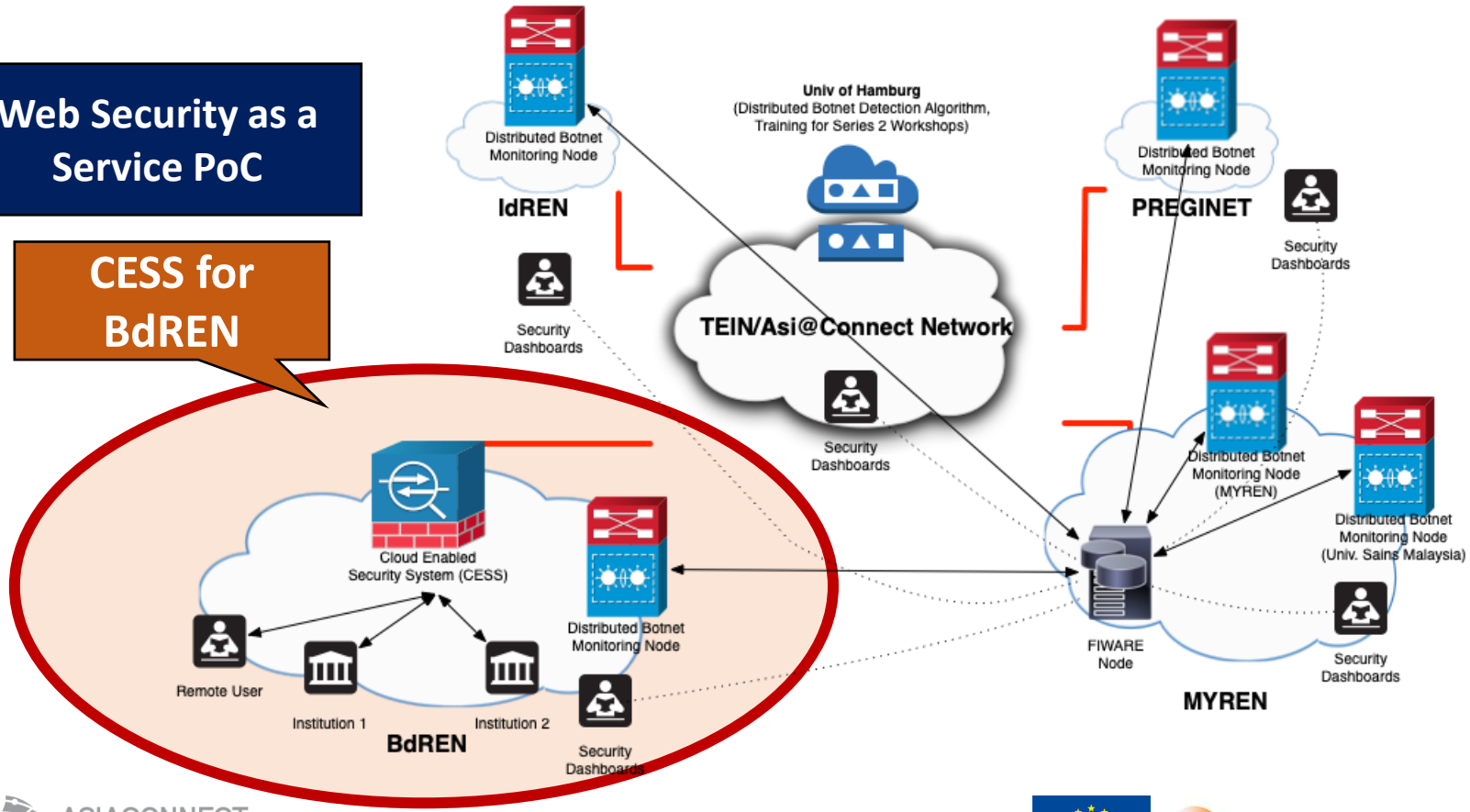
Series 1 (Web Security)	MY	BD	ID	PH	SubTotals
Male	14	28	23	29	94
Female	5	3	2	9	19
				Series 1 Total	**113**
Series 2 (Botnet Mitigation)	MY	BD	ID	PH (Virtual)	SubTotals
Male	17	28	28	27	100
Female	4	3	2	19	28
				Series 2 Total	**128**

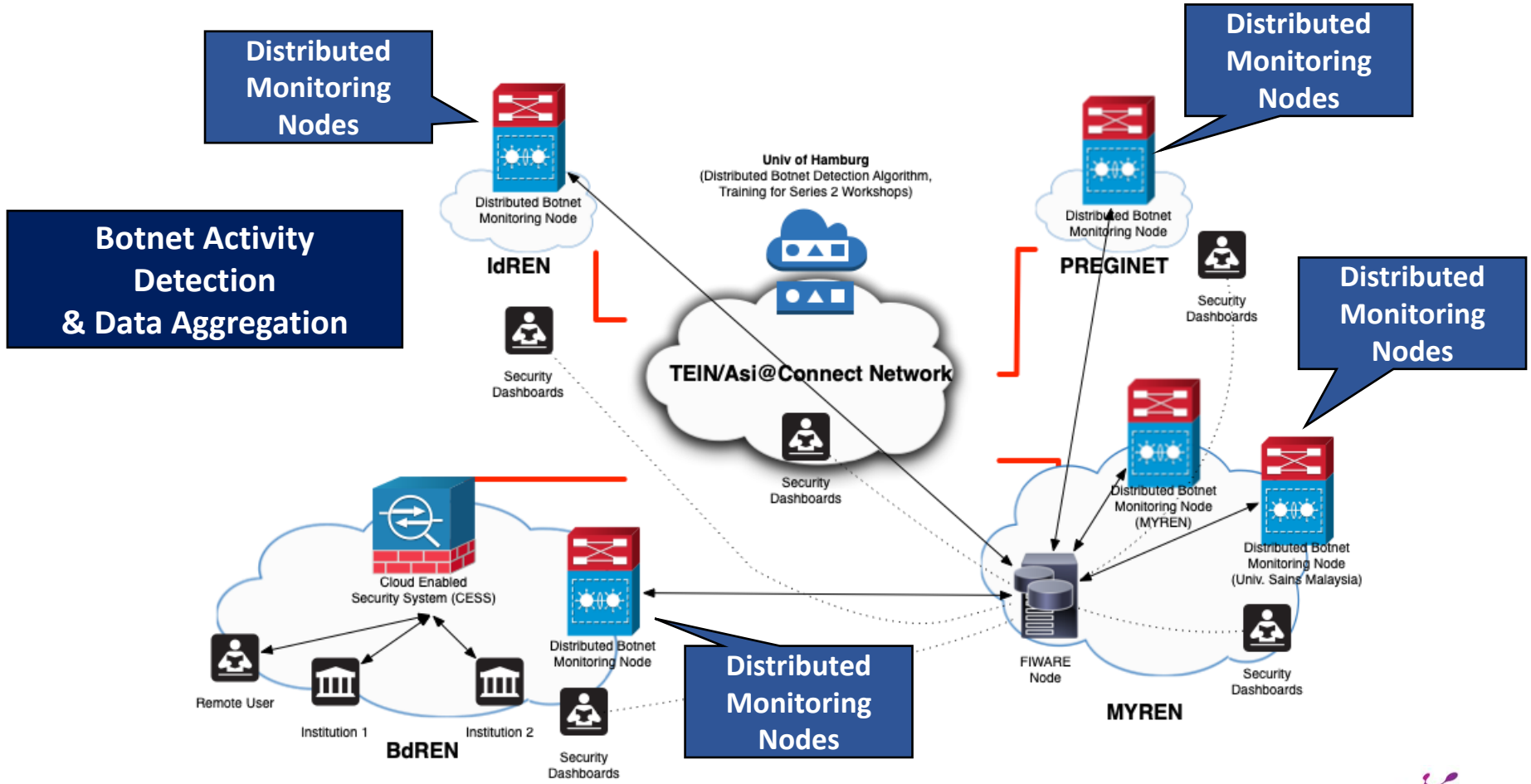
DCNDS Network Architecture



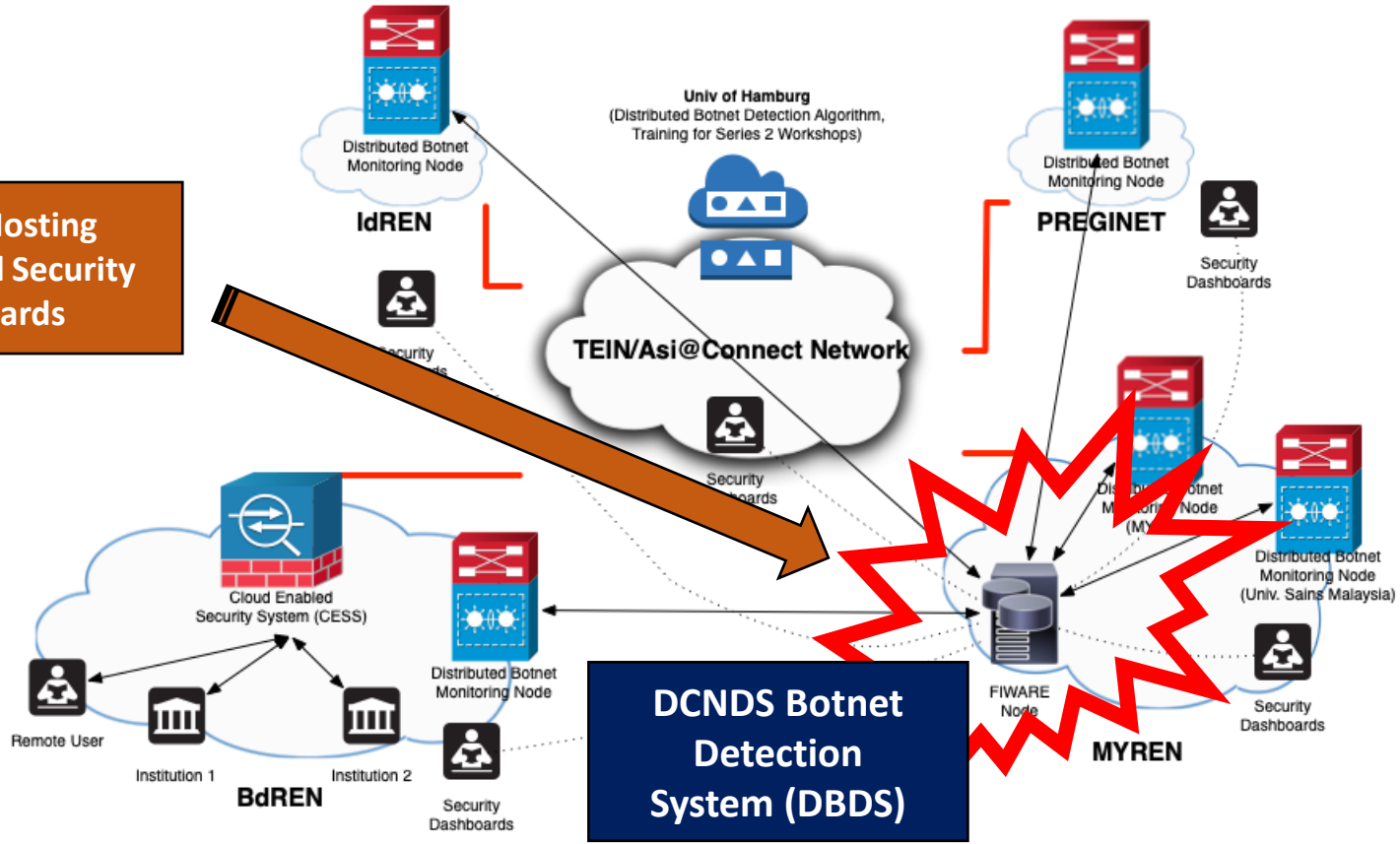
Web Security as a Service PoC

CESS for BdREN



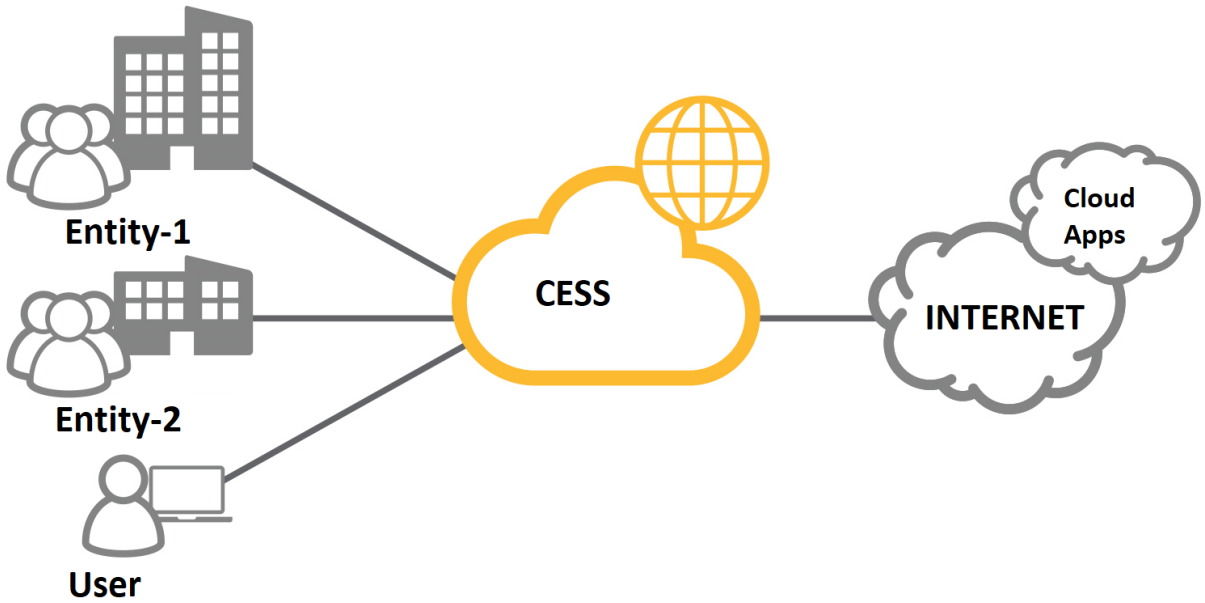

Distributed Monitoring Nodes
Distributed Monitoring Nodes
Botnet Activity Detection & Data Aggregation
Univ of Hamburg
(Distributed Botnet Detection Algorithm, Training for Series 2 Workshops)
Distributed Botnet Monitoring Node
IdREN
Distributed Botnet Monitoring Node
PREGINET
Security Dashboards
Security Dashboards
TEIN/Asi@Connect Network
Security Dashboards
Distributed Monitoring Nodes
Distributed Botnet Monitoring Node (MYREN)
Distributed Botnet Monitoring Node (Univ. Sains Malaysia)
Cloud Enabled Security System (CESS)
Distributed Botnet Monitoring Node
FIWARE Node
Security Dashboards
MYREN
Remote User
Institution 1
Institution 2
BdREN
Security Dashboards
Distributed Monitoring Nodes

FIWARE Hosting Cloud-based Security Dashboards



DCNDS Botnet Detection System (DBDS)

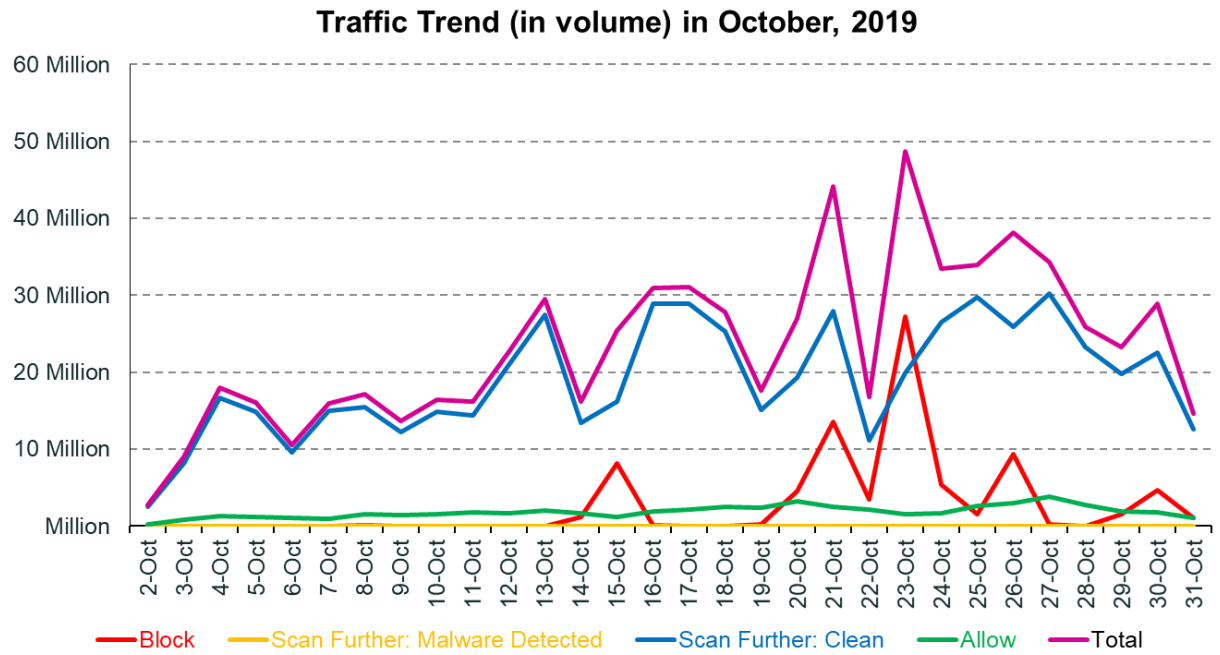
CESS Architecture for BdREN



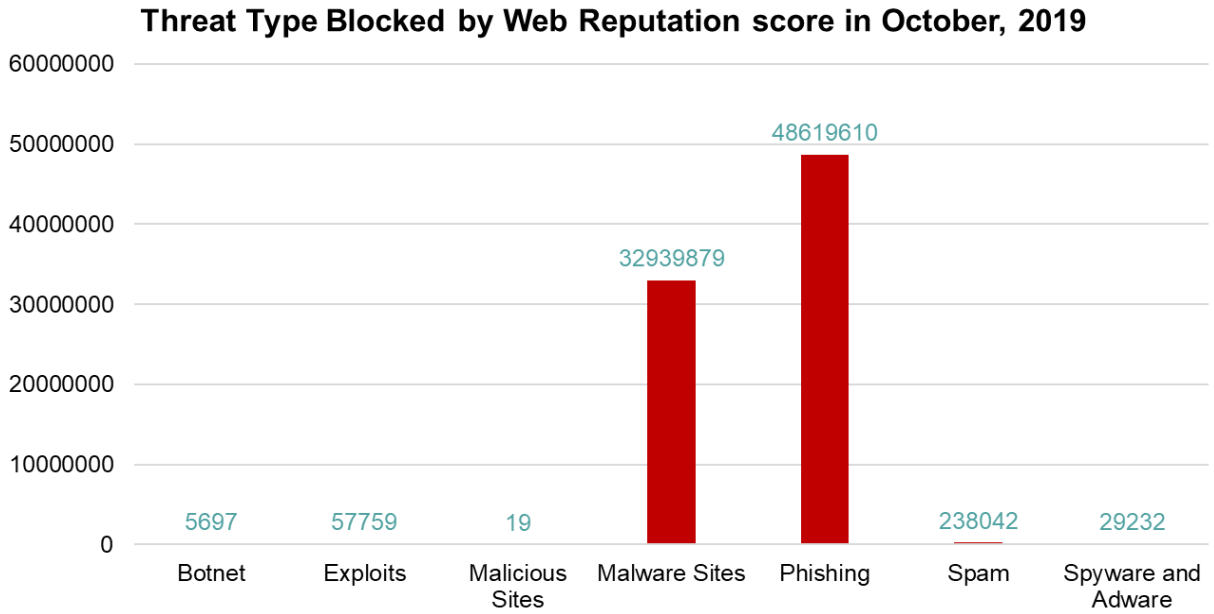
CESS System Information

- Product: Cisco Web Security Appliance (virtual)
- Model: S600V
- Operating System: AsyncOS
- OS Version: 11.7.0-418
- Deployment: Vmware ESXi (version 5.5)
- Disk Size: 1024 GB
- RAM: 24 GB
- Processor Core: 12
- Deployment Location: BdREN Cloud

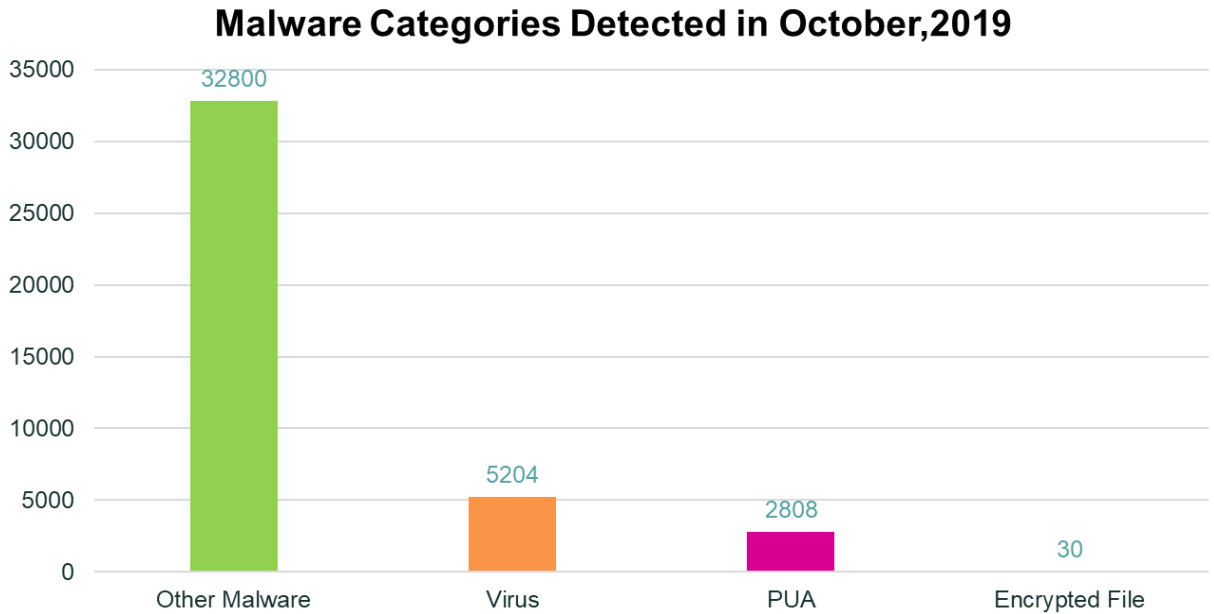
CESS Sample Statistics



CESS Web Traffic Threat Mitigation

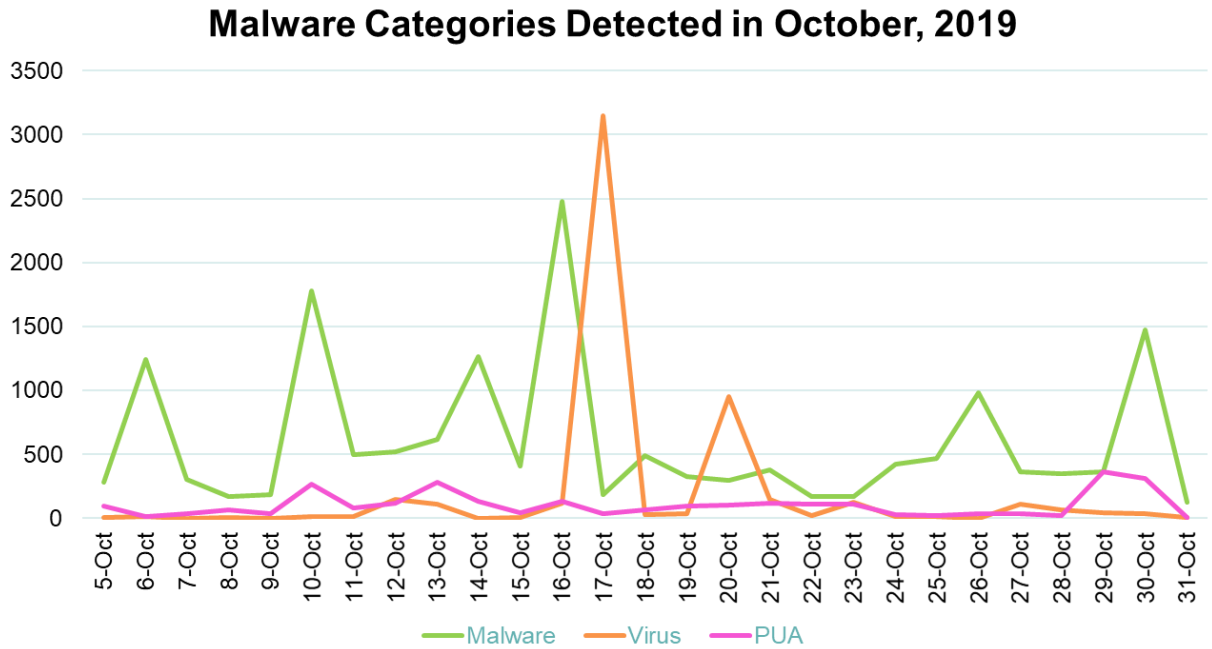


CESS Malware Detection



PUA: Potentially Unwanted Applications

CESS Malware Activity Detection



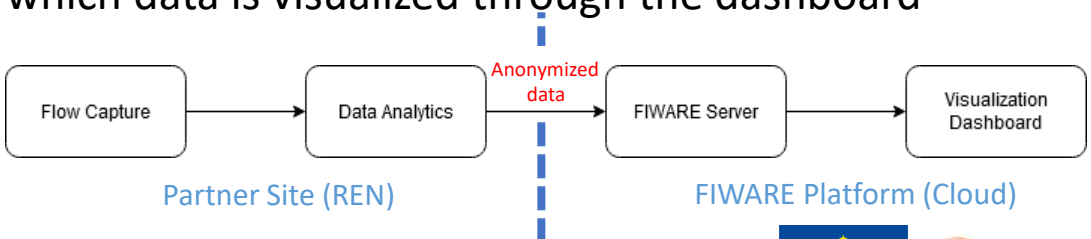
PUA: Potentially Unwanted Applications

CESS Lessons Learnt

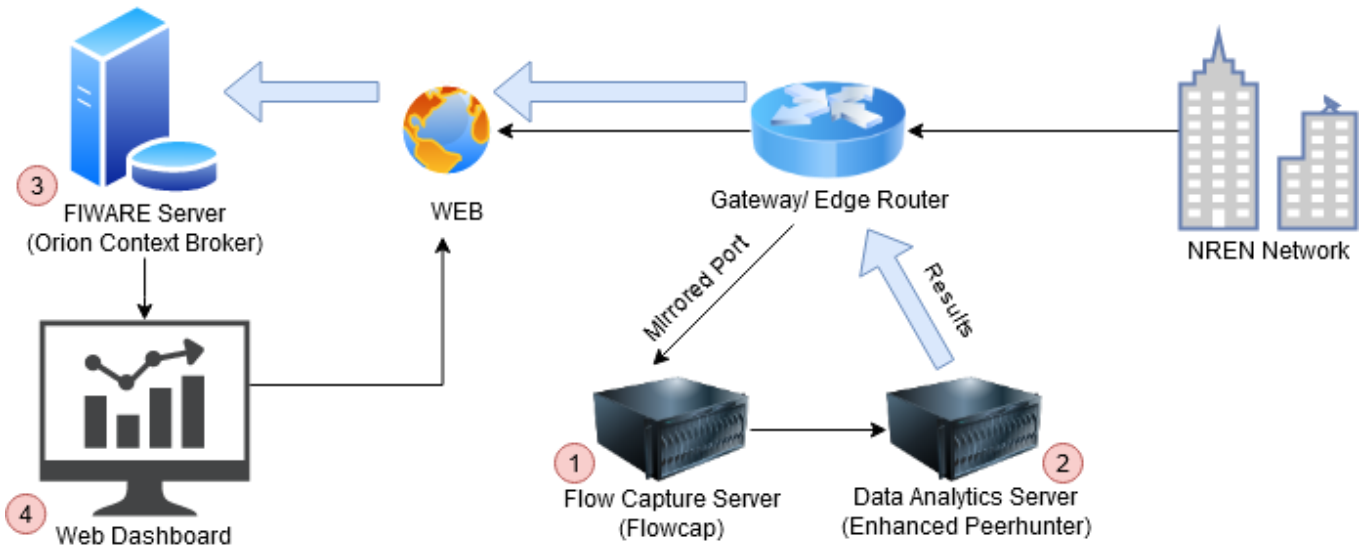
- It took several months to get the system up and running
- The initial plan was to pass network traffic to CESS in transparent mode
  - However, it was not possible to implement transparent mode, due to hardware limitations at BDRen routers and switches
- As a result, it was not possible to capture entire institutional network traffic data
- We only captured data to some BDRen servers in VLANs protected by CESS

DCNDS Botnet Detection System

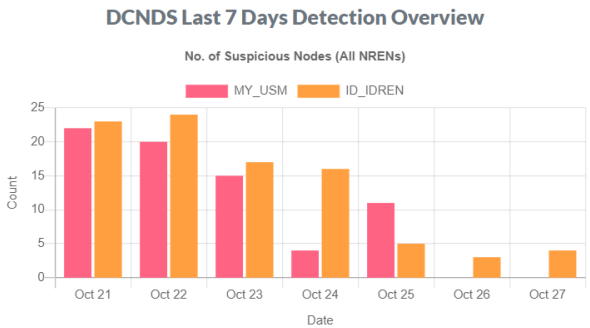
- The DCNDS Botnet Detection System (**DBDS**) is a collection of:
  - A network flow capture program
  - A Big Data Analysis system designed to detect suspicious P2P Botnet traffic
  - An anonymized data feed to collate detections from various sites
  - A Web Dashboard to visualize the data
- Each partner Research and Education Networks (RENs) deploy the components, capture flows, analyse and send anonymized results to a centralized FIWARE server from which data is visualized through the dashboard

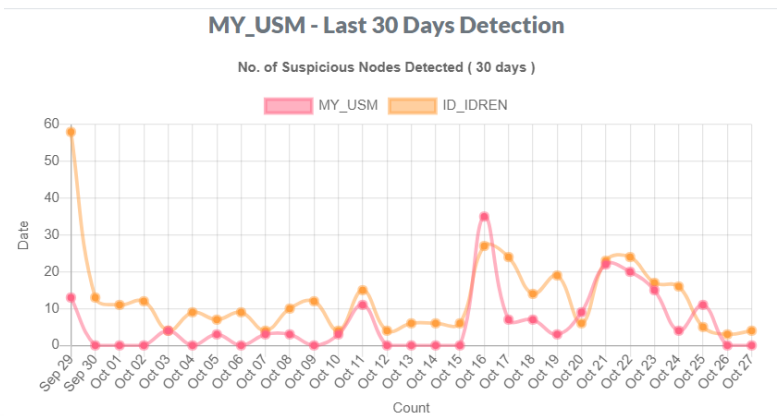


DBDS Overview



DBDS Web Dashboard



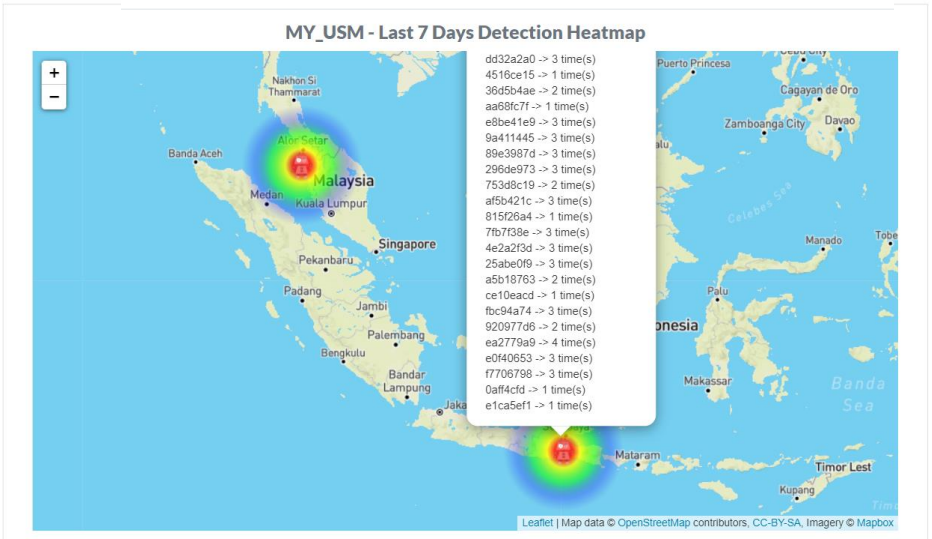


MY_USM - Last 7 Days Recurring Suspicious Nodes

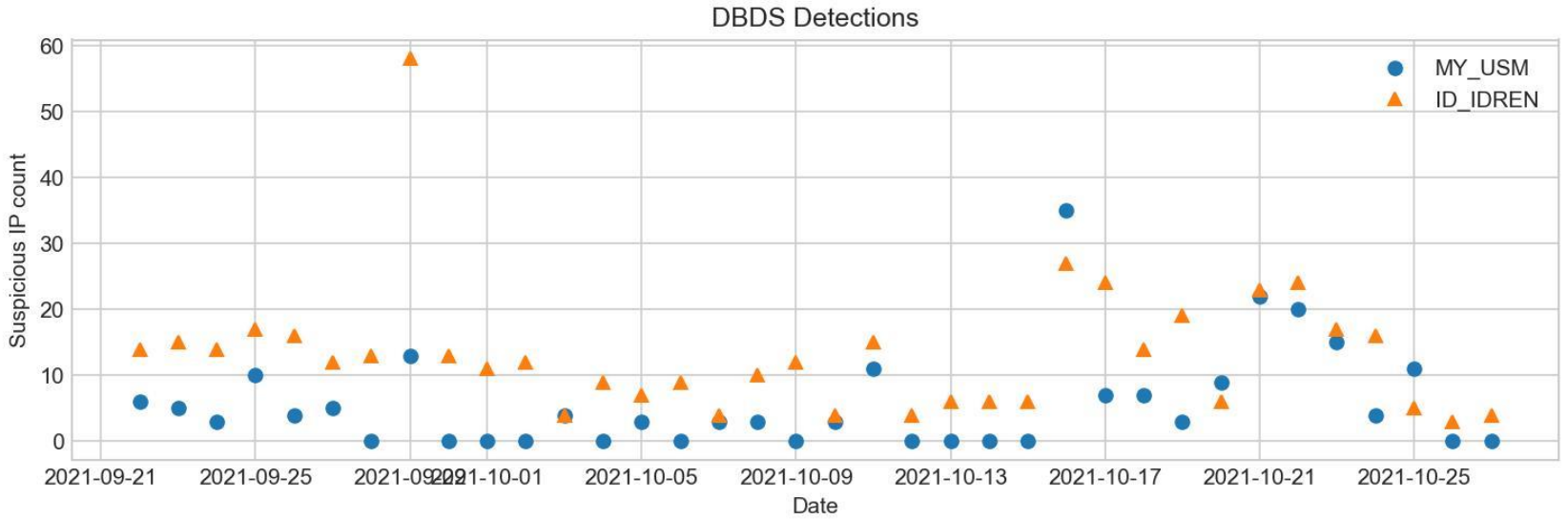
Show 10 entries Search:

IP Hash	Geolocation	Occurences
3a79f9b9	Malang, East Java, Indonesia.	4
ea2779a9	Malang, East Java, Indonesia.	4
f472806b	George Town, Penang, Malaysia.	3
7b1735c9	George Town, Penang, Malaysia.	3
cac29f45	George Town, Penang, Malaysia.	3
522c9d87	George Town, Penang, Malaysia.	3
ba18eddd	George Town, Penang, Malaysia.	3
3dc6e01d	George Town, Penang, Malaysia.	3
1ec75702	George Town, Penang, Malaysia.	3
836cd955	George Town, Penang, Malaysia.	3

Showing 1 to 10 of 56 entries Previous 1 2 3 4 5 6 Next



DBDS Statistics: Sept-Oct'21



DBDS Deployment Status

NREN	Hardware Setup	Software Setup	Data Feed @ FIWARE
BD-BdREN			 Since: -
ID-IdREN			 Since: 27th Feb'21
MY-USM			 Since: 8th Feb'21
PH-PREGINET			 Since: -

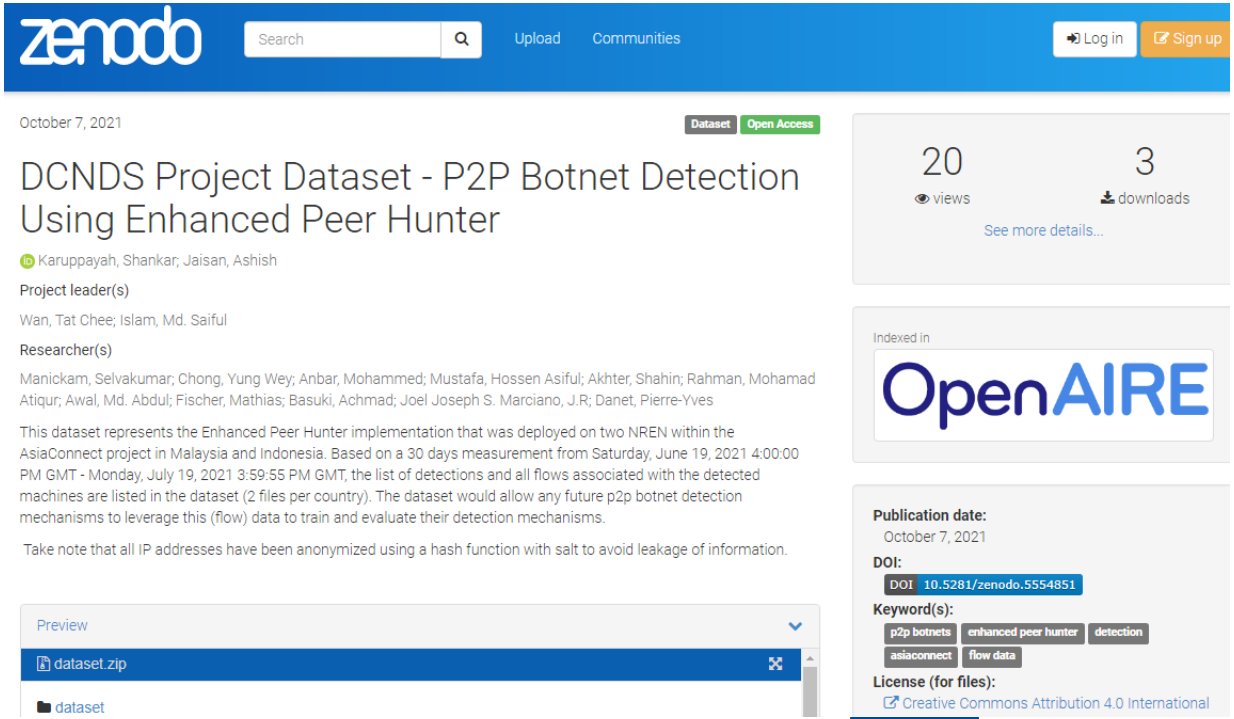
DBDS Summary

- DBDS Components have been deployed successfully at all four partner locations
- FIWARE services and Visualization Dashboard are up and running
- Intention for further development of DBDS via new funding

Acknowledgement of thanks to all partners, Advanced Science and Technology Institute (ASTI) [PH], Universitas Brawijaya [ID], Bangladesh University of Engineering and Technology (BUET) [BD], University of Hamburg [Germany], and FIWARE Foundation [EU] for their continuous support and efforts

Special thanks to members of NAv6 and USM Centre for Knowledge, Communication and Technology which provided technical support and fruitful discussions

Dataset has been published on Zenodo (June 19 – July 19, 2021)



ASIACONNECT
DISTRIBUTED AND CLOUD-BASED NETWORK DEFENSE SYSTEM FOR NRENs

European Union

Asi@Connect

TEIN COOPERATION CENTER

Project Lessons and Challenges

- In-country capacity building is able to attract more local participants
  - Many participants do not have opportunity to attend centralized training in another country due to funding issues
- Onset of COVID lockdowns in March 2020 affected the last workshop scheduled in Philippines
  - Final Workshop for Philippines conducted virtually in Nov 2020
  - Virtual training is the 'new norm' but is not easy for conducting hands-on workshops
- Extended COVID Lockdowns affected development of DBDS
  - Difficult to continue software development as no campus access for students and staff
  - Deployed to partners in Feb 2021, operational since Mar 2021
    - Local packet-capture and first-level analytics is running properly
    - Anonymized data feed to FIWARE is working for MY and ID sites
  - Deployment and testing of packet capture nodes in partner NREN sites need in-person attention
  - Coordination and communication is slow due to weak off-site Internet access for many partners

Training Materials and Research Dataset

- All workshop materials are available from http://dcnds.asia
  - Training session videos
  - Presentation slides
- Research Dataset for Botnet activity traces available from Zenodo.org
  - https://dcnds.asia/index.php/dbds-dataset/

Contact us

- **https://dcnds.asia**
- **Tat-Chee Wan**
  - Email: tcwan@usm.my
    National Advanced IPv6 Centre (NAV6)
    Universiti Sains Malaysia
    11800 USM, Penang, Malaysia
- **Md. Saiful Islam**
  - Email: mdsaifulislam@iict.buet.ac.bd
    Institute of Information and Communication Technology (IICT)
    Bangladesh University of Engineering and Technology
    West Palashi, Dhaka, Bangladesh

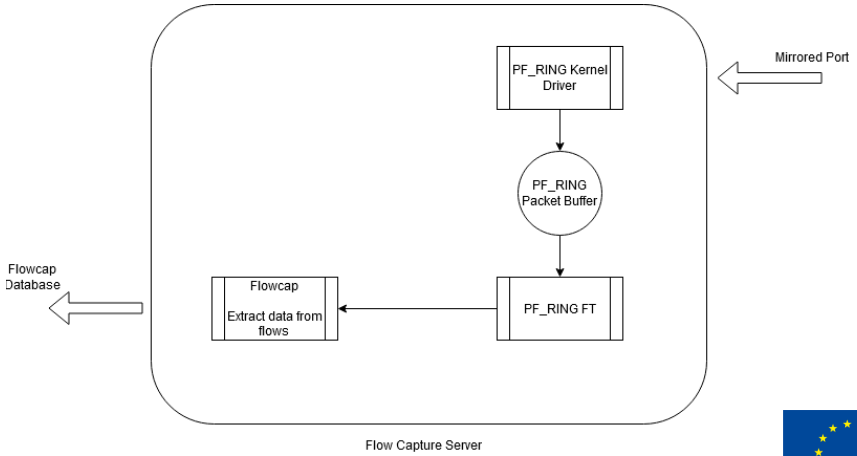
Thank you

1. Flow Capture Component Overview

- The network flow capture component, dubbed '**Flowcap**' is a program that captures network flow from a gateway/edge router at each partner location using a **mirrored network port**
- Flowcap captures the following attributes from a flow:
  1. ***First packet timestamp***
  2. ***Last packet timestamp***
  3. ***Source IP, Source Port***
  4. ***Destination IP***
  5. ***Destination Port***
  6. ***L4 Protocol (TCP, UDP)***
  7. ***Bytes-per-packet out***
  8. ***Bytes-per-packet in***
  9. ***Total no. of packets out***
  10. ***Total no. of packets in***
- Flowcap stores the captured network flow data on to the database hosted at the local **Data Analytics Server**

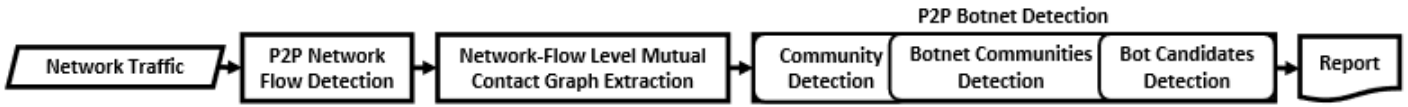
1. Flow Capture Component Overview

- Flowcap is written with PF_RING FT library [1], an **optimized high performance network flow capture library** capable of *10 Gbit* line rate on a single Server CPU core and can scale up to *100 Gbit* on multi core systems
- A regular packet capture wouldn't be efficient enough to handle the large amount of traffic from an NREN, and therefore PF_RING library is used for the project



2. Data Analytics Component

- The Data Analytics Component, Enhanced PeerHunter*, a network-flow level community-behavior-analysis based system, able to detect P2P botnets
- Enhanced PeerHunter analyses network flows captured by Flowcap
  - It starts from a P2P network flow detection component
  - Then, it uses “mutual contacts” to cluster bots into communities.
  - Finally, it uses network-flow level community behavior analysis to detect potential botnets



*D. Zhuang and J. M. Chang, "Enhanced PeerHunter: Detecting Peer-to-Peer Botnets Through Network-Flow Level Community Behavior Analysis," in *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 6, pp. 1485-1500, June 2019

2. Data Analytics Component (Cont.)

MapReduce Programming Model

- **MapReduce** is a programming model and an associated implementation for processing and generating big data sets with a parallel, distributed algorithm on a cluster
- A MapReduce program is composed of:
  1. A map procedure, which performs filtering and sorting (such as sorting students by first name into queues, one queue for each name)
  2. A reduce method, which performs a summary operation (such as counting the number of students in each queue, yielding name frequencies)
- The experimental implementation is written with Hadoop's MapReduce Framework and can be deployed in cloud computing platforms
- Therefore, the implementation is highly scalable (i.e., processing an average of 97 million network flows in about 20 minutes [2])

2. Data Analytics Component (Cont.)

Enhanced PeerHunter (EPH)

- Enhanced PeerHunter is a network-flow level community-behavior-analysis based system, able to detect P2P botnets [2]
- *Enhanced* PeerHunter is an extension to previous work: *PeerHunter*
- Enhanced PeerHunter considers the scenario **that P2P bots and legitimate P2P applications** could run on the same set of hosts whereas PeerHunter does not
  - Better accuracy (lower false positives)

2. Data Analytics Component (Cont.)

Enhanced PeerHunter (EPH)

- **Network flows used by EPH:**
  1. P2P network flow detection component.
     Uses the **distinct** number of **/16 IP prefix** to distinguish distinct P2P flows (**destination diversity - DD**).
  2. Use "*mutual contacts*" to cluster bots into communities.
  3. Use network-flow level **community behavior analysis** to detect potential botnets .
- A P2P host usually communicates with peers distributed in a large range of physical networks, which results in DD
- A /16 prefix is used rather than BGP (most BGP prefixes are /24) prefix to approximate DD, since /16 IP prefix is more likely to belong to different networks compared to a /24 prefix. Therefore, a /16 prefix considered to be a good approximation of network boundaries [2]
- Enhanced PeerHunter uses DD to decide if a flow is P2P

2. Data Analytics Component (Cont.)

Enhanced PeerHunter (EPH): Mutual Contacts

- The **mutual contacts (MC)** between a pair of hosts is a set of shared contacts between them
- Compared with legitimate hosts, a pair of bots within the same P2P botnet has higher probability to share mutual contacts
- In order to prevent bots (peers) from churning, the bot master must check each bot periodically, which results in a convergence of contacts among peers within the same botnet
- The basic idea of using mutual contacts is to build a **mutual contacts graph (MCG)** as shown in Fig. 1, a host level MCG, where A, B are linked together in Fig. 1b, since they have mutual contacts 1, 2 in Fig. 1a. Similarly, C, D, E are linked to each other in Fig. 1b, since every pair of them share at least one mutual contacts in Fig. 1a

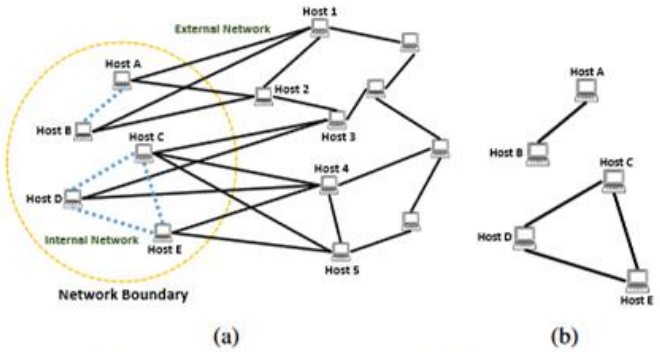


Fig. 1: Illustration of network (a) and its mutual contacts graph (b).

2. Data Analytics Component (Cont.)

Enhanced PeerHunter (EPH): Community Behavior Analysis

- Bots within the same P2P botnet always work together as a community, thus, should have distinguishable community behaviors
- Enhanced PeerHunter considers three types of community behaviors:
  - (a) flow statistical feature
  - (b) numerical community feature
  - (c) structural community feature
- PeerHunter uses the incoming and outgoing bytes-per-packets (BPP) of a flow in a P2P network as its community flow statistical feature

2. Data Analytics Component (Cont.)

Enhanced PeerHunter (EPH): Community Behavior Analysis

- PeerHunter considers two numerical community features: average destination diversity ratio(**AVGDDR**) and average mutual contacts ratio (**AVGMCR**)
- For structural community feature, EPH considers each host as a vertex and link an edge between a pair of hosts when they have mutual contacts, the bots within the same botnet tend to form cliques. On the contrary, the contacts of different legitimate hosts usually diverge into different networks

2. Data Analytics Component (Cont.)

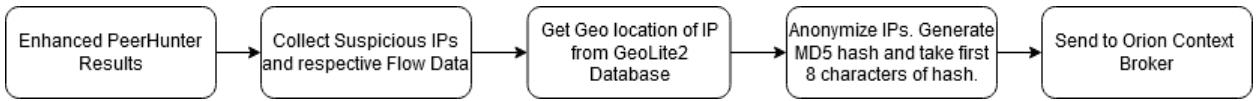
DCDNS Helper Scripts

- Besides Enhanced PeerHunter implementation, the data Analytics component has a collection of python scripts to aid in **data parsing, analysis** and **uploading** the results to FIWARE Cloud (hosted at USM)
- The scripts are scheduled to be executed everyday at **midnight (local time)** at each NREN location

2. Data Analytics Component (Cont.)

DCDNS Helper Scripts (cont.)

- Fetch the captured flow data from the database and write them to text files formatted for input to Enhanced PeerHunter
- Execute Enhanced PeerHunter Analysis
- Get all suspicious IPs and their 'neighbors' and create a list → *allFlows*
- Lookup Geo location information for them (using MaxMind's GeoLite2 DB [3])
- IP information in *allFlows* is anonymized by generating an MD5 hash.
  - e.g., 184.145.xxx.xxx ➔ d430xxxx
- *allFlows* to be sent to FIWARE Server (Component #3)



4. DCNDS Botnet Monitoring System (DBDS)

Dashboard

- The DBDS Dashboard is a web-based dashboard to visualize the results generated by all partner system
- Private dashboard available **exclusively** to DCNDS project partners to assist Network Administrators to locate specific infected hosts
- General dashboard with anonymized overview of DBDS detection statistics available to all partners
- The dashboard contains graphs, charts, geolocation and heatmaps to visualize suspicious botnet activity around the globe

References

1. https://www.ntop.org/products/packet-capture/pf_ring/pf_ring-ft-flow-table/
2. D. Zhuang and J. M. Chang, "Enhanced PeerHunter: Detecting Peer-to-Peer Botnets Through Network-Flow Level Community Behavior Analysis," in *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 6, pp. 1485-1500, June 2019, doi: 10.1109/TIFS.2018.2881657.
3. https://dev.maxmind.com/geoip/geoip2/geolite2/
4. https://github.com/telefonicaid/fiware-orion/