2022 International Conference on Recent Trends in Microelectronics, Automation, Computing and Communications Systems (ICMACC) | 978-1-6654-9604-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICMACC54824.2022.10093557

Security for IoT using Federated Learning

Harinath Bodagala
C4I Technologies INC
harinathbodgala@gmail.com

Dr. Priyanka H
Asst. Professor, PES University
Bengaluru, Karantaka, India
priyankahsachin@gmail.com

***Abstract*— In recent years, the government, universities, and industry have paid attention to the distributed approach in machine learning (ML) and cybersecurity for the developing Internet of Things (IoT). Federated cybersecurity (FC) is a strategy for making the Internet of Things (IoT); more secured and effective in the future. This new approach has an ability to detect problems regarding security, implement solutions, and across the IOT systems it manages effectively to limit the propagation of vulnerabilities. Forming a federation of shared information is one way to attain a cybersecurity. Federated learning (FL) is a machine learning paradigm which is especially effective for securing the sensitive IoT environment. The origin of FL, and also FL for cyber security are presented in this article. The various security assaults, and responses are also discussed as an outcome of this article. Experiments are carried out in Google Colaboratory, using well-known python libraries. The results shows that FL provides highest level of security in comparison with centralised learning.**

Keywords*— *IoT, Federated Learning, Cyber Secuirty and Machine Learing

I. INTRODUCTION

Today's society is increasingly reliant on facts, and protecting the data is a critical job. Individual, organizational, and government sensitive data must be transferred from one place to another point over a link of network. Netfilx has been using machine learning for their applications and services, from small devices to huge corporations like Google, Facebook, and so on. Machine learning not only becoming essential for improving user experience but also helpful in building models for business. It has also become essential for identifying and avoiding cyber risks and attacks. But in today's cyberspace, attack patterns and tactics have undergone a significant transformation. Polymorphic viral attacks are challenging to predict and detect since their signature is constantly changing. Thus, the use of ML is to identify and predict risks, abnormalities in the cyberspace and then take appropriate actions against the identified problems. Cybercrime has been identified in many fields including Cloud computing [1] [2], Iot and so on. All the data must be secured and privacy should be maintained for all users in the central server or in the cloud.

For ML-based cybersecurity, there are several models now being used, each with pros and cons. These models include centralised, decentralised, and federated [3]. Among these methods, the federated learning (FL) model is helpful in cybersecurity. Additionally, FL's potential applications in a number of fields have been investigated, including vehicle ad hoc networks, edge networks [4], smart cities , recommender systems, health care and many others. Since data produced in an endpoint does not exit the device, the FL architecture naturally promotes security and privacy (in contrast to the centralised learning framework). Figure 1 represents difference between centralised learning and federated learning.

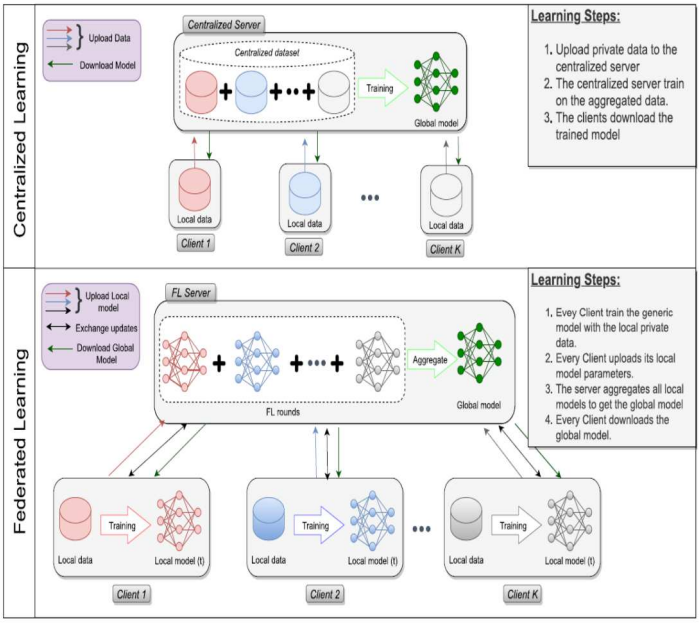

Fig. 1. Steps of federated learning and centralised learning

Centralized Learning: Traditionally, ML has been used by IOT applications for transferring all the information from each linked Iot system to the cloud servers in order to create a common model that can be used and deployed among multiple devices. The key advantage of this tyoe of learning is that the model can instantly function with additional relevant IoT devices after generalising utilising information from a set of IoT devices. However, typical centralised learning has some challenges, including security, delay, and networking.

Federated Learning: The fundamental idea behind FL is to build machine learning algorithm using dispersed data across various devices without letting any data leak. FL is a novel method that, specifically, downloads the most recent model and computes an extended model on IoT devices using the localized

978-1-6654-9604-9/22/$31.00 ©2022 IEEE

IoT data. The IoT devices then send these locally trained models back to the central computer for compilation.

There isn't a lot of exploratory research done on IoT applications using federated learning, and the majority of studies focus more on data protection. A safe federated learning architecture that combines “federated transfer learning, vertical federated learning, and horizontal federated learning” [5] was reviewed by Yang et al. Review of “federated learning algorithms” [5] by Aledhari et al. with use-cases, practical applications, and hardware platforms. “An approach to the integration of federated learning in the context of 6G communications” [6] was offered by Liu et al. [6] Difficulties and advantages of federated learning's use in smart city sensing was presented by Jiang et al [7]. A thorough examination of the privacy risks associated with federated learning was presented by Mothukuri et al. [8], but there was no experimental study on IoT networks.

A thorough overview of the most current developments in federated learning and IoT applications was presented by Nguyen et al. [9]. A categorization of multi-level federated ML in networking and communication systems was proposed by Wahab et al. [10]. The outline of “federated learning and blockchain for IoT application concepts” [11] was presented by Ali et al [11].

A division of the two main types of attacks are—poisoning and inference—for threat models. The analysis highlights the key takeaways, key strategies, and underlying presumptions that underlie the various attacks. Model poisoning, as opposed to conventional data poisoning, is a new threat that the FL context introduces and it also helped in classification of inputs. The threat detection was carried out by Bhagoji et al.[12] included targeted model, by enhancing the malicious entity and improving threat detection by using an alternate reduction method.

III FEDARATED LEARNING FOR CYBER SECURITY

FL supports collaborative learning approach that only distributes model updates that can be a good solution to increase safety and confidentiality because the IoT environment is more vulnerable to various types of cyber-attacks. Developing and upgrading cyber defence models and methods in response is made easier by fast learning and global sharing of information on various types of cyberattacks, such as spoofing, breach, anomalies, and Denial of Service attacks. Therefore, FL has enormous potential for effective network and device-level cyberspace security. In Fig 2, the use of FL as a defence against potential dangers is shown.

The fundamental idea behind federated learning is to develop machine learning models that are based on distributed datasets across several devices without letting data to leak. The new technique known as "federated learning" specifically involves downloading the most recent model and computing an updated model on IoT devices by utilizing IoT data which is available locally.

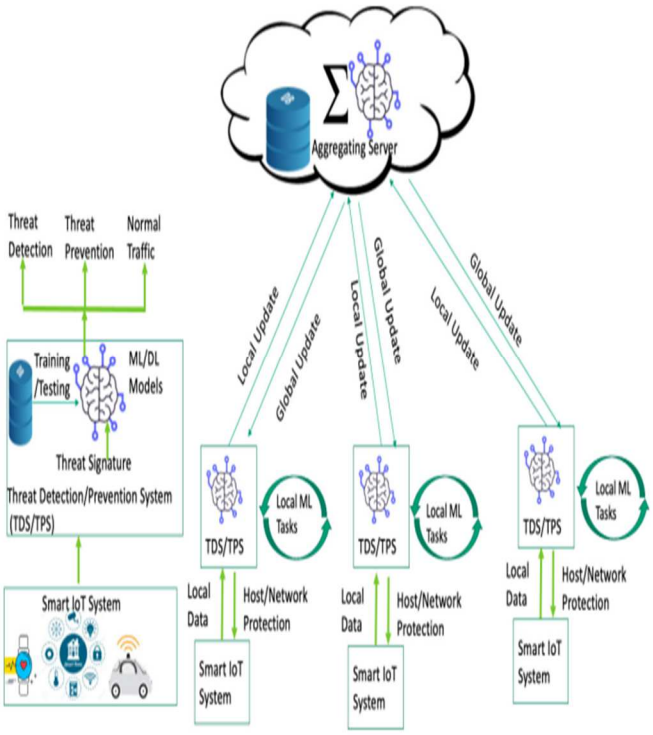

Fig. 2. IoT attacks

Regarding federated learning deployment and the related practical and technological problems, data distribution has become crucial. The three different types of federated learning that exist currently are depicted in Fig 3.

Horizontally FL is used when data sets' sampling spaces differ, and they share the same feature space.
Vertical FL is used when there are differences between the feature spaces of the two data sets but they share the same sample space.
Federated learning is used when there are differences in the feature and sampling spaces between the data sets.

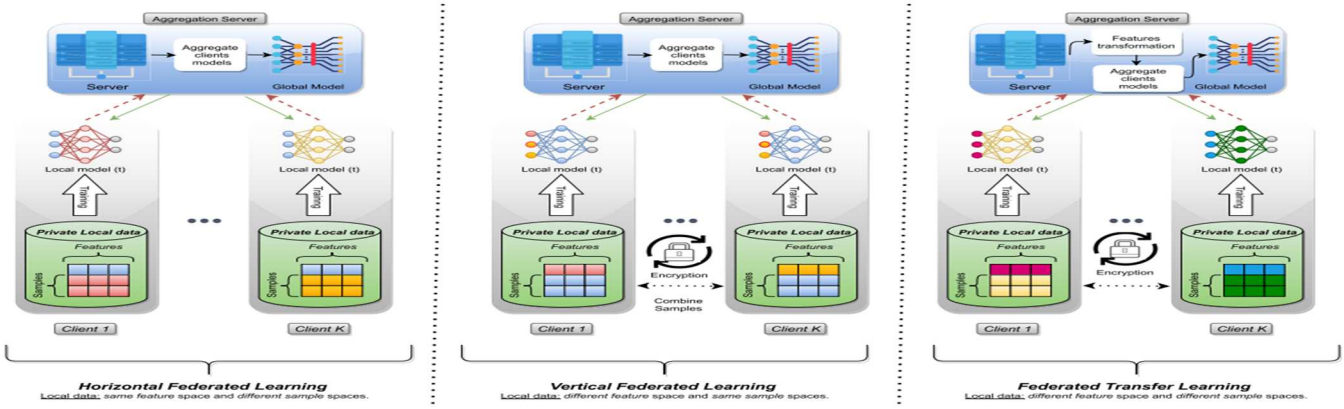

Fig. 3. Different types of FL

A. FL for Iot devices

IoT devices are being used more often in daily life. Most of these devices are susceptible to the attack due to risky installation, configuration, and design practices. The research by Ciuonzo et al. [13] , emphasized on the problem of decentralised identification of a faulty entity in a "wireless sensor network", demonstrates the number of current systems already contain IoT devices that are weak points in existing systems and are therefore vulnerable to hacking, which is dangerous for delicate tasks like monitoring.

The DIOT is autonomous self-learning decentralised method, introduced by Nguyen et al. [14] to find infected IoT devices, is built on a federated learning methodology. The federated learning technique is implemented using the “flask and flask socketio packages” [14]. The analysis of the DIOT scheme's performance reveals that it can identify 95.6 percent of threats in an average of 257 milliseconds. A federated learning-based intrusion detection system was created by Zhao et al. [15] and can be used to find hacked IoT devices. The suggested approach advises dispersing the user servers over the worldwide initial “long-short-term memory model” [15]. The user servers then create their own distinct models and begin transferring the settings for those models to the main server.

The central server then delivers the new aggregate global model to the user servers after combining the model settings to create it. The simulation results on the “SEA dataset”, which was created by the “AT&T Shannon Lab”, show that the suggested system achieves greater accuracy and coherence than the traditional systems. Mohammed et al. [16] initiated a stateful online heuristic that relies on federated learning in conjunction with an IoT application called client alarm which alerts users of any unauthorised access to devices within the Iot environment. Federated learning has been used to identify the appropriate clients and address the issue of accuracy optimization

III EXPERIMENTAL RESULTS

We train Deep Neural Network (DNN), Convolutional Neural Network (CNN), and Recurrent Neural Network (RNN) models for FL cyber threat detection in the IoT. Then, we contrast the outcomes with those from the traditional/centralized machine learning approaches. On Google Colaboratory, we carried out our experiments using recognised libraries like Keras, Pandas, TensorFlow, and NumPy. Various federated learning platforms that are available for modelling and testing federated learning algorithms.

The training procedure used in our federated learning model is shown in Fig 4. The algorithm. 1 shows the steps involved in training the dataset. The aggregation server initially selects a C percentage of M clients to participate in the FL plan and do computations for the R FL rounds. A common model with a set of initial weights (fw) is created randomly by the aggregation server. The generic model is then retrieved from the aggregate server by each of the clients K. Each client 'k' generates a new localized weights C1 for the newly created local model and retrains the model of generic locally using its own private data.

Algorithm 1: Federated learning

Server (K, C, R) :
  $w_1 \leftarrow GenericModel()$
  for $t = 1, .., R$ **do**
    $S_t \leftarrow$ Subset(max$(C \cdot K, 1)$, “$random$”)
    Parallel.for $k \in S_t$ **do**
      $w_{t+1}^k \leftarrow Client(w_t, k)$
    end
    $w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$
  end
Client (w, k) :
  $\mathcal{B} \leftarrow$ Split$(\mathcal{P}, B)$
  for $i = 1, .., E$ **do**
    for $b \in \mathcal{B}$ **do**
      $w \leftarrow w - \eta \nabla f(w, b)$
    end
  end
  Send w to Server

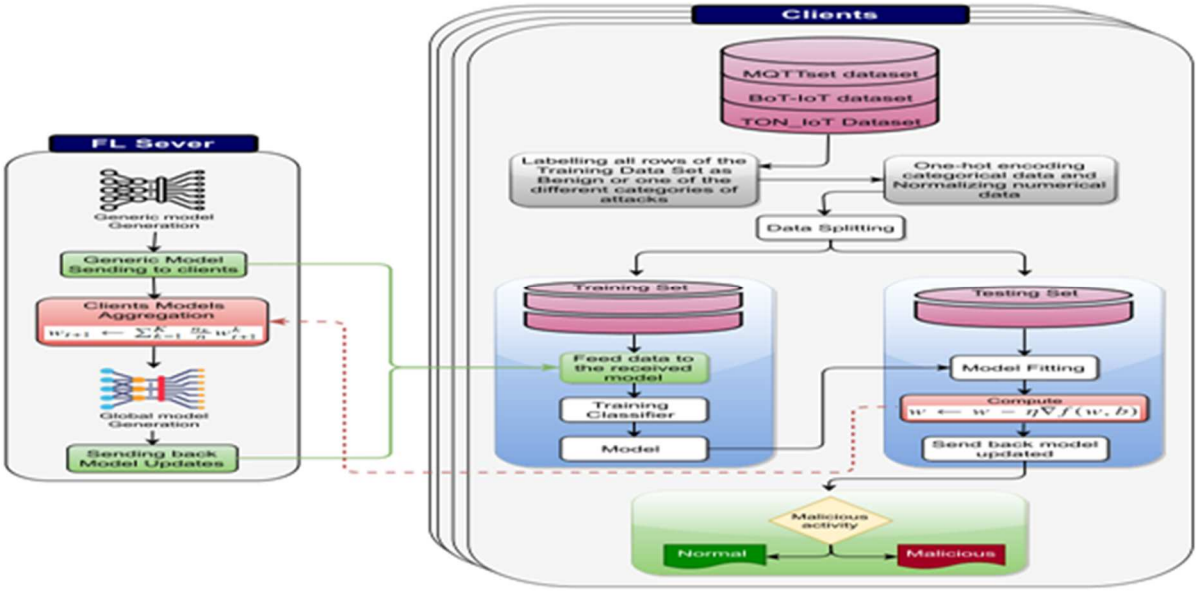

Fig. 4. Training procedure for FL model

The security research group employs various datasets to evaluate FL-based cybersecurity solutions in IoT networks: "TON IoT" [17], "Bot-IoT" [18] and "MQTTset" [19]. They were carefully chosen since they were created using a variety of data sources and were compiled from sensor telemetry datasets from the IoT and IIot.

1) The "BoT-IoT dataset [18]" was created at "UNSW Canberra's Cyber Range Lab" as an outcome of creating a real-world network environment that combined both botnet and regular traffic. The entire 70 GB set of PCAP files, containing more than 72 million data, were recorded. A CSV files, are provided for the dataset and it is used to carry out experiments.

2) To simulate a smart IoT environment, MQTT dataset consists of communication between multiple IoT devices that is based on the "Message Queue Telemetry Transport (MQTT) " protocol. It includes actual attacks made specifically to attack the IoT environment.

3) To gather and analyse data from a variety of IoT and Industrial IoT sources., the IoT Lab of the "UNSW Canberra Cyber, School of Engineering and Information Technology (SEIT), UNSW Canberra"[20], has introduced .The performance is increased, after 47 FL runs. The centralized models did not perform well for the non-distribution data. Randomly, the given samples of data were assigned to each. "TON_IoT dataset" [17] (IIoT). The cross-layer communication is addressed across the three technologies: i) IIoT ii) Cloud, and iii) Edge/Fog systems For an IoT sensor data service

In this experimental configuration, a federated learning strategy is used, where the shared knowledge between the aggregation server and the subscribing users to begin the learning process rather than collecting all the information in one location and the subscribing users to begin the learning process rather than collecting all the information in one location and starting from there. The data never exit from the client side. The validation accuracy for each global model versus the centralized model is shown in fig. 5 by combining all datasets. Fig 5 displays the accuracy rate for each FL models in comparison to the centralized model. Fig 5 illustrates the accuracy rate for the Bot-Iot dataset utilizing FL classifier such as "Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Deep Neural Network (DNN)". For distributed datasets, the efficacy of FL models was comparable to that of centralized learning.

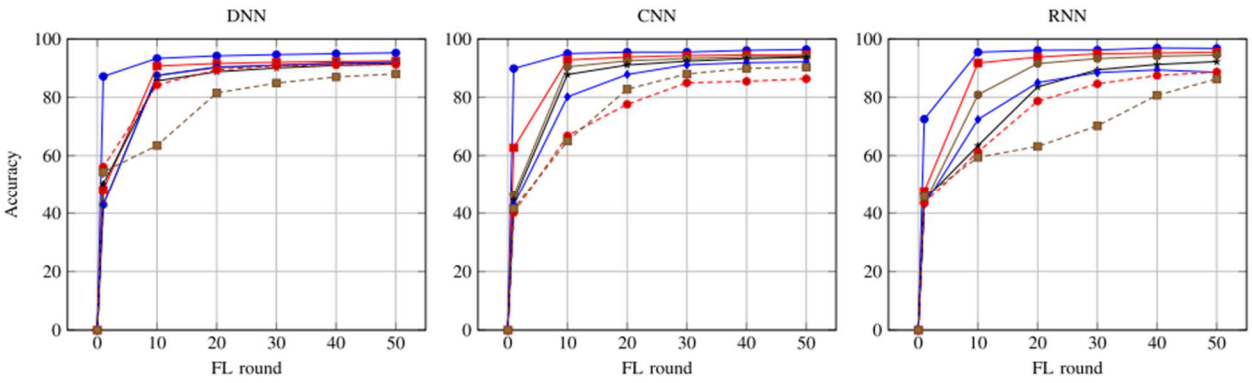

Fig .5. Bot-IoT dataset

Table. 1. Evaluation results of FL

		Precision			Recall			F_1-Score		
Dataset	**Class**	**DNN**	**CNN**	**RNN**	**DNN**	**CNN**	**RNN**	**DNN**	**CNN**	**RNN**
Bot-IoT	Benign	100%	100%	100%	100%	100%	100%	100%	100%	100%
	DDoS	97%	95%	97%	94%	97%	95%	96%	96%	96%
	DoS	94%	97%	95%	96%	93%	97%	95%	95%	96%
	Reconnaissance	96%	97%	99%	99%	99%	99%	97%	98%	99%
	Theft	00%	00%	100%	00%	00%	36%	00%	00%	53%

The main objective of this research is to compare the efficiency of FL with conventional machine learning. It mainly concerned with providing cybersecurity for Internet of Things systems. Based on the dataset from the "Bot-IoT," was done as shown in table 1. The study demonstrates that federated learning techniques such as “DNN, CNN, and RNN” perform better than conventional/centralized ML methods. The use of FL will improve the accuracy in spotting threats and safeguard the confidentiality of data from IoT devices.

VI CONCLUSION

In this work, we compared and experimentally analysed federated learning techniques for IoT application cybersecurity. For a variety of IoT applications, we specifically looked into the FL-based data security in the IoT devices. We looked at the security and privacy flaws in FL based systems. By utilizing the 3 deep learning methods RNN, CNN, and DNN, we carried out an experimental examination of FL. We looked at each deep learning model's performance using the Bot-IoT dataset by applying into FL and centralised learning. According to the results, FL algorithms can ensure the confidentiality of IoT device and offer the greatest level of security better than traditional/centralized machine learning.

REFERENCES

[1] Priyanka, H., Cherian, M. (2020). Efficient Utilization of Resources of Virtual Machines Through Monitoring the Cloud Data Center. In: Bindhu, V., Chen, J., Tavares, J. (eds) International Conference on Communication, Computing and Electronics Systems. Lecture Notes in Electrical Engineering, vol 637. Springer, Singapore. https://doi.org/10.1007/978-981- 15-2612-1_62

[2] Priyanka, H. and Cherian, M., 2021. Effective Utilization of Resources through Optimal Allocation and Opportunistic Migration of Virtual Machines in Cloud Environment. International Journal of Cloud Applications and Computing (IJCAC), vol 11, Issue (3), pp.72-91.

[3] M. A. Ferrag and L. Maglaras, ‘‘DeepCoin: A novel deep learning and blockchain-based energy exchange framework

[4] M. A. Ferrag, L. Maglaras, S. Moschoyiannis, and H. Janicke, ‘‘Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study,’’ J. Inf. Secur. Appl., vol. 50, Feb. 2020, Art. no. 102419.

[5] Q. Yang, Y. Liu, T. Chen, and Y. Tong, ‘‘Federated machine learning: Concept and applications,’’ ACM Trans. Intell. Syst. Technol., vol. 10, no. 2, pp. 1–19, 2019.

[6] Y. Liu, X. Yuan, Z. Xiong, J. Kang, X. Wang, and D. Niyato, ‘‘Federated learning for 6G communications: Challenges, methods, and future directions,’’ China Commun., vol. 17, no. 9, pp. 105–118, Sep. 2020.

[7] J. C. Jiang, B. Kantarci, S. Oktug, and T. Soyata, “Federated learning in smart city sensing: Challenges and opportunities,” Sensors, vol. 20, no. 21, p. 6230, Oct. 2020.

[8] V. Mothukuri, R. M. Parizi, S. Pouriyeh, Y. Huang, A. Dehghantanha, and G. Srivastava, ‘‘A survey on security and privacy of federated learning,’’ Future Gener. Comput. Syst., vol. 115, pp. 619–640, Feb. 2021.

[13] D. Ciuonzo, P. S. Rossi, and P. K. Varshney, ‘‘Distributed detection in wireless sensor networks under multiplicative fading via generalized score tests,’’ IEEE Internet Things J., vol. 8, no. 11, pp. 9059–9071, Jun. 2021.

[14] T. D. Nguyen, S. Marchal, M. Miettinen, H. Fereidooni, N. Asokan, and A.-R. Sadeghi, ‘‘DÏoT: A federated self-learning anomaly detection system for IoT,’’ in Proc. IEEE 39th Int. Conf. Distrib. Comput. Syst. (ICDCS), Jul. 2019, pp. 756–767

[15] R. Zhao, Y. Yin, Y. Shi, and Z. Xue, ‘‘Intelligent intrusion detection based on federated learning aided long short-term memory,’’ Phys. Commun., vol. 42, Oct. 2020, Art. no. 101157.

[16] I. Mohammed, S. Tabatabai, A. Al-Fuqaha, F. E. Bouanani, J. Qadir, B. Qolomany, and M. Guizani, ‘‘Budgeted online selection of candidate IoT clients to participate in federated learning,’’ IEEE Internet Things J., vol. 8, no. 7, pp. 5938–5952, Apr. 2020

[17] N. Moustafa, ‘‘New generations of Internet of Things datasets for cybersecurity applications based machine learning: TON_IoT datasets,’’ in Proc. eResearch Australasia Conf., Brisbane, QLD, Australia, 2019, pp. 21–25.

[18] N. Koroniotis, N. Moustafa, E. Sitnikova, and B. Turnbull, “Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics: Bot-IoT dataset,’’ Future Gener. Comput. Syst., vol. 100, pp. 779–796, Nov. 2019.

[19] I. Vaccari, G. Chiola, M. Aiello, M. Mongelli, and E. Cambiaso, “MQTTset, a new dataset for machine learning techniques on MQTT,’’ Sensors, vol. 20, no. 22, p. 6578, Nov. 2020.

[20] N. Moustafa, “A new distributed architecture for evaluating AI-based security systems at the edge: Network TON_IoT datasets,’’ Sustain. Cities Soc., vol. 72, Sep. 2021, Art. no. 102994