Machine Learning for Intrusion Detection: Stream Classification Guided by Clustering for Sustainable Security in IoT

Martin Manuel Lopez
martinmlopez@arizona.edu
University of Arizona
Tucson, Arizona, USA

Sicong Shao
sicongshao@arizona.edu
University of Arizona
Tucson, Arizona, USA

Salim Hariri
hariri@arizona.edu
University of Arizona
Tucson, Arizona, USA

Soheil Salehi
ssalehi@arizona.edu
University of Arizona
Tucson, Arizona, USA

ABSTRACT

The Internet of Things (IoT) has brought about unprecedented connectivity and convenience in our daily lives, but with this newfound interconnectedness comes the threat of cyber-attacks. With ever-increasing IoT devices being connected to the internet, securing IoT devices is becoming increasingly urgent. Machine learning (ML) is among the most popular techniques used by intrusion detection systems (IDS) to enhance their detection performance when securing IoT. However, a key obstacle of ML-based IDS for IoT is learning from nonstationary streaming data, also known as concept drift. One of the most challenging learning scenarios under concept drift is extreme verification latency (EVL), which occurs when only unlabeled nonstationary streaming data is available after a small set of initial labeled data. Stream Classification Algorithm Guided by Clustering (SCARGC) is an algorithm that can effectively deal with the nonstationary data streams in EVL scenarios. Applying an EVL implementation provides the capability of adapting to nonstationary environments within the IoT domain. The SCARGC model, as an integrated IoT intrusion detection system, allows for sustainable security as new threats are identified in this non-stationary environment. Hence, in this project, we develop an innovative IoT intrusion detection approach by natively integrating SCARGC and intrusion detection to address the EVL challenges to provide sustainable security as the model adapts to nonstationary environments. We evaluated the proposed approach on real-world IoT cybersecurity datasets. The results demonstrate the feasibility of the proposed approach, which can lead to the development of sophisticated intrusion detection systems for IoT.

CCS CONCEPTS

• **Security and privacy → Intrusion detection systems**; • **Computing methodologies → Machine learning**.

GLSVLSI '23, June 5–7, 2023, Knoxville, TN, USA

ACM ISBN 979-8-4007-0125-2/23/06.
https://doi.org/10.1145/3583781.3590271

KEYWORDS

IoT Security, Extreme Verification Latency, Nonstationary Environments, Datastream, Machine Learning for Security, Intrusion Detection Systems

ACM Reference Format:
Martin Manuel Lopez, Sicong Shao, Salim Hariri, and Soheil Salehi. 2023. Machine Learning for Intrusion Detection: Stream Classification Guided by Clustering for Sustainable Security in IoT. In *Proceedings of the Great Lakes Symposium on VLSI 2023 (GLSVLSI '23), June 5–7, 2023, Knoxville, TN, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3583781.3590271

1 INTRODUCTION

The Internet of Things (IoT) is rapidly expanding and becoming an integral part of our daily lives, with a growing number of connected devices and appliances communicating and exchanging data over the Internet. The increasing demand for IoT technology has resulted in a proliferation of connected devices, thus escalating the security risks in IoT [2, 6–8]. It will be increasingly rewarding for attackers to compromise IoT systems as there will be greater rewards for successfully breaching IoT systems. Furthermore, without proper security for the data processed/generated by these IoT devices, the data may be stolen for financial gain or, worst, threaten people's lives [10, 15, 18]. Naturally, the public's adoption of IoT systems is exploited and abused by cybercriminals using sophisticated and advanced hacking methods such as Botnets.

Intrusion detection systems (IDS) are one of the promising security solutions which can be used to protect IoT systems. Many researchers use machine learning techniques for Intrusion detection since ML-based IDS is advantageous over traditional signature-based IDS due to its capability to detect unknown and zero-day attacks. However, IoT systems' heterogeneous, dynamic, complex, and evolving features results in a non-stationary IoT data stream. Therefore, the conventional ML-based IDS for IoT trained based on the data obtained from the stationary data stream (i.e., data sampled from a fixed probability distribution) fails to address the challenge. Although several IoT intrusion detection methods have been proposed to handle the non-stationary stream, the limitations remain. One of the most critical limitations is assuming the existence of class labels immediately or with some delay after the detection. Unfortunately, this assumption is often not feasible in practice because the cost of labeling the huge unlabeled amount of IoT data during the detection phase is exceptionally high. To

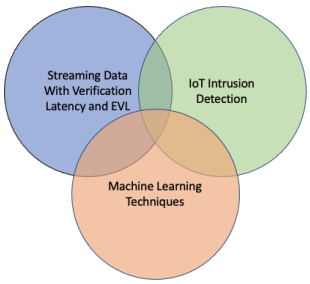

Figure 1: Commonality of EVL, IoT IDS, and Machine Learning Techniques

address this limitation, recent studies have been conducted to deal with extreme verification latency (EVL) [4, 5, 17], where the labeled data are only available at the initial step. The following data are not labeled, providing a more practical learning scenario for ML-based intrusion detection for sustainable IoT security.

The burning platform of implementing sustainable security within the IoT platform involves policies and practices prioritizing long-term solutions that counteract cyber-attacks and cyber-threats while considering the efficiency, economic, and environmental impacts of the system it tries to protect. The goals of sustainable security include prevention, resilience, and adaptability. Our research aims to address these goals as we utilize robust machine learning techniques to develop an intrusion detection system that can detect and prevent cyber-attacks by learning through previous behaviors and data. To address the adaptability, we integrate the EVL method into our detection model for learning under streaming data that entails IoT devices, including the capability of adapting to concept drift while providing a robust and resilient system. Our desire to determine if using an EVL method offers an efficient solution that can recognize cyber-attacks in a non-stationary environment allows an intrusion detection system to incorporate an adaptable machine learning technique that provides sustainable security to IoT systems.

This paper's main contribution is to design intrusion detection techniques for sustainable IoT security by proposing a novel approach containing two main modules: an ML-based detection model and an EVL method, namely Stream Classification Algorithm Guided by Clustering (SCARGC) [17]. In this context, SCARGC performs clustering followed by a detection step that is repeated continuously in a closed-loop process, using the current and previous cluster positions derived by clustering unlabeled observations in order to monitor drifting classes over time, thus adapting the ML detection model in nonstationary IoT environment. A second contribution of the paper is that we implement the proposed approach and conduct case studies on real-world IoT security datasets. Through the experiments, we show the feasibility of the proposed approach that can successfully perform intrusion detection over time and provide more sustainable IoT security. In our final contribution, we also compare the detection performance of the IoT intrusion detection approach implemented with/without SCARGC and demonstrate the improvement on three popular ML algorithms for IDS, including logistic regression (LR), support vector machine (SVM), and multilayer perceptron (MLP). This manuscript is organized as follows: Section 2 covers related work in machine learning for IoT and IDS, Section 3 presents the proposed approach, Section 4 presents the experiments and the results, and Section 5 concludes this paper.

2 RELATED WORK

This section briefly introduces ML in IoT and ML for IDS.

2.1 Machine Learning in IoT Systems

Machine learning has become an increasingly important tool in developing IoT systems due to the opportunity of utilizing computing resources to identify patterns and trends from network traffic or behavior trends. Machine Learning in IoT allows for increasing scalability by processing large amounts of data and deploying machine learning models on multiple devices. IoT systems provide the scalability to allow multiple devices to share the computational workload, which provides lower latency [9]. There can be opportunities to improve data privacy maintained locally in the device and develop a security feature that prevents sharing of critical information across the IoT system. In developing IoT systems, the benefits of having a distributed system are tied to ensuring that machine learning algorithms can be distributed across the devices within the IoT system [9]. There is a delineation between local training and distributed training where the local training will host and run a model within a single machine or device. Distributed training allows the model to be performed simultaneously or in parallel [9]. Determining which execution paradigm and what machine learning models are essential before implementing models within the IoT system. Machine learning techniques in IoT systems must identify the model's needs and requirements to learn and classify. In many use cases, machine learning models are used in IoT to provide scalable solutions for computational needs [9]. We see in Figure 2 how the three areas of IoT intrusion detection, streaming data with verification latency, and the use of machine learning techniques intersect. Our research seeks to understand how to utilize machine learning models to prevent cyber-attacks within our IoT system. We hope to find a solution where we can provide an acceptable approach and solution to solving the issues found in these three categories.

2.2 Threat Detection for IoT Using Machine Learning

IoT devices collect diverse information, including electricity consumption, location information, sensor data, sensor networks, and potentially social networks. As IoT provides users the capability of offering capability and ease of use for people's lives, it also threatens personal privacy, and potentially national security [14]. As cyber-attacks continue to pose a threat to IoT systems, there have been studies that sought to capture the opportunities of implementing machine learning techniques for dangerous attacks such as man-in-the-middle (MITM) and Distribute Denial of Service (DDoS) attacks [7]. Due to the nature of machine learning algorithms, the IoT systems implementation needs to adhere to best practices, such as feature selection techniques for cyber-attacks [7].

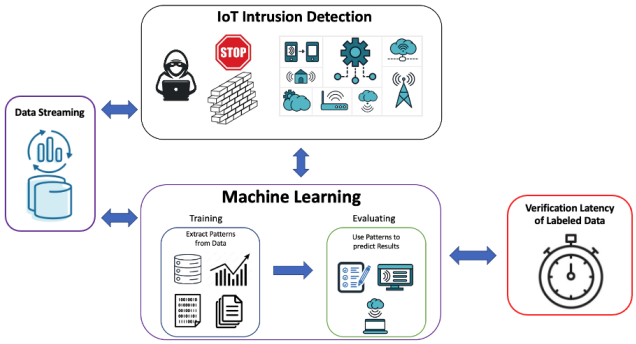

Figure 2: Learning Using EVL Frameworks for IoT

Evaluation of these machine learning techniques and models within the IoT networks, there are design constraints that affect the selection of the machine learning algorithm that can assist in detecting an IoT cyber-attack. Feature selection is based entirely on the dataset for machine learning model training. Understanding the need for machine learning for IoT for intrusion detection, the requirement for a testbed, and training data for our machine learning models is critically important. Due to the growing interconnectivity of many devices, it merits examining the sources of these datasets to train the IoT machine learning models. The cyber threats are diverse enough that various studies address botnet attacks [11, 16]. Other research has sought to develop an IoT dataset that standardizes feature descriptions and cyber-attack classes. The purpose is to utilize a representative heterogeneously IoT dataset to assist with the learning of machine learning intrusion detection systems(IDS) [3, 11, 12]. Based on previous studies, it is imperative to utilize data from a taxonomy of IoT devices where we could implement a machine learning model [13]. Recognized model categories include supervised learning, unsupervised learning, and semi-supervised learning. There are various pros and cons for each model selection, as the efficiency and effectiveness of each model will dictate how well an intrusion attack is recognized within the IoT system. There are commonalities between EVL frameworks and implementations and utilizing machine learning techniques for IoT IDS. EVL frameworks offer opportunities to address cyber-attack datasets' concept drift. Therefore, We seek to determine if EVL frameworks will provide a robust prediction for IoT IDS applications.

3 PROPOSED APPROACH

3.1 Extreme Verification Latency Environment for IoT Intrusion detection

The goal of machine learning is to learn from data. The common assumption for ML-based IDS for IoT is that the underlying data are sampled from a fixed distribution [17]. This assumption is often violated in IoT security applications as varying streaming data from IoT devices will contain varying distributions due to the complicated concept of IoT environments. For example, the normal/abnormal IoT devices' behaviors shift in volume, angle, and location of the streaming without feedback, rendering the intrusion detection model ineffective when only learning from fixed training data distributions. Unfortunately, most IoT applications lack regular access to labeled streaming data. There is a more typical learning setting in which there are limited labeled data for training intrusion detectors and plenty of unlabeled data. If a detector predicts the unlabeled data, a lag period occurs before receiving the true labels. We can refer to this scenario as verification latency [4, 17]. In the extreme scenario, labels of IoT data are only available at the beginning, and afterward, only unlabeled IoT data are present in the stream. This scenario can be referred to as EVL use case within a non-stationary environment for IoT Intrusion detection.

3.2 Design of the Proposed IoT Intrusion Detection System

In contrast to most EVL framework studies which run synthetic datasets, we address EVL by running real-world cybersecurity datasets (i.e., IoT intrusion detection datasets). The EVL algorithms provide a framework for utilizing the varying IoT testbeds to evaluate the classification accuracy, computational complexity, and potential parameter sensitivity. We use the Stream Classification Algorithm Guided by Clustering (SCARGC) as the EVL method integrating with ML classifiers, including Support Vector Machine (SVM), Logistic Regression (LR), and Multi-Layer Perceptron (MLP). These standard classifiers are used as the base classifier for SCARGC to compare the detection performance with real-world IoT datasets. Utilizing EVL algorithms such as SCARGC could allow learning in a non-stationary environment. Intrusion detection systems recognize normal behaviors and attacks based on past data. Our premise of implementing an EVL framework to compare the varying base classifier performance allows us to down-select an implementation that provides robustness for the varying IoT cyber-attack datasets. As shown in Figure 2 our environment entails receiving streaming data that contains both normal and attack data. We address the learning issues under non-stationary environments with EVL seniors.

Our development identifies two IoT datasets that were developed under an IoT system. The goal of using these datasets is to provide a real-world environment from IoT devices under normal behaviors and cyber-attacks. Table 1 demonstrates previous work conducted in IoT utilizing machine learning techniques. The IoT datasets have historically been subjected to traditional machine learning techniques by implementing popular models such as SVMs, Logistic Regression, and neural networks. However, the previous works conducted under EVL and non-stationary environment often utilizes synthetic datasets. The proposed intrusion detection techniques for sustainable IoT security contain two main modules: an ML-based detection model and an EVL method, namely Stream Classification Algorithm Guided by Clustering (SCARGC). Therefore, our approach herein conducted the popular ML classifiers on all the IoT datasets while performing adaptive learning through the EVL method that can deal with non-stationary environments.

3.3 SCARGC for IoT Intrusion Detection

The EVL algorithm SCARGC as a framework addresses the non-stationary environments by clustering the unlabeled streaming data to track drifting classes over time. The SCARGC algorithm begins with building an initial classification model utilizing the available

Table 1: Comparison with the Previous Works

Work Conducted	Standard Machine Learning	Adaptive Machine Learning	EVL	Real-World Data
Bot IoT Dataset [11]	✔	✗	✗	✔
ToN IoT Dataset [3] [12]	✔	✗	✗	✔
Synthetic Datasets [1] [17]	✔	✔	✔	✗
SCARGC [17]	✔	✔	✔	✗
This Work	✔	✔	✔	✔

labeled data while using a clustering approach to associate the clusters to a single class. After SCARGC receives new data, the actions occur: each sample is stored in a pool, and then predict the data stream by the initial classification model. Once the minimum number of examples of each class is stored in the pool, the pool is then clustered again. The goal of clustering from the pool is to map the previous model to the new model found in the clustering phase [17]. The newly labeled examples create an updated classification model that replaces the initial one, allowing the classification phase to adapt to the concept drift. The SCARGC as a framework is built over the past labeled data from the labels provided by the association of clusters in the current iteration [17]. The clustering algorithm used implemented the k-means algorithm as it is a simple and computationally efficient clustering model. Due to this clustering method, the SCARGC framework stores the centroids of the past clustering iteration and uses them as seeds for the current clustering step. This method prevents instability of k-means due to the stochastic nature [17]. The clustering process also allows the SCARGC framework to use the centroids to calculate the Euclidean distance to perform the mapping between the clusters. [17].

4 EXPERIMENTS AND RESULTS

4.1 Experimental Setting

Understanding our IoT datasets allowed us to select proper machine learning models to learn from the data and make predictions. This study utilized two different IoT datasets: an IoT Botnet dataset [11], and an IoT telemetry based on heterogeneous data sources [12].

4.1.1 UNSW Bot IoT Dataset. The Bot IoT dataset consists of three components: network platforms, simulated IoT services, extracting features, and forensics analytics. The network platforms include normal and attacking virtual machines (VMs) with additional network devices such as a firewall and tap [11]. The Bot IoT dataset was developed by simulating five smart devices to operate locally. The testbed dataset used a cloud infrastructure for generating normal/benign network traffic. The IoT platforms were composed of a realistic smart-home network with five IoT devices: 1) Smart Refrigerator, 2) Smart Garage door, 3) Weather Monitoring System, 4) Smart Lights, and 5) Smart thermostat deployed in smart home [11]. The Botnet scenarios simulated normal data and the cyber-attacks including Probing Attacks, Denial of Service, and Information Theft. Probing attacks are malicious attacks that gather information from victims by scanning remote systems. The attacks include port scanning, where an attacker passively probes the network. Active probing includes Operating System (OS) probing in which the attacker scans and gathers information by comparing responses to pre-existing ones and the differences between TCP/IP stack implementations [11]. The Denial of Service (DoS) and Distributed Denial of Service (DDoS) attacks include malicious attacks where bots target a remote machine and disrupt the services by increasing the volume of bots accessing the remote device and increasing the network traffic. The DoS attacks were generated in the Bot IoT dataset through SYN TCP floods. And finally, the information attacks included data theft, and keylogging [11].

4.1.2 UNSW TON IoT Dataset. The generated telemetry dataset, formally named the TON IoT dataset, comprises heterogeneous data sources from the telemetry of IoT services that were orchestrated to demonstrate the interconnections of edge, fog, and cloud layers. The edge layer includes physical devices and operating systems employed as the infrastructure. The fog layer involves the virtualization technology which controls the virtual machines. The cloud layer allows the telemetry data to publish and subscribe capabilities to the network [12]. The normal and attack scenarios were developed to generate normal and cyber-attack scenarios in the testbed to collect the experimental data. In this dataset, there are nine attacks, including 1) scanning attacks, 2) Denial of Service (DoS) attacks, 3) Distributed Denial of Service (DDoS), 4) Ransomware attacks, 5) Backdoor attacks, 6) Injection attacks, 7) Cross-site scripting attacks, 8) Password cracking attack, and 9) Man-In-The-Middle (MITM) attacks. This ToN IoT dataset provides information from six IoT devices and provides heterogeneous and concurrent data sources network traffic [12]. The dataset consists of an IoT system with a smart home that includes a Smart Fridge, Garage, GPS, Light Monitor, Smart Thermostat, Weather Monitoring System, and the data from the Modbus of the IoT system.

4.2 Results

The Bot IoT dataset consists of test and training data for the normal scenarios and various attack scenarios within the IoT system. We divided the training and test data into batches to demonstrate a non-stationary environment. We aim to understand if utilizing EVL machine learning techniques can provide an IoT IDS that is capable of detecting varying cyber-attacks. The batch sizes allowed our experiments to be run in 100 different timesteps.

4.2.1 SCARGC and Base Classifiers. Our approach implements EVL-based IoT IDS by combining the SCARGC and the base classifiers, including LR, MLP, and SVM. We conduct experiments on the UNSW IoT datasets by running the base classifiers with/without integrating SCARGC. Table 2 contains the average accuracy

Table 2: Average Accuracy for IoT Datasets

Dataset	Logistic Regression (LR)	Multi-Layer Perceptron (MLP)	SVM	SCARGC (LR)	SCARGC (MLP)	SCARGC (SVM)
Bot IoT	0.9974	0.9929	0.9880	0.9971	**0.9999**	0.9850
ToN Fridge IoT	0.9981	0.9978	0.9977	0.9988	**0.9989**	**0.9989**
ToN Garage IoT	0.9445	0.9749	0.9956	0.9504	0.9988	**0.9989**
ToN GPS IoT	0.9354	0.7736	0.8568	0.9399	**0.9986**	0.9162
ToN Light IoT	0.9540	0.9720	0.9919	0.9612	**0.9989**	0.9989
ToN Modbus IoT	0.8968	0.5475	0.6924	**0.9985**	0.7070	0.6821
ToN Thermostat IoT	0.9247	0.9249	0.8578	0.9567	**0.9988**	0.9547
ToN Weather IoT	0.9290	0.9291	0.6327	0.9274	**0.9988**	0.9368
Overall Average	0.9475	0.8891	0.8766	**0.9663**	0.9625	0.9340
Rank Average	4.00	4.50	4.75	3.25	**1.50**	3.00

Table 3: Average ROC AUC Scores for IoT Datasets

Dataset	Logistic Regression (LR)	Multi-Layer Perceptron (MLP)	SVM	SCARGC (LR)	SCARGC (MLP)	SCARGC (SVM)
Bot IoT	0.9178	0.7641	0.5108	0.8946	**0.9999**	0.5000
ToN Fridge IoT	0.9985	0.9982	0.9982	0.9989	**0.9990**	**0.9990**
ToN Garage IoT	0.9347	0.9699	0.9954	0.9403	0.9988	**0.9990**
ToN GPS IoT	0.9248	0.7841	0.8507	0.9299	**0.9986**	0.9207
ToN Light IoT	0.9441	0.9668	0.9912	0.9529	**0.9991**	0.9990
ToN Modbus IoT	0.8423	0.5177	0.5056	**0.9983**	0.5377	0.5000
ToN Thermostat IoT	0.8852	0.8880	0.7890	0.9366	**0.9988**	0.9335
ToN Weather IoT	0.9137	0.9138	0.5600	0.9119	**0.9988**	0.9231
Overall Average	0.9201	0.8503	0.7751	**0.9454**	0.9413	0.8468
Rank Average	4.00	4.25	4.88	3.25	**1.38**	3.25

results, while Table 3 contains the average Area Under the Receiver Operating Characteristic Curve (ROC AUC). The utilization of the SCARGC framework as our EVL algorithm for most datasets resulted in increased accuracy across all three implementations. However, there were a few instances where the base classifier outperformed the SCARGC implementation marginally. Notably, SCARGC with an MLP base classifier performed better than the other five implementations. Figure 3 compares the classifiers for the Bot IoT dataset and shows that the SCARGC implementation offers a higher performance value than that of only the base classifiers. Figure 4 provides the increased performance utilizing the SCARGC framework while using MLP. Our results highlight the usage of an EVL method with our base classifiers provides a higher classification accuracy than the simple machine learning models. Both figures show the steps and the classification accuracy at each step. The strategy for developing the data stream was developing various batches with a randomized seed of the UNSW datasets. Our randomization enabled us to create randomized attacks on the streaming system. We evaluated the classifier accuracy for all six implementations. Our results demonstrate how the MLP implementation utilizing the SCARGC framework performs more effectively. We demonstrate how the SCARGC framework provides a higher accuracy rate and ROC/AUC metric scores. In running the Multi-Layer Perceptron classifier, we see a substantial improvement in accuracy and ROC/AUC scores when running the SCARGC framework. The Multi-Layer Perceptron (MLP) base classifier utilizing the SCARGC framework performs better than without the EVL framework. The MLP implementation allows using an artificial neural net within the SCARGC framework. The MLP implementation has outperformed the base classifiers and the SCARGC implementation. We have also performed a rank average on all six implementations and found that the MLP SCARGC implementation has the best overall. We see that the SCARGC implementation exceeds the performance of the base classifiers and is ranked accordingly, as found in Table 2 and Table 3.

5 CONCLUSION AND FUTURE WORK

This paper proposes an intrusion detection approach for sustainable security in IoT, which utilizes an extreme verification latency algorithm (i.e., SCARGC) to handle non-stationary environments where concept drift could be present. The proposed approach is evaluated using the SCARGC framework, which combines multiple base classifiers, including SVM, MLP, and LR. By incorporating the SCARGC framework, we sought to provide additional capability to address concept drift for IoT. We demonstrated that this added capability offers effective and sustainable security for IoT devices when incorporating an intrusion detection system that utilizes the SCARGC framework. The paper utilizes two IoT datasets, namely the Bot IoT and the ToN IoT datasets, which consist of data from various smart devices, such as Fridges, Garages, GPS, Modbus, Lights, Thermostats, and Weather devices. The proposed method

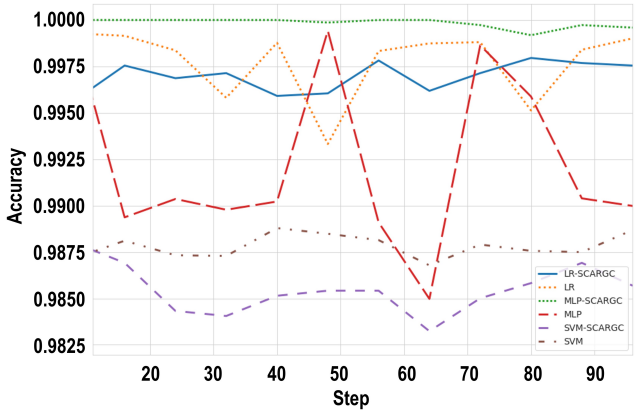

Figure 3: Bot IoT Dataset Comparison Accuracy Between Implementation

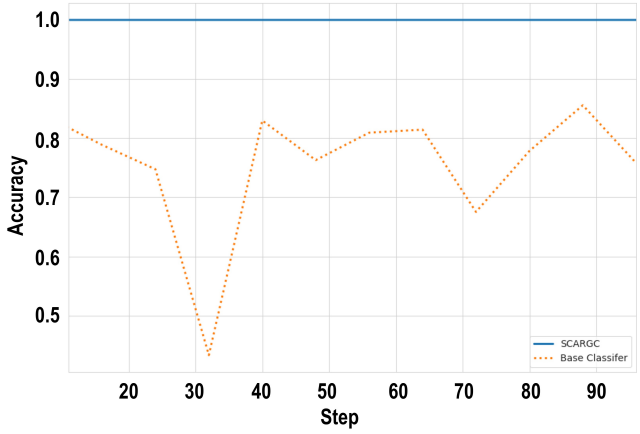

Figure 4: Multi-Layer Perceptron Accuracy Comparison of ToN GPS Dataset

provides higher accuracy or similar classification metrics than the base classifiers. Overall, the proposed method offers a promising approach to detect cyber-attacks accurately in IoT environments, which can help enhance the security of IoT devices and networks. Our future work entails utilizing other EVL methods to evaluate the cyber-attack IoT datasets. Using these real-world IoT datasets provides various opportunities for developing Intrusion Detection Systems that learn incrementally and provide sustainable security for IoT systems. Following our success in running a simple neural network using the MLP, we plan to expand on this implementation, incorporate the EVL framework in neural networks, and potentially incorporate it in deep learning models.

ACKNOWLEDGEMENT

This work is partly supported by National Science Foundation (NSF) projects 1624668 and 1921485, as well as Department of Energy/-National Nuclear Security Administration under Award Number DE-NA0003946 and AGILITY project 4263090 sponsored by Korea Institute for Advancement of Technology (KIAT South Korea).

REFERENCES

[1] Maria Arostegi, Ana I. Torre-Bastida, Jesus L. Lobo, Miren Nekane Bilbao, and Javier Del Ser. 2018. Concept tracking and adaptation for drifting data streams under extreme verification latency. *Intelligent Distributed Computing* (2018), 11–25. https://doi.org/10.1007/978-3-319-99626-4_2

[2] Charan Bandi, Soheil Salehi, Rakibul Hassan, Sai Manoj P D, Houman Homayoun, and Setareh Rafatirad. 2021. Ontology-Driven Framework for Trend Analysis of Vulnerabilities and Impacts in IoT Hardware. In *2021 IEEE 15th International Conference on Semantic Computing (ICSC)*. 211–214. https://doi.org/10.1109/ICSC50631.2021.00045

[3] Tim M. Booij, Irina Chiscop, Erik Meeuwissen, Nour Moustafa, and Frank T.H.Den Hartog. 2022. ToN_IoT: The Role of Heterogeneity and the Need for Standardization of Features and Attack Types in IoT Network Intrusion Data Sets. *IEEE Internet of Things Journal* 9, 1 (1 2022), 485–496. https://doi.org/10.1109/JIOT.2021.3085194

[4] Gregory Ditzler. 2016. A Study of an Incremental Spectral Meta-Learner for Nonstationary Environments. *Proceedings of the International Joint Conference on Neural Networks* 2016-October (9 2016), 38–44. https://doi.org/10.1109/IJCNN.2016.7727178

[5] Gregory Ditzler, Manuel Roveri, Cesare Alippi, and Robi Polikar. 2015. Learning in Nonstationary Environments: A Survey. *IEEE Computational Intelligence Magazine* 10, 4 (2015), 12–25. https://doi.org/10.1109/MCI.2015.2471196

[6] Kevin Immanuel Gubbi, Banafsheh Saber Latibari, Anirudh Srikanth, Tyler Sheaves, Sayed Arash Beheshti-Shirazi, Sai Manoj P D, Satareh Rafatirad, Avesta Sasan, Houman Homayoun, and Soheil Salehi. 2023. Hardware Trojan Detection Using Machine Learning: A Tutorial. *ACM Trans. Embed. Comput. Syst.* (Jan 2023). https://doi.org/10.1145/3579823

[7] Mahmudul Hasan, Md Milon Islam, Md Ishrak Islam Zarif, and Mma Hashem. 2019. Attack and Anomaly Detection in IoT Sensors in IoT Sites using Machine Learning Approaches. (2019). https://doi.org/10.1016/j.iot.2019.10

[8] Rakibul Hassan, Charan Bandi, Meng-Tien Tsai, Shahriar Golchin, Sai Manoj P D, Setareh Rafatirad, and Soheil Salehi. 2023. Automated Supervised Topic Modeling Framework for Hardware Weaknesses. In *2023 IEEE 24th International Symposium on Quality Electronic Design (ISQED'23)*. 125–132. https://doi.org/10.1145/3583781.3590271

[9] Volodymyr Kadzhaia, Aistis Raudys, and Aistis Raudys. 2022. Distributed Machine Learning for IoT. *Vilnius University Open Series* (5 2022), 36–43. https://doi.org/10.15388/lmitt.2022.4

[10] Gaurav Kolhe, Tyler Sheaves, Kevin Immanuel Gubbi, Soheil Salehi, Setareh Rafatirad, Sai Manoj PD, Avesta Sasan, and Houman Homayoun. 2022. LOCK&ROLL: Deep-Learning Power Side-Channel Attack Mitigation Using Emerging Reconfigurable Devices and Logic Locking. In *Proceedings of the 59th ACM/IEEE Design Automation Conference* (San Francisco, California) *(DAC '22)*. ACM, New York, NY, USA, 85–90. https://doi.org/10.1145/3489517.3530414

[11] Nickolaos Koroniotis, Nour Moustafa, Elena Sitnikova, and Benjamin Turnbull. 2018. Towards the Development of Realistic Botnet Dataset in the Internet of Things for Network Forensic Analytics: Bot-IoT Dataset. (11 2018). http://arxiv.org/abs/1811.00701

[12] Nour Moustafa. 2021. A New Distributed Architecture for Evaluating AI-based Security Systems at the Edge: Network ToN_IoT Datasets. *Sustainable Cities and Society* 72 (9 2021). https://doi.org/10.1016/j.scs.2021.102994

[13] Bin Qian, Jie Su, Zhenyu Wen, Devki Nandan Jha, Yinhao Li, Yu Guan, Deepak Puthal, Philip James, Renyu Yang, Albert Y. Zomaya, Omer Rana, Lizhe Wang, Maciej Koutny, and Rajiv Ranjan. 2020. Orchestrating the Development Lifecycle of Machine Learning-based IoT Applications: A Taxonomy and Survey. *Comput. Surveys* 53, 4 (9 2020). https://doi.org/10.1145/3398020

[14] Jing Qiu, Zhihong Tian, Chunlai Du, Qi Zuo, Shen Su, and Binxing Fang. 2020. A Survey on Access Control in the Age of Internet of Things. *IEEE Internet of Things Journal* 7, 6 (2020), 4682–4696. https://doi.org/10.1109/JIOT.2020.2969326

[15] Soheil Salehi, Tyler Sheaves, Kevin Immanuel Gubbi, Sayed Arash Beheshti, Sai Manoj P D, Setareh Rafatirad, Avesta Sasan, Tinoosh Mohsenin, and Houman Homayoun. 2022. Neuromorphic-Enabled Security for IoT. In *2022 20th IEEE Interregional NEWCAS Conference (NEWCAS)*. 153–157. https://doi.org/10.1109/NEWCAS52662.2022.9842256

[16] Muhammad Shafiq, Zhihong Tian, Ali Kashif Bashir, Xiaojiang Du, and Mohsen Guizani. 2021. CorrAUC: A Malicious Bot-IoT Traffic Detection Method in IoT Network Using Machine-Learning Techniques. *IEEE Internet of Things Journal* 8, 5 (3 2021), 3242–3254. https://doi.org/10.1109/JIOT.2020.3002255

[17] Vinicius M A Souza, Diego F Silva, Joäo Gama, and Gustavo E.A.P.A. Batista. 2015. Data Stream Classification Guided by Clustering on Nonstationary Environments and Extreme Verification Latency. *SIAM International Conference on Data Mining 2015, SDM 2015* (2015), 873–881. https://doi.org/10.1137/1.9781611974010.98

[18] Ryan Tsang, Doreen Joseph, Asmita Asmita, Soheil Salehi, Nadir Carreon, Prasant Mohapatra, and Houman Homayoun. 2022. FANDEMIC: Firmware Attack Construction and Deployment on Power Management Integrated Circuit and Impacts on IoT Applications. *2022 Network and Distributed System Security (NDSS) Symposium* (2022). https://doi.org/10.14722/ndss.2022.24349