

OWASP IoT Top 10 based Attack Dataset for Machine Learning

Nay Myat Min*, Vasaka Visoottiviseth*, Songpon Teerakanok*, Nariyoshi Yamai**

*Faculty of Information and Communication Technology, Mahidol University, Nakhon Pathom, Thailand

**Institute of Engineering, Tokyo University of Agriculture and Technology (TUAT), Tokyo, Japan

nay.min@student.mahidol.ac.th, {vasaka.vis, songpon.tee}@mahidol.edu, nyamai@cc.tuat.ac.jp

Abstract— Internet of Things (IoT) systems are highly susceptible to cyberattacks by nature with minimal security protections. Providing a massive attack surface for attackers, they automatically become easy targets with potentially catastrophic impacts. Researchers are currently focusing on developing various anomaly detection systems for IoT networks to deal with this situation. However, these systems require a comprehensive labeled attack dataset to classify the malicious traffic correctly. This paper presents a systematic approach to design and develop an IoT testbed to generate such an attack dataset, namely the AIoT-Sol Dataset. Our proposed dataset contains the benign traffic as well as the often-overlooked attack techniques in the existing IoT datasets. It encompasses 17 attack types from several categories: network attacks, web attacks, and web IoT message protocol attacks. We selected these attacks by referencing the Open Web Application Security Project (OWASP) IoT Top Ten. Also, we provide a mapping of possible attacks for all ten security risks.

Keywords— Internet of Things, IoT, Attack Dataset, Anomaly Detection, Machine Learning, Security, OWASP IoT Top 10

I. INTRODUCTION

With the latest technological advancements, billions of Internets of Things (IoT) systems rapidly come into our everyday lives. Their ability to autonomously decide and deliver information between end-users, machine-to-machine, and machine-to-users makes them critical assets in today's world. In contrast, the rise of IoT systems integrations exposes countless new attack surfaces. They are largely vulnerable because manufacturers fail to provide basic security checks before production. Also, the developers may even reuse insecure code snippets or unverified third-party libraries due to insufficient resources and tight constraints.

We can no longer underestimate the impact of IoT systems on our daily life and avoid information security challenges. They are under constant cyberattacks since they have a large attack surface and are insecure by nature. However, due to their computational resource constraints, implementing traditional lines of defense is not feasible. Thus, IoT becomes the target of various attacks such as botnets, ransomware, crypto-jacking, advanced persistent threats (APT), distributed-denial-of-service (DDoS), identity or data theft, man-in-the-middle (MITM), and social engineering attacks. Many studies

have been conducted to address the security challenges in the IoT ecosystem with new solutions such as spam and fraud detection, malware detection, and malicious network detection using Machine Learning (ML).

Supervised or unsupervised learning algorithms can be used to train the ML models. A significant difference between them is that the former requires a labeled dataset while the latter learns by inferring from an unlabeled dataset. Therefore, training on a correctly labeled dataset is vital when using supervised learning algorithms to receive a reliable predictive model. There are several labeled datasets available from various previous research works. Nonetheless, a well-defined IoT attack dataset for anomaly detection systems is extremely rare, making it challenging to analyse and evaluate diverse methods for malicious traffic detection in IoT networks.

Researchers evaluate their proposed anomaly detection and prevention systems with attack datasets generated using automated attacking tools in most existing works. However, from a hacker's perspective, the attacking tools are the programs and scripts that only help find and exploit the computer systems, web applications, servers, and network weaknesses. Therefore, the results generated from the automation do not represent the actual security status of the target as there can be numerous false positives and negatives. Skilled hackers with malicious intentions are persistent and utilize both exploitation tools and manual techniques to attack targets until it is successfully compromised. The strength of the hacking tools depends on the skill of the operator.

Moreover, the existing datasets do not contain sufficient attack types on IoT networks and systems. Thus, evaluating the defense systems using incomplete attack datasets can produce a false sense of security, making the researchers believe that their systems are safer than they are in reality. This unintended false sense of security is harmful as it makes people be overconfident in their security mechanisms and overestimate the precautions, causing them to take unnecessary risks.

In this paper, a new IoT attack dataset is proposed. We generate our proposed labeled AIoT-Sol Dataset using our IoT testbed. This dataset includes several often-overlooked attack techniques in the IoT datasets. A total of 17 attack types are selected from the attack categories, such as network attacks, web attacks, web IoT message protocol attacks, to address the

lack of a comprehensive dataset reflecting the real-world IoT attacks. It also includes the typical attack types available in most IoT datasets, such as denial-of-service (DoS) attacks. All of these attacks are generated utilizing the most suitable tools and manually tuning them. Moreover, it contains a simulation of realistic IoT network traffic in a normal operation scenario. We will provide the dataset in its raw capture format (pcapng) and the extracted features.

The main contributions of this paper are as follows:

- We share our experience in designing the IoT testbed for AIoT-Sol Dataset Generation.
- We provide a mapping of possible attacks for each critical security risk in the OWASP IoT Top 10.
- We produce a comprehensive dataset reflecting the dynamics and characteristics of real-world attacks. In addition, our dataset includes benign traffic and more crafted attack scenarios and types than the existing datasets by referencing the Open Web Application Security Project (OWASP) IoT Top 10 - 2018.
- The selected attack types are network scanning, network logon brute force, ARP MITM attack, PPPD remote code execution, SSDP flood attack, SYN flood attack, SSL regeneration attack, web directory brute force, SQL injection, command injection, server-side request forgery, directory traversal, open redirect, XML external entity, cross-site scripting, cross-site request forgery, and MQTT brute force.

The rest of the paper is organized as follows: related works are discussed in section II and preceded in section III by our testbed design. Next, we provide the mapping of the attacks in section IV, followed by a detailed explanation of the dataset generation in section V. Then, the discussion about the generated dataset is presented in Section VI. Finally, we conclude the paper in Section VII and propose ideas for future.

II. BACKGROUND

A. Internet of Things (IoT)

IoT can be viewed as "A global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving interoperable information and communication technologies. [1]" IoT has a nature of System of Systems (SoS) in which different IoT systems are connected to produce a further added value. Generally, IoT refers to the use of sensors and communication technology embedded into devices allowing them to be tracked and controlled over the Internet. IoT management solutions require advanced communication platforms and cloud solutions that facilitate seamless integration of devices, networks, applications, and services.

B. OWASP IoT Top 10 – 2018

OWASP Internet of Things Project is developed to ensure that the developers, creators, and users have awareness and knowledge about IoT-related security issues [2]. It also aims to help consumers make better security decisions when using the Internet of Things. Table 1 defines the critical security risks for various IoT systems as listed in OWASP IoT Top 10.

TABLE 1. OWASP IoT TOP TEN SECURITY RISKS

Topic	Critical Security Risks
I1	Weak, Guessable, or Hard-coded Passwords
I2	Insecure Network Services
I3	Insecure Ecosystem Interfaces
I4	Lack of Secure Update Mechanism
I5	Use of Insecure or Outdated Components
I6	Insufficient Privacy Protection
I7	Insecure Data Transfer and Storage
I8	Lack of Device Management
I9	Insecure Default Settings
I10	Lack of Physical Hardening

C. Related Works

This section reviews related works on network-based labeled datasets for anomaly detection systems. Some existing research works for the dataset are a by-product of developing their Intrusion Prevention Systems (IPS) using ML. However, most of these datasets were not specific to IoT security.

DARPA 1998/99 [3], [4] datasets are among the most well-known datasets for anomaly detection. They contain various attack types such as denial-of-service (DoS), buffer overflow, port scans, or rootkits. KDD CUP 99 [5] is based on the DARPA and is equally popular. It contains primary attributes about TCP connections, high-level attributes such as the number of failed logins and contains more than 20 different types of attacks. However, this dataset is criticized for its redundancy. Therefore, NSL-KDD [6] comes with an enhancement by removing the duplicates. The last one to come in this DARPA family is the PU-IDS dataset. PU-IDS [7] comes with a generator capable of extracting statistical features from the dataset to generate new synthetic instances.

UNSW-NB15 [8] has both normal and malicious traffic in its dataset. The malicious one can be separated into nine classes: backdoors, DoS, exploits, fuzzer, or worms. The generation of CIC-IDS2017 [9] was done in an emulated environment for 5 days. CIC datasets encompass normal and malicious user behaviours. CIC-IDS2017 and CSE-CIC-IDS2018 [10] comprise several attack varieties like SSH brute force, Heartbleed, Botnet, DoS, Web, and Infiltration Attacks. In addition, CIC introduced a network traffic flow generator called CICFlowMeter that extracts 80 statistical features.

CIDDS-001 [11] dataset contains SSH brute force, DoS, port scan attacks, and several other attacks from an emulated small business environment for four weeks. CIDDS-002 [12] only collected the port scan attacks for two weeks. Both include normal, and attack traffics. However, they are provided as a unidirectional flow-based traffic opposite to CIC datasets which contain bi-directional flow-based traffic.

Moreover, IoTID20 [13] dataset is generated in a simulated smart home testbed consisting of SKT NGU and EZVIZ Wi-Fi cameras. It has three label features which are binary, category, and subcategory. It utilizes the CICFlowMeter to extract features. The binary contains normal and anomaly. The anomaly is further classified into DoS, Mirai, Man-in-the-middle attack (MITM), and Scan. The subcategory includes

TCP SYN Flooding, Brute Force attack, HTTP Flooding, UDP Flooding, ARP Spoofing, Host Port and OS attacks.

Lastly, MQTT-IoT-IDS2020 [14] provides a dataset with Message Queuing Telemetry Transport (MQTT) protocol-related benign or attack instances by experimenting on IDS for IoT models. It includes generic scanning, User Datagram Protocol scan, MQTT brute force, and SSH brute force attacks. Among the reviewed datasets, DARPA 98-99, PU-IDS, and UNSW-NB are available in the traditional packet-based network traffic format, while the rest except KDD CUP 99 and NSL-KDD are in the flow-based format.

III. AIOT-SOL TESTBED DESIGN

A. Testbed Logical Design

We produce the AIoT-Sol dataset using our testbed comprising multiple IoT Devices (router, raspberry pi, smart camera, smart socket, smart plug, smart phone, smart watch, and MQTT clients), MQTT broker, monitoring server with Wireshark, and an attacker machine. In addition, we run deliberately vulnerable applications like Damn Vulnerable Web Application (DVWA), OWASP WebGoat, OWASP IoTGoat, and the Metasploitable2 on the virtual machines. Since they contain Web and IoT Top 10 vulnerabilities, we can simulate various attacks against them as real-world targets.

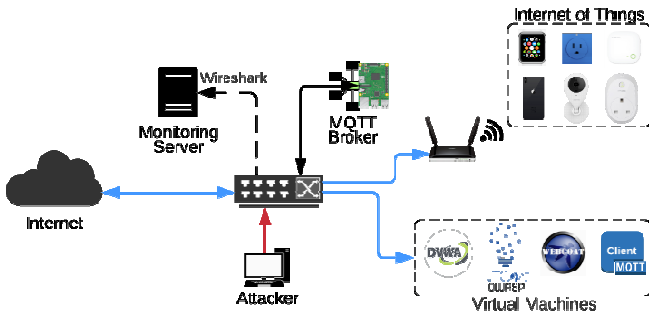


Figure 1. AIoT-Sol Testbed Design

Figure 1 illustrates the design of our AIoT-Sol Testbed. Two types of events contribute to dataset generation: traffic during the normal operations in the IoT network and traffic while the IoT devices are under attack. Firstly, the devices are reset and booted from the factory default state every first-time setup to collect the network activity during its normal operation. Next, we collect all traffic from the device and its network activity when actively communicating in the network. Devices generate network traffic corresponding to user interactions and messages communicated with other devices/servers. This baseline traffic is collected by setting up a device and leaving it in a connected state for a certain time interval. We assume each device is inherently benign; therefore, the traffic generated during standby and normal user interaction is considered benign network behaviours.

B. Test Bed Components Specifications

Table 2 and 3 show the detailed hardware and software specifications of the testbed components. We utilize the IoT devices and four deliberately vulnerable applications as our victim machines. MQTT components include 1 MQTT Broker,

TABLE 2. SPECIFICATIONS OF TESTBED PHYSICAL DEVICES

Device Name	Hardware Specifications	Software Specifications
XPS	Dell XPS 13 9370 Core i7 8550U (8CPU) ~2.0 GHz, 16 GB RAM	Windows 10 Wireshark, VMware Player 16
Monitoring Server IoTGoat-Host WebGoat-Host	HP EliteDesk 800 G1 Core i7 4790U (8CPU) ~3.6 GHz, 16 GB RAM	
DVWA Metasploitable-Host	Dell OptiPlex 7040 Core i7 6700 (8CPU) ~3.4 GHz, 16 GB RAM	
MQTT Broker (RaspberryPi3 B+)	4× ARM Cortex-A53, 1.2GHz, 1 GB RAM	Pi OS, MQTT 5.0,
TP-Link Archer MR200	AC750 Wireless Dual Band 4G LTE Router	-
Cisco Switch	24 Ports Switch	Cisco IOS
Smart Socket by Lampton	Wi-Fi 2.4 GHz b/g/n WEP/WPA/WPA2-PSK	-
Smart Wi-Fi Plug HS110 by TP-Link	Wi-Fi 2.4 GHz b/g/n	-
Smart Camera by Lampton	30 FPS, 5V DC/1A 720P/VGA/QVGA	-
TP-Link Tapo C200	100-240VAC, 50/60Hz, 1080p Full HD	-
iPhone XS Max	A12 Bionic, 4GB RAM	ios-14.7
Apple Watch 3	S3 SiP dual-core	watchOS-7.6.1

TABLE 3. SPECIFICATIONS OF TESTBED VIRTUAL MACHINES

Software	Virtual Machine Specifications
DVWA	2 Cores, 8 GB, Bridged
OWASP WebGoat	
OWASP IoTGoat	
Metasploitable2	
4 Paho MQTT Python Clients	2 Cores, 4 GB, Bridged
Kali Linux (Attacker Machine)	4 Cores, 8 GB, Bridged

4 MQTT Sensors, 4 MQTT Subscribers. We simulate two-state MQTT sensors with a Python script and a simple subscriber client with mosquitto_sub. Sensors running inside Ubuntu publish with four different topics while the mosquitto_sub running inside Kali subscribe to each topic.

IV. ATTACKS MAPPING

To be able to defend the incoming attacks against the existing security risks successfully, one must know the possible attacks against each vulnerability. Table 4 lists the mapping between the possible attacks and the critical risks from the OWASP IoT Top Ten - 2018. The topics supported by the AIoT-Sol dataset are the selected attacks among I1, I2, I3, and I5. Mainly, the attack types that will generate network traffic and contribute to the dataset are chosen.

TABLE 4. POSSIBLE ATTACKS FOR EACH OWASP IoT TOP 10

No	Possible Attacks
I1	Brute force, Dictionary, Keylogger, Credential Stuffing, Password Spraying, Phishing, MITM Attack
I2	Ping Sweep, Port Scan, Packet Sniff, Traffic Analysis, DoS, Masquerade Attacks, Message Modification, Message Replay
I3	SQL Injection, Broken Authentication/Authorization, Directory Traversal, Command Injection, Logic Bug, Server-side Request Forgery (SSRF), XML External Entity (XXE), Cross-site Scripting (XSS), Cross-site Request Forgery (CSRF), WebSocket attacks
I4	Altering Boot Process, Direct Memory Access, Security Bypass, Processor Level Exploit, Firmware Update Attacks, Firmware Hijacking, Privilege Escalation
I5	Shodan searches for weaknesses, Exploit Unpatched Flaws, Use of publicly available Exploit Codes
I6	Eavesdropping, Unauthorized Access, Social Engineering, Information Distortion, Identity Theft, Sabotage
I7	Identity Theft, Fraud, Bypass SSL, Redirect Attacks, Intellectual Property Theft, Stealing Sensitive Information
I8	Exploit Unpatched Flaws, Firmware Update Attacks, Information Recovery for Enumeration,
I9	Accounts Takeover, Exploit Unpatched Flaws, Privilege Escalation, Root Level Services Exposed
I10	Physical access to physical assets to damage, disable, steal, or maliciously use the system.

V. THE PROPOSED AIOT-SOL DATASET

This section discusses the production of the benign traffic and the attack traffic reflecting the characteristics of real-world attacks for our proposed AIoT-Sol Dataset.

A. Benign Dataset Generations

We perform the following activities for two hours to generate the benign network traffic:

- 1) Generate packets from legitimate MQTT clients and server. A total of four sensors with topic names of Door-one, Door-two, Light-one, and Light-two. Door sensors publish their state of open/closed, while Light sensors publish their state of on/off. Each topic has one subscriber and is published for every 60 seconds.
- 2) Control the sensors' state change using Mosquitto Publish. This is done randomly, mimicking a real user.
- 3) Randomly browse through locally hosted websites such as DVWA, WebGoat, and IoTGoat.
- 4) Browse through Internet websites such as Wikipedia occasionally. We also download some files.
- 5) Send and receive emails using dummy accounts.
- 6) SSH login and logout from the legitimate users.
- 7) Turn on/off the smart socket and smart plug.
- 8) Rotate/View/Capture using smart cameras.
- 9) Share, Access, Edit, and Delete via Shared Network.
- 10) Video Conferencing using Messenger and MS Teams.
- 11) Video Streaming on YouTube and Netflix.

B. Attack Dataset Generation

Table 5 lists the attack types selected from Table 4, together with the relevant tools or techniques to generate them.

TABLE 5. ATTACK PLAN (1 MINUTE PER ATTACK – MULTIPLE TIMES)

Attack Types	Tools	Time	Victims
Network Scanning	Nessus, OpenVAS, Nmap	30	IoTGoat, WebGoat, Metasploitable 2, DVWA
Network Logon Bruteforce	thc-Hydra, Medusa, Metasploit	40	
ARP MITM Attack	Ettercap	20	WebGoat-Host
PPPD RCE	pptppoc.py	10	IoTGoat
SSDP Flood Attack	Saddam	30	IoTGoat, WebGoat
SYN Flood Attack	Hping3	30	
SSL Regeneration Attack	thc-ssl-dos	20	IoTGoat, WebGoat, Metasploitable 2, DVWA
Web Directory Listing	Gobuster	10	
SQL Injection	SQL Map	10	
Directory Traversal	Dotdotpwn	10	
Server-side Request Forgery	SSRFmap	10	
Command Injection	Burp Intruder	10	
Open Redirect		10	
XML External Entity		10	
Cross-site Scripting	Wfuzz	10	
Cross-side Request Forgery	XsrfProbe	10	
MQTT Bruteforce	mqttsa, mqtt-pwn	80	MQTT Broker

We execute each attack type for 1-minute multiple times. Most tools in the attack plan are available in the Kali Linux operating system (OS). However, automated scanners such as Nessus and OpenVAS are installed separately from their commercial websites. Similarly, mqttsa and mqtt-pwn require an independent installation. As for Burp Intruder, we initially install Burp Suite and run the Intruder inside Windows OS. Except the automated tools, we manually configure the tools or scripts to mimic an actual attacker.

Our AIoT-Sol dataset contains four major attack categories with several unusual attack types to contain an attack using publicly available exploit code, multiple web application-specific attack types, and an attack against IoT-specific protocol. We need these attacks in the dataset to correctly classify the attack types using ML in upcoming research.

We use CICFlowMeter [15] to extract features from the pcapng files and generate a CSV format of the AIoT-Sol dataset. We modify this tool to extract nine additional MQTT-related features making our final dataset contains 85 network statistical features and three label features. Table 6 shows the label features of our dataset in detail. The label features are:

TABLE 6. AIoT-SOL FEATURE LABELS

Binary Classification	Category Classification	Subcategory Classification
Benign	Benign	Benign
Anomaly	Network Attacks	Network Scanning Network Logon Bruteforce ARP MITM Attack PPPD Remote Code Execution
	Denial of Service Attacks	SSDP Flood Attack SYN Flood Attack SSL Regeneration Attack
	Web Attacks	Web Directory Bruteforce SQL Injection Command Injection Server-side Request Forgery Directory Traversal Open Redirect XML External Entity Cross-site Scripting Cross-site Request Forgery
	Web IoT Message Protocol Attack	MQTT Bruteforce

- 1) Binary Classification classifies the dataset as benign network traffic or anomaly network traffic.
- 2) Category Classification organizes the dataset as benign network traffic or any of the following four attack categories: Network attacks, Denial of Service attacks, Web attacks, and Web IoT Message Protocol attacks.
- 3) Subcategory Classification categorizes the dataset as benign network traffic or any of the following attacks:
 - a) *Network Attacks*: Network Scanning, Network Logon Bruteforce, ARP MITM Attack, PPPD RCE
 - b) *Denial of Service Attacks*: SSDP Flood Attack, SYN Flood Attack, SSL Regeneration Attack
 - c) *Server-side Web Attacks*: SQL Injection, Directory Traversal, Command Injection, Server-side Request Forgery, Open Redirect, and XML External Entity. Client-side Web Attacks: Cross-site Scripting and Cross-site Request Forgery
 - d) *Web IoT Message Protocol Attack*: MQTT bruteforce

VI. DISCUSSIONS

This section discusses the characteristics of the dataset generated from our proposed testbed and compares against the pre-existing datasets that we reviewed in Section II C. Table 7 shows the accumulated characteristics such as file size, number of packets, and number of flows of the captured pcapng files. Our dataset is scalable since we capture our attack traffics separately and we neglect the timestamp or protocol to avoid bias in the extracted features. In the future, we can add more modern attack types or remove the outdated attacks from this dataset. It is provided in its raw capture format (pcapng files) as well as the processed features.

TABLE 7. CHARACTERISTICS OF THE AIoT-SOL DATASET

Pcapng Files	File Size	Packets	Flows
Benign_Traffic	5.18 GB	5435132	1309920
Network Scanning	323 MB	991024	389774
Network Logon Bruteforce	664 MB	925685	27172
ARP MITM Attack	66.4 MB	102089	48632
PPPD RCE	14.8 KB	20	12
SSDP Flood Attack	3.18 GB	19033076	67941
SYN Flood Attack	9.09 GB	60150090	29839230
SSL Regeneration Attack	102 MB	199063	6000
Web Directory Listing	1.03 MB	2640410	13543
SQL Injection	238 MB	426669	8345
Command Injection	21.5 MB	82977	7442
Server-side Request Forgery	3.8 MB	15454	1435
Directory Traversal	309 MB	392692	3408
Open Redirect	36.0 MB	45564	996
XML External Entity	374 MB	496810	25750
Cross-site Scripting	516 MB	643605	2120
Cross-side Request Forgery	10.3 MB	42228	4925
MQTT Bruteforce	925 MB	391686	391686

TABLE 8. ATTACK TYPE COMPARISON AMONG DATASETS

Dataset	Attack Types
DARPA 98/99 [3][4]	DoS, Probing, Privilege Escalation (Remote to Local and User to Root)
KDD CUP 99 [5]	
NSL-KDD [6]	
PU-IDS [7]	
UNSW-NB15 [8]	backdoors, DoS, exploits, fuzzers, port scans, recon, shellcode, spam, worms
CIC-IDS2017 [9]	botnet, XSS, DoS, DDoS, heartbleed, infiltration, SSH bruteforce, SQLi
CSE-CIC-IDS2018 [10]	botnet, bruteforce, port scan, DDoS, DoS, web attack, infiltration attack
CIDDS-001 [11]	DoS, Ping & SYN scan, SSH bruteforce
CIDDS-002 [12]	{ACK, FIN, UDP, SYN, Ping} Scans
IoTID20 [13]	Syn Flooding, Brute Force, HTTP Flooding, UDP Flooding, ARP Spoofing, Host Port, OS Scan
MQTT-IoT-IDS2020 [14]	Aggressive scan, UDP scan, MQTT brute force, SSH brute force
AIoT-Sol	MITM, PPPD RCE, SSDP flood, SYN flood, SSL regeneration, directory brute, directory traversal, command injection, open redirect, SQLi, XXE, XSS, SSRF, CSRF, MQTT brute force

Table 8 compares the attack types consisted in each dataset. AIoT-Sol dataset contains more various attack types than others. However, our dataset does not contain botnet attacks, even though it is one of the most common IoT-related attacks because it is readily available in other datasets. Note that we do not aim to replace the existing datasets; instead, we intend

to incorporate more Web and IoT-specific attacks that are not available yet in existing datasets. Although some attack types may overlap with the existing ones, we have performed more crafted attack scenarios for our dataset to be more realistic and comprehensive. As a result, researchers will be able to evaluate their IoT security mechanism using our dataset together with datasets of their choice.

VII. CONCLUSIONS

This paper presented a systematic approach to designing a realistic IoT testbed incorporating physical IoT devices, gateways, switches, computers, and virtual machines configured to interact in several attacking scenarios to generate a new attack dataset for anomaly detection in IoT networks. We have also mapped attacks for the critical security risk in the OWASP IoT top 10. Several relevant attack types are selected to produce a comprehensive dataset that matches the real-world attack scenarios. Our proposed AIoT-Sol dataset contains a total of 17 attack types to evaluate anomaly detection algorithms in the IoT network. We have shared this dataset at [16]. Our main contribution is the new dataset to complement the existing network datasets with more attack types to help evaluate various ML algorithms.

In the future, we wish to generate a new IoT attack dataset comprising a wider range of attack types. Furthermore, we plan to utilize this AIoT-Sol dataset to develop a system capable of detecting and filtering malicious IoT network traffic using Machine Learning. Note that this research project was partially supported by Faculty of Information and Communication Technology, Mahidol University.

REFERENCES

- [1] S. Villamil, C. Hernández, and G. Tarazona, "An overview of internet of things," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 18, no. 5, pp. 2320–2327, 2020.
- [2] OWASP, "OWASP Top 10 Internet of Things," 2018. [Online]. Available: https://wiki.owasp.org/index.php/OWASP_Internet_of_Things_Project#tab=IoT_Top_10
- [3] R. P. Lippmann et al., "Evaluating intrusion detection systems: The 1998 DARPA off-line intrusion detection evaluation," *Proceedings - DARPA Information Survivability Conference and Exposition, DISCEX 2000*, vol. 2, pp. 12–26, 2000, doi: 10.1109/DISCEX.2000.821506.
- [4] R. Lippmann, J. W. Haines, D. J. Fried, J. Korba, and K. Das, "The 1999 DARPA off-line intrusion detection evaluation," *Computer Networks*, vol. 34, no. 4, pp. 579–595, Oct. 2000, doi: 10.1016/S1389-1286(00)00139-0.
- [5] "KDD Cup 1999 Data." <http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html> (accessed Sep. 28, 2021)
- [6] M. Tavallaei, E. Bagheri, W. Lu, and A. A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set A Detailed Analysis of the KDD CUP 99 Data Set," 2009.
- [7] R. Singh, H. Kumar, and R. K. Singla, "A reference dataset for network traffic activity based intrusion detection system," *International Journal of Computers, Communications and Control*, vol. 10, no. 3, pp. 390–402, 2015, doi: 10.15837/ijccc.2015.3.1924.
- [8] N. Moustafa and J. Slay, "UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," 2015 Military Communications and Information Systems Conference, MilCIS 2015 - Proceedings, Dec. 2015.
- [9] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," *ICISSP 2018 - Proceedings of the 4th International Conference on Information Systems Security and Privacy*, vol. 108–116, 2018, doi: 10.5220/0006639801080116.
- [10] "IDS 2018 | Datasets | Canadian Institute for Cybersecurity | UNB." <https://www.unb.ca/cic/datasets/ids-2018.html> (accessed Sep. 28, 2021)
- [11] M. Ring, S. Wunderlich, D. Grödl, D. Landes, and A. Hotho, "Flow-based benchmark data sets for intrusion detection," undefined, 2017.
- [12] "(PDF) Creation of Flow-Based Data Sets for Intrusion Detection." https://www.researchgate.net/publication/322337341_Creation_of_Flow-Based_Data_Sets_for_Intrusion_Detection (accessed Sep. 28, 2021).
- [13] I. Ullah and Q. H. Mahmoud, A Scheme for Generating a Dataset for Anomalous Activity Detection in IoT Networks, vol. 12109 LNAI. Springer International Publishing, 2020. doi: 10.1007/978-3-030-47358-7_52.
- [14] H. Hindy, E. Bayne, M. Bures, R. Atkinson, C. Tachtatzis, and X. Bellekens, "Machine Learning Based IoT Intrusion Detection System: An MQTT Case Study (MQTT-IoT-IDS2020 Dataset)," *Lecture Notes in Networks and Systems*, vol. 180, pp. 73–84, 2021, doi: 10.1007/978-3-030-64758-2_6.
- [15] "Applications | Research | Canadian Institute for Cybersecurity | UNB." <https://www.unb.ca/cic/research/applications.html#CICFlowMeter> (accessed Sep. 29, 2021).
- [16] "AIoT-Sol." <https://github.com/NayMyatMin/AIoT-Sol> (accessed Dec. 28, 2021)



Nay Myat Min is a research student at Mahidol University, Thailand. He is an M.Sc. Degree in Cyber Security and Information Assurance candidate. He received his B.Sc. in Computer and Network Technology from Northumbria University in 2018. Since then, he has worked as a security professional in the Banking Industry. His prior work experience includes penetration testing and risk assurance. His research interests are offensive security, IoT security, and machine learning.



Vasaka Visoottiviseth is an associate professor at Mahidol University, Thailand. She received her Ph.D. degree in computer engineering from Nara Institute of Science and Technology (NAIST), Japan in 2003, M.E. and B.E. degree from Tokyo University of Agriculture and Technology (TUAT), Japan in 1999 and 1997, respectively. Her current research interests are mobile and wireless computing, Internet traffic measurement, and network security. She has been a member of IEEE since 1998.



Songpon Teerakanok received his B.E. from Prince of Songkla University, Thailand in 2013, and M.E. and D.Eng. degrees in Information Science and Engineering from Ritsumeikan University in 2016 and 2019, respectively. He was a former assistant professor at Ritsumeikan University before joining the Faculty of ICT, Mahidol University, Thailand in May 2021. His research interest covers Cryptography, Privacy, Location-based Service (LBS), and Digital Forensics.



Nariyoshi Yamai received the B.E. and M.E. degrees in electronic engineering and the Ph.D. degree in information and computer science from Osaka University in 1984, 1986, and 1993, respectively. After working at Nara National College of Technology, Osaka University, and Okayama University, he has been a Professor at the Institute of Engineering, Tokyo University of Agriculture and Technology since April 2014. His research interests include distributed systems, network architecture, and Internet. He is a member of IEEE CS, IEEE ComSoc, IEICE and IPSJ.