Industrial Placement Proposal

Exploring Multiple IoT Security Attacks with Machine Learning Based Schemes

Hanlin Cai

Supervisor: Prof. Zhezhuang Xu

Fujian Huading Intelligent Manufacturing Technology Co. LTD

Maynooth International Engineering College, Fuzhou University

February 25th, 2023

CONTENTS

1. INTRODUCTION

Nowadays, the internet of things (IoT) has grown into a global giant, grabbing hold of every facet of our everyday lives and benefiting people with its unrestricted intelligent technology. The IoT integration of the real world with computer networks and software applications such as health monitoring and home automation makes personal privacy and security techniques critical for future IoT development. However, due to the ease of access, lack of restrictions and fast-growing demand for more facilities, IoT systems can be vulnerable to collapse in the face of some massive attacks, such as spoofing, denial of service (DoS) and intrusion attacks.[1]

In order to maintain the security requirements of IoT systems, machine learning (ML) based security schemes could be a promising alternative. ML schemes can be applied to train IoT system to recognize specific attacks and adopt the corresponding defense strategies. Additionally, ML algorithms would be useful to predict future attacks, which are usually mutations of previous attacks, through analyzing and learning from the existing data. Therefore, ML-based security schemes are important for us to address the challenging attacks in real-world IoT security tasks.

2. LITERATURE REVIEW

In this section, we will first go through key concepts, including the IoT security attack, machine learning, and review previous literature studies and then analyze the gap in existing knowledge based on the review.

2.1 Key Concepts, Theories and Studies

[1] IoT Security Attack

The IoT security attack is one of the most crucial challenges in IoT research, Sengupta, et al.[2] have broadly divided the IoT attacks into the following four domains, as shown in Fig.1.

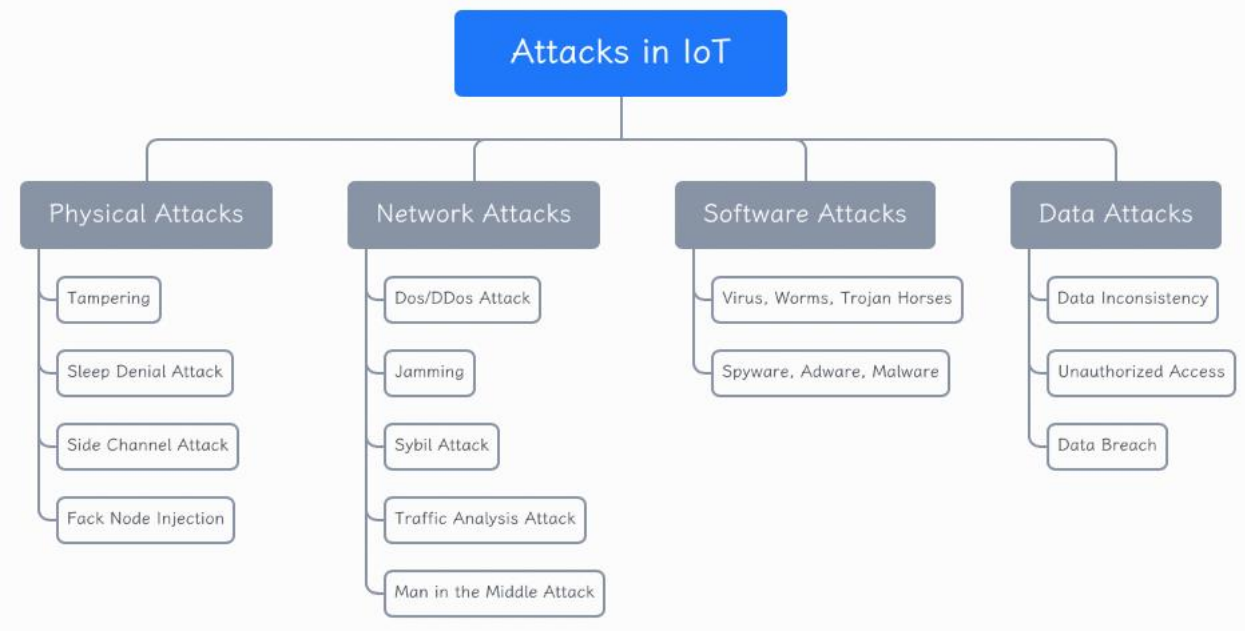

FIG 1 ATTACKS IN IOT

Most recently, many literature studies have discussed various IoT security attacks and corresponding defensive schemes. For instance, Vishwakarma, et al.[3] discussed the principle of botnets and malware being deployed to 'Distributed' DoS attacks and proposed a more dependable DDoS defense technique to recognize the security gaps. Arshad, et al.[4] concentrated on the detection of malicious Sybil nodes and proposed the THC-RPL scheme, which significantly reduced the packet loss rate while maintaining lower power consumption. Besides, there were many papers focusing on the potential security challenges of IoT systems.[5-7] However, most of these works addressed only specific solutions to specific security attacks. So far, very few articles have proposed a general approach to defend against multiple IoT attacks.

[2] Machine Learning

Machine learning is a branch of artificial intelligence (AI) that focuses on optimizing model performance and effectiveness through learning existing data and previous experience. Generally speaking, ML algorithms can be classified into the following four types, as shown in the following Fig.2. ML methods do not need explicit programming and possess outstanding performance in dynamic networks, which are very suitable for maintaining the security of IoT facilities without continuous human supervision.

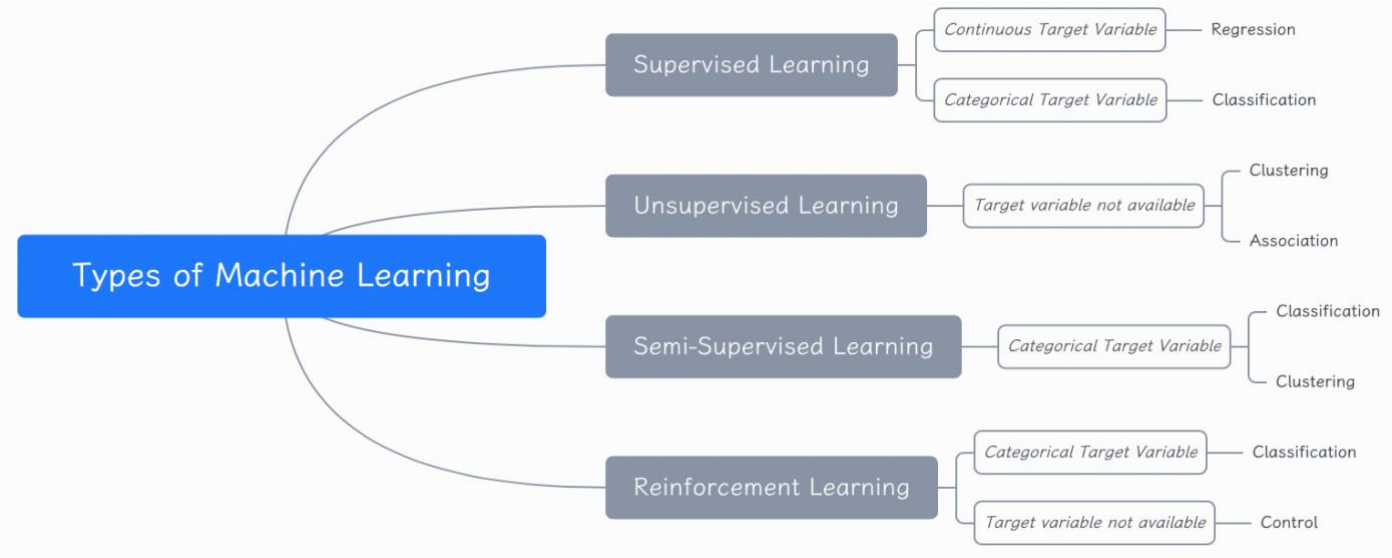

FIG 2 TYPES OF MACHINE LEARNING

As illustrated in Table 1, many papers have proposed various ML-based methods to prevent specific IoT attacks and improve IoT security. [3, 4, 8-14] Among them, the Q-learning schemes proposed by Xiao, et al.[8] performed well in the face of both spoofing and jamming attacks, while the SVM schemes proposed by Ozay, et al.[9] could effectively identify and defend against intrusion and spoofing attacks. Besides, it is worth noting that the intelligent architecture combining CEP and ML proposed by Roldán, et al.[15] can manage dynamic patterns for identifying IoT security threats, which is one of the most advanced security schemes.

TABLE 1 IOT SECURITY SCHEMES.

Attacks	Security Schemes	ML Methods	Performance
Spoofing	Authentication	Q-learning[8]	Average loss rate
	Authentication	SVM[9]	Classification accuracy
	Authentication	DNN[10]	False alarm rate
	Authentication	dFW[11]	Misdetection rate
DoS	Secure IoT offloading	MLP[12]	Detection accuracy
	Access Control	MCA[13]	Root mean error
	Flow Detection	NFS[3]	Storage efficiency
Intrusion	Access Control	Naive Bayes[14]	False alarm rate
	Access Control	SVM[9]	Classification accuracy
Sybil	Dual Identity	THC-RPL[4]	Power consumption
Jamming	Secure IoT offloading	Q-learning[8]	Energy consumption

2.2 Gap in Existing Knowledge

According to the literature exploration above, one of the most promising solutions to IoT security attacks is ML-based security schemes (e.g. SVM[9], CEP-ML architecture[15]) to identify and defend against specific IoT threats such as spoofing, DoS and intrusion attacks. (Fig.3) However, most of the solutions proposed by these studies could only address specific security attacks and would be unable to define more patterns for detecting dynamic multiple attacks. Inspired by Hussain, et al.[5] and Ahmad, et al.[16], it is feasible for us to explore a hybrid defense scheme on the strength of the advanced ML-based security algorithms.

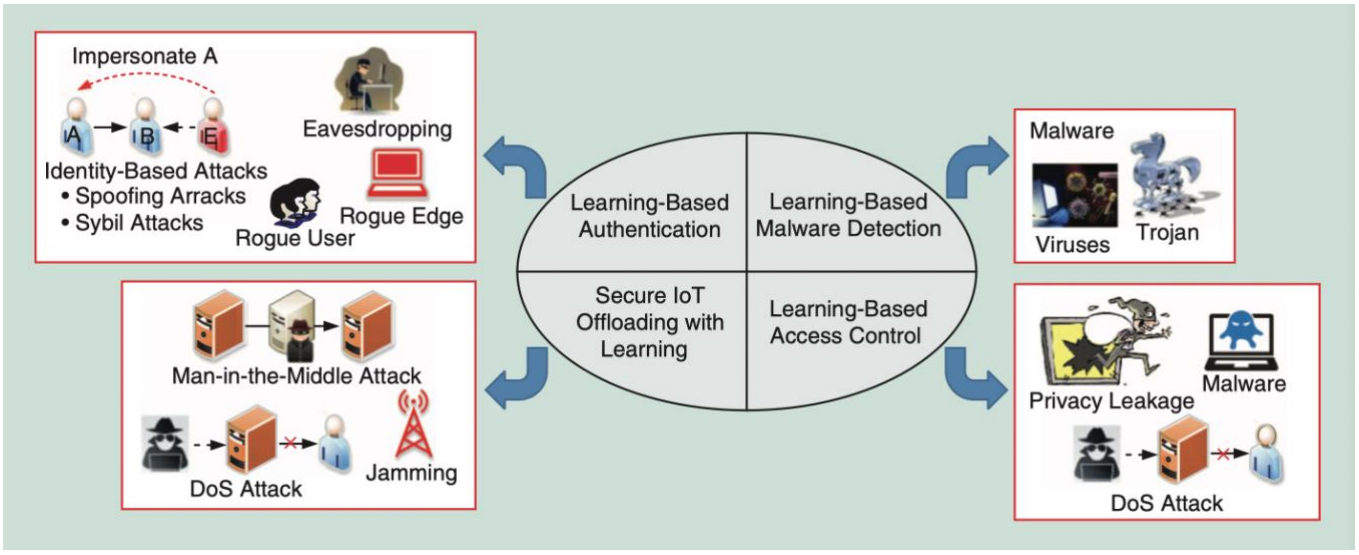

FIG 3 LEARNING-BASED SECURITY SCHEMES IN THE IOT. [7]

3. RESEARCH QUESTIONS

After analyzing the background and existing gaps, in order to address the research problem scientifically and effectively, I have divided it into the specific questions listed in Table 2.

TABLE 2 SPECIFIC QUESTIONS AND OBJECTIVE

	Question	Objective
RQ1	What are the common security vulnerabilities of current IoT systems and the principles of specific security attacks, such as spoofing, denial of service (DoS) and intrusion attacks?	To understand different security requirements, challenges, and need to secure IoT systems from hostile and massive attacks.
RQ2	What is machine learning based security scheme and the state-of-the-art algorithms for preventing specific IoT attacks, such as spoofing attacks and DoS attacks?	To identify vastly adopted and dependable ML-based methods by existing research to protect IoT systems from specific attacks.
RQ3	What is multiple IoT attacks and how can we combine the advantages of various ML-based security methods to defend against multiple IoT attacks?	To analyze the complicated security attacks in the real-operating environment of the IoT systems and integrate the most advanced defense schemes to improve the security.

4. RESEARCH METHODOLOGY

4.1 Aim and Objectives

The main goal of my research is to explore a dependable hybrid IoT defense scheme with state-of-the-art ML-based security algorithms to secure IoT systems from multiple IoT attacks, such as spoofing, denial of service (DoS) and intrusion attacks.

4.2 Methods and Sources

To achieve the research aim, first of all, it requires an in-depth literature review of the current development in the field of IoT security, the fundamentals of various IoT security attacks and the different methods researchers proposed to defend the specific attacks. Secondly, in consideration of the ML-based methods will demand a variety of training data and extensive experiments to ensure the efficiency and effectiveness of the ML algorithms, I would like to adopt quantitative and experimental research methods, which systematically intervene in the training process and evaluate the accuracy and precision of each security scheme. The following presents my fundamental research methodology implementation flow:

[1] Testbed Establishment

Due to the special characteristics of the IoT attack testbed, thorough model building and training for intrusion detection will be necessary. Therefore, I utilize FACT[1] and Stratosphere Lab[2] open-source projects to test multiple simulated IoT attack detection scenarios. As shown in Fig.4, the DDoS detection pipeline[17] available on Stratosphere Lab is a suitable test framework for evaluating the accuracy and precision of various DDoS classification algorithms, which would be useful to modify and improve the performance of classifiers.

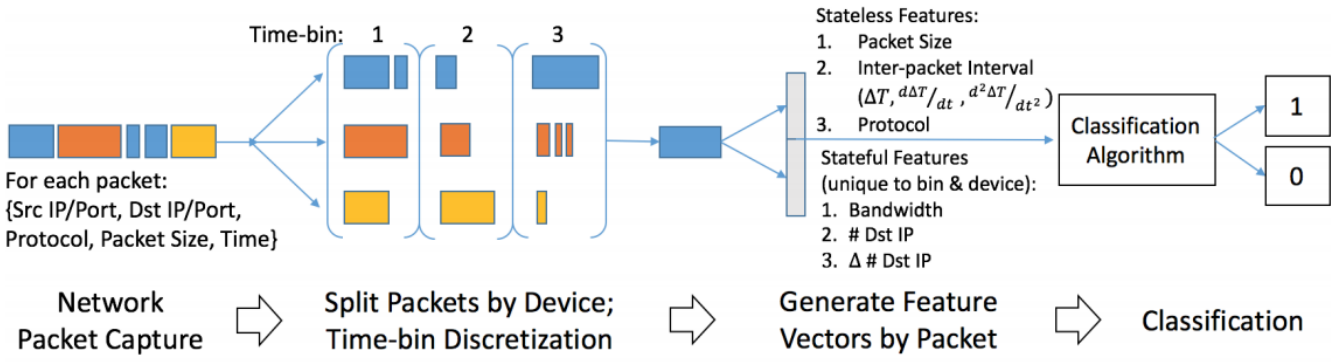

FIG 4 IOT DDOS DETECTION PIPELINE[17]

[2] Training Datasets Collection

It is significant and fair to train different ML-based security models on an unbiased dataset that is IoT specific and involves various types of attack traffic. Since IoT-related companies rarely release the primary data for the reasons of confidentiality laws and user privacy Restrictions, I adopt some of the reliable secondary datasets, as illustrated in Table 3. [18-22] However, instead of using all of the following datasets, I would select the most appropriate one based on the peer review and specific circumstances.

TABLE 3 DATASET ANALYSIS

Dataset (Year)	Benign Records	Malicious Records	Description
Bot-IoT (2018)[18]	8,892	71,342,700	Appropriate Labels
*CICIDS (2017)[19]	2,263,123	567,601	Unique, Complex Features
UNSW (2015)[20]	581,232	79,322	Large-scale DDoS attacks
*SCADA (2019)[21]	32,789	89,374	Full Packet Capture
IoT-23 (2020)[22]	30,854,735	298,490,308	Recent, but without HTTPS

The asterisked (*) datasets are the alternatives.

FACT[1]: https://fkie-cad.github.io/FACT_core/ Stratosphere Lab[2]: https://www.stratosphereips.org/

[3] ML Model Training

Based on the in-depth literature review and inspired by Ahmad and Alsmadi[16], I intend to utilize and integrate advanced ML algorithms, such as Random Forest, Naive Bayes, Support Vector Machine (SVM), distributed Frank-Wolfe (dFW), and K-Nearest Neighbors (K-NN).

At present, I have carried out a preliminary test for these ML models and obtained some elementary results. As shown in Table 4[9, 11, 23-25], the Random Forest model got the best accuracy of 99.2%, and the combination of SVM and dFW models achieved the highest F1-Score[26] of 97.2%, which represented the good balance between precision and recall. And it should be noted that, due to the limited time, this preliminary test was based on a small dataset, which may be biased and lack sufficient scenarios. In the future, I will adopt some larger and more reliable datasets (as shown in Table 3) to ensure the validity of the ML model training.

TABLE 4 PRELIMINARY MODEL EVALUATION

ML Model	Accuracy	Precision	Recall	F1-Score
SVM[9]	97.6%	94.9%	98.2%	96.5%
K-NN[23]	94.4%	92.0%	100%	96.0%
Random Forest[25]	99.2%	/	98.2%	/
GNB[23]	87.1%	/	90.7%	/
SVM[24]+dFW[11]	96.2%	95.2%	99.3%	97.2%
	$\frac{CI}{TI} \times 100\%$	$\frac{TP}{TP + FP}$	$\frac{TP}{TP + FN}$	$2 \times \frac{Precision \times Recall}{Precision + Recall}$

CI: Correctly classified intrusion; TI: Total number of inputs.

TP: The number of true positives; FP: The number of false positives; FN: The number of false negatives.

[4] Training and Data Analysis

As illustrated in Table 4, the main criteria for evaluating the efficiency and effectiveness of ML security models are accuracy, precision, recall, and F1-Score. After model training, I will collect and analyze the experimental results by using Excel forms, confusion matrices [27] and ROC curves [26] as the following steps:

- Organize experimental data results in Excel form and calculate statistical criteria (Table 4).
- Utilize TensorFlow, MATLAB and MMDetection to visualize and compare the confusion matrices between different ML classification algorithms. (Fig.5)
- Utilize Python packages such as Pyplot, Matplotlib or Tensorboard to plot and compare the ROC curves of different ML security models. (Fig.6)

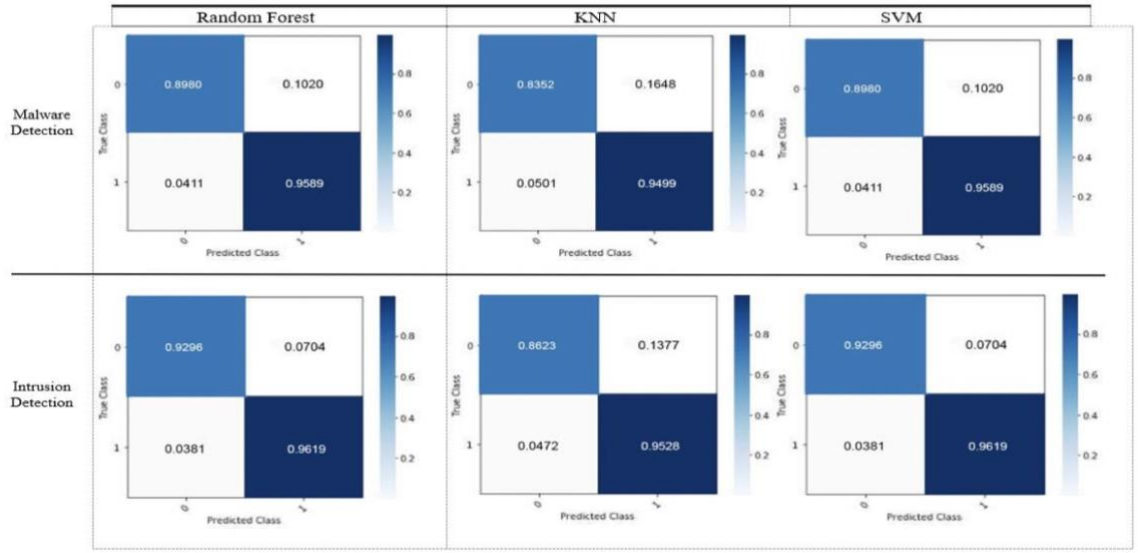

FIG 5 CONFUSION MATRIX COMPARISON

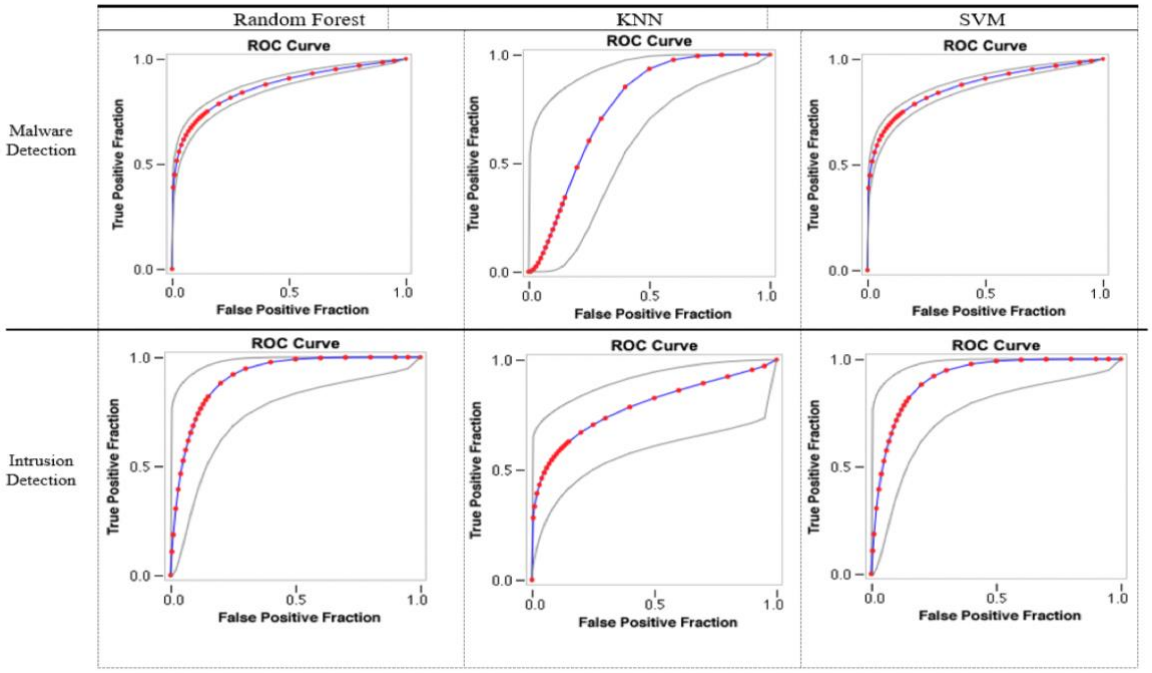

FIG 6 ROC CURVE COMPARISON

The confusion matrix is a widely used technique for summarizing and comparing the accuracy of various classification algorithms. Fig.5 shows the confusion matrix comparison between three different classification algorithms. It indicates which models are getting higher accuracy and what types of mistakes they are making.

The ROC curve is a very useful tool for visualizing and evaluating classifier performance. Fig.6 illustrates the ROC curve comparison of three different classifiers, which calculates the ratio between the true positive (TP) and false positive (FP) rates. Higher ROC values determine that the classifiers can successfully distinguish between TP and FP instances in the dataset. By comparing the confusion matrix and ROC curve of different ML security models, we can know how to modify and improve the algorithms further.

4.3 Advantages and Potential Obstacles

After presenting my methodology implementation flow, I would like to discuss the advantages and potential obstacles of my research. As to the advantages, my methodology is designed by taking advantage of existing advanced studies, reliable datasets, and efficient visualization tools to address the research questions step by step.

As to the potential obstacles, firstly, how to make my training datasets fairer and more scenario-rich would be a potential challenge. Secondly, more expertise and GPU resources would be required to support my further research. Finally, time is always the biggest obstacle. How to advance my research efficiently and scientifically in the next three months would greatly influence my results.

5. EXPECTED RESULT

There are still a variety of technologies to be discovered in the area of ML-based security methods and they have promising application prospects in the field of IoT security. I expect to utilize the research achievement of the IoT security schemes to explore approaches to realize a more privacy conscious and reliable IoT system, which would have a promising publication opportunity and broad market prospect.

6. TIME SCHEDULE

In order to advance my research scientifically and effectively, I developed a specific research schedule, as shown in Fig.7. And to be honest, I understand the research cannot always proceed smoothly as planned, so I will modify the schedule according to the actual situation.

Besides, my IP Record can be found here: https://github.com/GuangLun2000/Intern-2023

Industrial Placement Schedule

Project Lead: Zhezhuang Xu, Hanlin Cai

WBS	Task	Priority	Resource	Start	Finish	Duration	Done	% Complete
1	**In-depth literature review**	NORMAL	**FZU, Hanlin Cai**	**Fri 10-Feb-23**	**Fri 24-Feb-23**	**12**		**60%**
1.1	Go through the advanced paper	NORMAL	**FZU, Hanlin Cai**	Fri 10-Feb-23	Tue 21-Feb-23	8		80%
1.2	Repeat the experiments	LOW	**FZU, Hanlin Cai**	Mon 20-Feb-23	Thu 23-Feb-23	4		60%
1.3	Explore the gaps of existing methods	HIGH	**FZU, Hanlin Cai**	Fri 24-Feb-23	Fri 24-Feb-23	0		40%
2	Redesign some advanced ML algorithms	NORMAL	**FZU, Hanlin Cai**	Fri 24-Feb-23	Wed 29-Mar-23	24		20%
3	Try to integrate different security modelsd	LOW	**FZU, Hanlin Cai**	Fri 10-Mar-23	Fri 31-Mar-23	16		10%
4	**Comprehensive experiments**	NORMAL	**FZU, Hanlin Cai**	**Sat 01-Apr-23**	**Mon 24-Apr-23**	**23**		**2%**
4.1	Collect the data	NORMAL	**FZU, Hanlin Cai**	Sat 01-Apr-23	Thu 20-Apr-23	14		5%
4.2	Contrast various results	NORMAL	**FZU, Hanlin Cai**	Wed 12-Apr-23	Mon 24-Apr-23	9		2%
4.3	Analyze and comment on the experiments	HIGH	**FZU, Hanlin Cai**	Mon 24-Apr-23	Mon 24-Apr-23	0		0%
5	Report writing and revise (Repeatedly)	HIGH	**Hanlin Cai**	Thu 04-May-23	Wed 24-May-23	21		0%
6	**Final demo slides perpared for the research defence**	NORMAL	Hanlin Cai	Sat 20-May-23	Sun 28-May-23	9		0%
7	**Final research report revision and submission**	HIGH	Hanlin Cai	Tue 30-May-23	Mon 12-Jun-23	14		0%

FIG 7 TIME SCHEDULE FOR RESEARCH

REFERENCE

[1] Tahsien S M, Karimipour H, et al. Machine learning based solutions for security of Internet of Things (IoT): A survey [J]. Journal of Network and Computer Applications, 2020, 161.

[2] Sengupta J, Ruj S, et al. A Comprehensive Survey on Attacks, Security Issues and Blockchain Solutions for IoT and IIoT [J]. Journal of Network and Computer Applications, 2020, 149.

[3] Vishwakarma R, Jain A K. A survey of DDoS attacking techniques and defence mechanisms in the IoT network [J]. Telecommunication Systems, 2019, 73(1): 3-25.

[4] Arshad D, Asim M, et al. THC-RPL: A lightweight Trust-enabled routing in RPL-based IoT networks against Sybil attack [J]. PLoS One, 2022, 17(7): e0271277.

[5] Hussain F, Hussain R, et al. Machine learning in IoT security: Current solutions and future challenges [J]. IEEE Communications Surveys & Tutorials, 2020, 22(3): 1686-1721.

[6] Mohamad Noor M b, Hassan W H. Current research on Internet of Things (IoT) security: A survey [J]. Computer Networks, 2019, 148: 283-294.

[7] Xiao L, Wan X, et al. IoT Security Techniques Based on Machine Learning: How Do IoT Devices Use AI to Enhance Security? [J]. IEEE Signal Processing Magazine, 2018, 35(5): 41-49.

[8] Xiao L, Li Y, et al. PHY-Layer Spoofing Detection With Reinforcement Learning in Wireless Networks [J]. IEEE Transactions on Vehicular Technology, 2016, 65(12): 10037-10047.

[9] Ozay M, Esnaola I, et al. Machine Learning Methods for Attack Detection in the Smart Grid [J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 27(8): 1773-1786.

[10] Shi C, Liu J, et al. Smart User Authentication through Actuation of Daily Activities Leveraging WiFi-enabled IoT [Z]. Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing. Chennai, India; Association for Computing Machinery. 2017: Article 5.10.1145/3084041.3084061

[11] Xiao L, Wan X, et al. PHY-Layer Authentication With Multiple Landmarks With Reduced Overhead [J]. IEEE Transactions on Wireless Communications, 2018, 17(3): 1676-1687.

[12] Kulkarni R V, Venayagamoorthy G K. Neural network based secure media access control protocol for wireless sensor networks; proceedings of the 2009 International Joint Conference on Neural Networks, F 14-19 June 2009, 2009 [C].

[13] Tan Z, Jamdagni A, et al. A System for Denial-of-Service Attack Detection Based on Multivariate Correlation Analysis [J]. IEEE Transactions on Parallel and Distributed Systems, 2014, 25(2): 447-456.

[14] Alsheikh M A, Lin S, et al. Machine Learning in Wireless Sensor Networks: Algorithms, Strategies, and Applications [J]. IEEE Communications Surveys & Tutorials, 2014, 16(4): 1996-2018.

[15] Roldán J, Boubeta-Puig J, et al. Integrating complex event processing and machine learning: An intelligent architecture for detecting IoT security attacks [J]. Expert Systems with Applications, 2020, 149.

[16] Ahmad R, Alsmadi I. Machine learning approaches to IoT security: A systematic literature review [J]. Internet of Things, 2021, 14.

[17] Doshi R, Apthorpe N, et al. Machine Learning DDoS Detection for Consumer Internet of Things Devices [Z]. 2018 IEEE Security and Privacy Workshops (SPW). 2018: 29-35.10.1109/spw.2018.00013

[18] Koroniotis N, Moustafa N, et al. Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics: Bot-IoT dataset [J]. Future Generation Computer Systems, 2019, 100: 779-796.
[19] Chaabouni N, Mosbah M, et al. Network Intrusion Detection for IoT Security Based on Learning Techniques [J]. IEEE Communications Surveys & Tutorials, 2019, 21(3): 2671-2701.
[20] Yavanoglu O, Aydos M. A review on cyber security datasets for machine learning algorithms; proceedings of the 2017 IEEE International Conference on Big Data (Big Data), F 11-14 Dec. 2017, 2017 [C].
[21] Zolanvari M, Teixeira M A, et al. Machine Learning-Based Network Vulnerability Analysis of Industrial Internet of Things [J]. IEEE Internet of Things Journal, 2019, 6(4): 6822-6834.
[22] Sebastian Garcia A P, & Maria Jose Erquiaga. (2020). IoT-23: A labeled dataset with malicious and benign IoT network traffic (Version 1.0.0) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.4743746.
[23] Kumar A, Lim T J. EDIMA: Early Detection of IoT Malware Network Activity Using Machine Learning Techniques; proceedings of the 2019 IEEE 5th World Forum on Internet of Things (WF-IoT), F 15-18 April 2019, 2019 [C].
[24] Ioannou C, Vassiliou V. Classifying Security Attacks in IoT Networks Using Supervised Learning [Z]. 2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS). 2019: 652-658.10.1109/dcoss.2019.00118
[25] Hasan M, Islam M M, et al. Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches [J]. Internet of Things, 2019, 7.
[26] Fawcett T. An introduction to ROC analysis [J]. Pattern Recognition Letters, 2006, 27(8): 861-874.
[27] Luque A, Carrasco A, et al. The impact of class imbalance in classification performance metrics based on the binary confusion matrix [J]. Pattern Recognition, 2019, 91: 216-231.

ABOUT THE AUTHOR

Mr. Hanlin Cai is a junior majoring in Automation at Fuzhou University (China) and Robotics and Intelligent Devices at Maynooth University (Ireland, Combined Degrees). His research interests are on the Machine Learning and its applications in Industrial IoT. Web: https://caihanlin.com/

Dr. Zhezhuang Xu is currently a Professor with the School of Electrical Engineering and Automation, Fuzhou University, Fuzhou, China. He received the Ph.D. degree in control science and engineering from Shanghai Jiao Tong University, Shanghai, China, in 2012. Web: https://dqxy.fzu.edu.cn/info/1102/3547.htm