An Ensemble of Deep Recurrent Neural Networks for Detecting IoT Cyber Attacks Using Network Traffic

Mahdis Saharkhizan, Amin Azmoodeh, Ali Dehghantanha, *Senior Member, IEEE*, Kim-Kwang Raymond Choo, *Senior Member, IEEE*, and Reza M. Parizi, *Senior Member, IEEE*

***Abstract*—Internet-of-Things (IoT) devices and systems will be increasingly targeted by cybercriminals (including nation state-sponsored or affiliated threat actors) as they become an integral part of our connected society and ecosystem. However, the challenges in securing these devices and systems are compounded by the scale and diversity of deployment, the fast-paced cyber threat landscape, and many other factors. Thus, in this article, we design an approach using advanced deep learning to detect cyber attacks against IoT systems. Specifically, our approach integrates a set of long short-term memory (LSTM) modules into an ensemble of detectors. These modules are then merged using a decision tree to arrive at an aggregated output at the final stage. We evaluate the effectiveness of our approach using a real-world data set of Modbus network traffic and obtain an accuracy rate of over 99% in the detection of cyber attacks against IoT devices.**

***Index Terms*—Deep learning (DL), Internet of Things (IoT), IoT security, network traffic, recurrent neural networks (RNNs).**

I. Introduction

THE Internet of Things (IoT) can be broadly defined as a pervasive network of a (broad) range of connected smart nodes that offer diverse digital services, including the collection of environmental and user data. For example, IoT nodes can sense, process, and communicate (complex) information through IoT infrastructures to improve the quality and quantity of services and user experience in sectors ranging from healthcare to transportation to power management to military, etc. On the flip side, IoT devices and systems can also be an attack vector where an attacker (or an adversary) can seek to obtain information, target other entities (e.g., governments), and/or facilitate nefarious activities.

Manuscript received January 20, 2020; revised April 22, 2020; accepted May 19, 2020. Date of publication May 21, 2020; date of current version September 15, 2020. The work of Kim-Kwang Raymond Choo was supported in part by the National Science Foundation under Award 1925723. *(Corresponding author: Kim-Kwang Raymond Choo.)*

Mahdis Saharkhizan is with the School of Electrical and Computer Engineering, Shiraz University, Shiraz 71348-51154, Iran (e-mail: mahdis@cybersciencelab.org).

Amin Azmoodeh and Ali Dehghantanha are with the Cyber Science Lab, University of Guelph, Guelph, ON N1G 4S7, Canada (e-mail: aazmoode@uoguelph.ca; adehghan@uoguelph.ca).

Kim-Kwang Raymond Choo is with the Department of Information Systems and Cyber Security, University of Texas at San Antonio, San Antonio, TX 78249 USA (e-mail: raymond.choo@fulbrightmail.org).

Reza M. Parizi is with the College of Computing and Software Engineering, Kennesaw State University, Marietta, GA 30060 USA (e-mail: rparizi1@kennesaw.edu).

Digital Object Identifier 10.1109/JIOT.2020.2996425

In existing networks (including those comprising IoT devices), security systems such as intrusion detection systems (IDSs) are typically used to monitor the network traffic and identify suspicious activities within the traffic [1], [2]. IDSs can be either signature based or anomaly based, where signature-based IDSs recognize intrusions (or suspicious activities) by finding the relationship between the previously learned rules/signatures of known attacks' rules. Anomaly-based IDSs, on the other hand, monitor network traffic and compare the traffic with previously learned patterns to spot malicious activities [3]. It is known that signature-based methods are not very effective in detecting new and unknown attacks. Anomaly-based methods have shown to be able to recognize known and new attacks [4] to some degree, but they often have high false-positive rates.

In recent years, there has been renewed interest in exploring the utility of artificial intelligence (AI) techniques, such as machine learning (ML) and deep learning (DL), in designing cybersecurity solutions, such as malware detection [5]–[11] and threat intelligence [12], forensic investigation [13], and privacy-preserving techniques [14]. DL-based approaches typically include a learning model with several layers, and each layer contains a significant number of computational nodes. However, designing efficient and effective AI-based IoT attack detection systems remains an open research challenge.

In this article, we propose a new approach that monitors the network traffic of IoT networks over the Modbus protocol [15] and extracts network packets to train an ensemble of long short-term memory (LSTM) models. From there, it aggregates the output of LSTMs by a decision tree (DT) and assigns the right label to each network connection. The proposed approach is characterized by the following capabilities.

1) A significantly high level of accuracy in detecting different attacks within IoT networks.
2) The capability to detect attacks for different periods, including right from the start of an attack.
3) The marginal false-positive (FP) rate with respect to the ensemble of detection modules.

The remainder of this article is organized as follows. Section II briefly reviews the extant literature. In Section III, we outline our research methodology. In Sections IV and V, we present our proposed approach, describe the evaluation

setup, and discuss the findings. Finally, Section VI concludes this article.

II. Related Literature

There has been an increasing focus on ensuring the security of IoT networks, partly due to the popularity of IoT devices in our society [16]. For example, Oh *et al.* [17] introduced a signature-based intrusion detection method for IoT systems, using multiple pattern-matching algorithms. Using a data set of Snort and ClamAV extracted rules, their approach has a detection rate between 81% and 90% in different experimental settings. In another study, Anthi *et al.* [18] proposed an anomaly-based IDS for smart home IoT devices using a three-layer IDS that leverages a supervised learning mechanism. They achieved an F-measure between 90% and 98%. However, signature-based mechanisms can be bypassed by modifying the attack's rules without degrading its harmfulness.

To facilitate malicious activity detection, Azmoodeh *et al.* [19] proposed a DL-based approach that extracts a graph of executable files' operation codes. They then introduced a feature selection method and used it to generate an adjacency matrix of extracted graphs prior to training a convolutional neural network (CNN) in order to identify malicious and benign applications. This approach needs binary executable as input. In order to enhance the security of wireless IoT networks, Aminanto *et al.* [20] proposed a deep-feature extraction and selection model using deep autoencoders. Then, they combined the autoencoder with a supervised classification algorithm and reportedly achieved a detection rate of 99.918% and a false alarm rate of 0.012% for detection impersonation attacks. This method is designed to have maximum performance for impersonation attacks while our approach considered a broader range of IoT's cyber attacks.

Activities within IoT systems rely on sequential data, such as the sequence of network packets, operational codes, or environmental sensed variables. Hence, HaddadPajouh *et al.* [21] presented a deep recurrent neural network (RNN)-based approach that uses the sequence of IoT executable's operational codes for training (together with a data set of IoT malware and benign samples). Findings from their evaluation reported an accuracy rate of 98.8% in recognizing malicious payloads. For device-level detection in IoT systems, Azmoodeh *et al.* [22] collected power consumption signals of infected IoT nodes. Then, they applied a grinding algorithm to these signals and trained different classifiers to identify infected nodes. They reportedly achieved an accuracy rate of 94.27% in detecting IoT nodes infected by crypto-ransomware. However, HaddadPajouh *et al.* [21] and Azmoodeh *et al.* [22] required device-level information to detect malicious activities. For enhanced scalability and robustness, Diro and Chilamkurti [23] presented an LSTM-based model for distributed cyber attack detection in fog-to-things communication. They reportedly achieved an accuracy rate of 99.91% and 98.22% on the ISCX and AWID data sets, respectively. This method has not been proposed to apply to Modbus network traffic. In addition, the proposed method considers a window of network session to enhance its detection rate as well to reduce false alarms.

Supervisory control and data acquisition (SCADA) and Modbus protocol are two fundamental building blocks in IoT-based systems, particularly those deployed in critical infrastructures and industrial systems [24], [25]. Anton *et al.* [26] evaluated the performance of ML-based anomaly detection systems on the industrial Modbus data set of cyber attacks. They investigated the utility of support-vector machine (SVM), random forest (RF), k-nearest neighbor (kNN), and k-means clustering on a synthetic data set, and the findings suggested that SVM has the highest accuracy rate of 100% in the majority of their experimental settings. Despite the significant performance, this approach was designed and assessed on a synthetic data set which degrades its reliability to work on real environments. In another study, Goldenberg and Wool [15] modeled the behavior of the Modbus protocol to detect intrusions using deterministic finite automaton (DFA). Their algorithm modeled the traffic of the Modbus protocol and was sensitive to anomalies. The authors reportedly obtained accuracy rates between 65% and 99% for different numbers of DFAs. Ullah and Mahmoud [27] proposed a hybrid model for detecting anomalous SCADA data using an ML-based model, which eliminates irrelevant features to increase the accuracy of detection. According to the authors, their approach achieved a precision rate of 100% in the majority of their experiments on the KDD99 data set. Anton *et al.* [26], Goldenberg and Wool [15], and Ullah and Mahmoud [27] made an effort to identify anomalies that are more likely to have a higher false alarm while the proposed approach is proposed to learn attack behaviors and detect them.

III. Research Methodology

In this section, we first describe our data set, its content, and our approach to preparing it for the learning task (Section III-A). Then, we explain our approach to extract data for the learning task (Section III-B) and introduce our evaluation approach to evaluate the competency of the learning task (Section III-C).

A. Data Set

To have a clear view of the used data set in this article, Section III-A1 provides a description of the Modbus protocol, and Section III-A2 gives the information about the characteristics of the data set.

1) Modbus Over TCP/IP: The Modbus[1] protocol is widely deployed in industrial control systems (ICSs), and it works in a master/slave mode. Although it was initially developed for serial communication, it is now often used over the transmission control protocol (TCP). There have been different versions of Modbus proposed over the years, namely, *Modbus RTU*, *Modbus ASCII*, and *Modbus over TCP/IP*.

The Modbus/TCP is a recognized and approved protocol by the Internet assigned number authority (IANA) since 1996, with its default port number of 502. Instead of using the device address, Modbus uses an IP address to communicate and interact between the master and slave nodes. As shown

[1] http://www.modbus.org

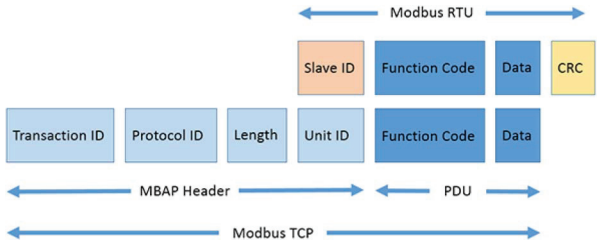

Fig. 1. Modbus PDU over TCP/IP.

TABLE I
DATA SET INFORMATION (NUMBER OF PCAP FILES FOR EACH CLASS)

Class	Number of pcap files
1- Clean	3
2- MITM	22
3- ModbusQueryFlooding	52
4- PingFloodDDoS	37
5- TcpSYNFloodDDoS	37

in Fig. 1, protocol data unit (PDU) frames that include function code to run on the device, is the fundamental part of the Modbus/TCP packet [28]. The majority of Modbus messages include commands, such as read and write to control industrial nodes.

2) Data Set Description: The data set [29] used in this article contains Modbus/TCP network traffic data, which have been simulated based on a small-sized process industrial automation scenario. The data set includes five categories of network traffic, namely, *Clean* traffic, man-in-the-middle (MITM) attack, *Ping DDoS Flood* attack, *Modbus Query Flood* attack, and *TCP SYN DDoS Flood* attack [30]. Network traffics of data set were captured into *pcap*[2] files. Table I gives information about the number of pcap files corresponding to each class.

B. Extracting Modbus Flows

In order to extract the captured Modbus network traffics, CICFlowmeter[3] toolset [31] was first utilized. As a result, 83 features were extracted for each network packet. Table II provides information about the number of samples (network packet) belong to each class. Then, to eliminate features that were highly correlated to environment setup and were suspected of causing bias in resulting ML model, *FlowID, SourceIP, DestinationIP, SourcePort, DestinationPort,* and *Timestamp* were removed. Finally, the column normalization on the prepared data set was applied.

C. Evaluation Metrics

The following criteria are used to evaluate the utility of ML-aided techniques in intrusion detection.

1) *True Positive (TP):* It indicates that an intrusion is correctly identified.
2) *True Negative (TN):* It indicates that a benign activity is detected as a nonmalicious activity correctly.

[2]https://en.wikipedia.org/wiki/Pcap
[3]http://www.netflowmeter.ca/netflowmeter.html

TABLE II
EXTRACTED SAMPLES FROM PCAP FILES

Class	Number of Samples
1- Clean	259,635
2- MITM	230,330
3- ModbusQueryFlooding	4,021,403
4- PingFloodDDoS	616,746
5- TcpSYNFloodDDoS	730,971
Total	5,859,085

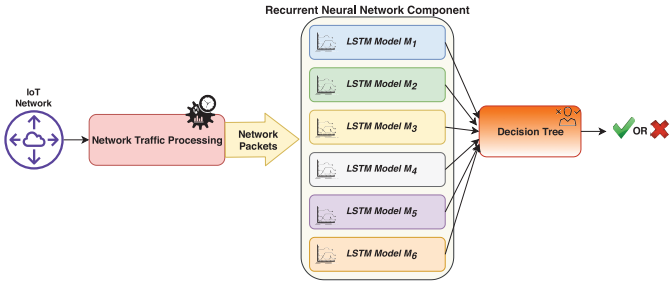

Fig. 2. Proposed method overview.

3) *FP:* It indicates that a benign activity is falsely detected as a malicious activity.
4) *False Negative (FN):* It indicates that an intrusion is not detected and labeled as a nonmalicious activity.

Based on the criteria described above, the following metrics are introduced to quantify the effectiveness of a given system.

1) *Accuracy:* It indicates the number of samples that a classifier correctly detects, divided by the number of all samples

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{1}$$

2) *Precision:* It is another metric that indicates the ratio of predicted intrusion samples that are correctly predicted

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{2}$$

3) *Recall:* It indicates the ratio of intrusion samples that are correctly predicted

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{3}$$

4) *F-Measure:* It is the harmonic mean of precision and recall, and is defined as follows:

$$F\text{-Measure} = \frac{2 * \text{TP}}{2 * \text{TP} + \text{FP} + \text{FN}}. \tag{4}$$

IV. PROPOSED METHOD

The proposed method includes a stack of deep RNNs (Section IV-A) that are trained with the prepared data set and a DT that aggregates the output of RNNs (Section IV-B). Fig. 2 illustrates the conceptual view of our proposed method.

A. LSTM Models

Deep RNNs are a fundamental category of DL models that are proposed to apply learning tasks on sequential data. Despite the considerable capability of RNNs to learn from

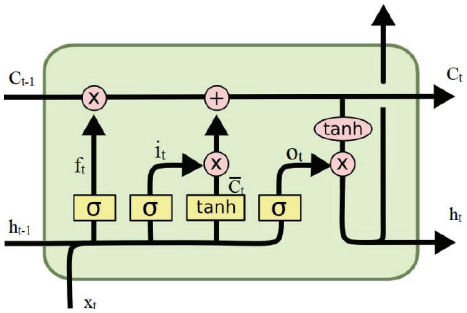

Fig. 3. LSTM cell.

TABLE III
PROPOSED METHOD'S LSTMS SETTINGS

Mode Name	Number of Layers	Layer Size
LSTM-1	1	(100)
LSTM-2	1	(200)
LSTM-3	2	(100,100)
LSTM-4	2	(200,100)
LSTM-5	3	(100,100,100)
LSTM-6	3	(100,50,20)

sequential information, they suffer from the problem of missing data dependency during the long-term data patterns [32]. LSTM [33] is the fundamental and widely applied architecture of RNNs that is capable of recognizing the pattern of dependency between the sequence of input data and learn the long-term pattern of data. Fig. 3 shows the structure of an LSTM cell. An LSTM is formed by a set of cells, and each cell includes three main layers, namely, *forget gate, input gate,* and *output gate*. The forget gate is responsible for removing the previous information of each cell and functions as follows:

$$f_t = \sigma\left(w_f\left[h_{t-1}, x_t\right] + b_f\right) \tag{5}$$

where w_f and b_f are weights and bias of the cell that are learned during the training phase of the LSTM. Then, the input gate that updates information of cell is calculated as follows:

$$i_t = \sigma\left(w_i\left[h_{t-1}, x_t\right] + b_i\right)\ \bar{C}_t = \tanh\left(w_c\left[h_{t-1}, x_t\right] + b_c\right). \tag{6}$$

Finally, the *output gate* generates the cell output for the next cell and the output of network as follows:

$$C_t = f_t * C_{t-1} + i_t * \bar{C}_t\ h_t = \sigma\left(w_o\left[h_{t-1}, x_t\right] + b_o\right) * \tanh(C_t) \tag{7}$$

Generally, an LSTM network includes one layer of cells. However, increasing the depth of the network elevates its performance and accuracy for learning and recognizing complex sequential patterns [34]. The proposed method includes a stack of LSTMs having various settings to learn the various pattern of network traffic associated with clean and attack scenarios. The number of layers and the capacity of networks are two main settings to design our LSTMs. Table III gives information about the proposed method's LSTMs. Besides, as described in Section III, we train the proposed method for different window sizes of the packet, and therefore, for each window size, the input size of each LSTMs varies. Fig. 4 illustrates an LSTM for a window size of 5.

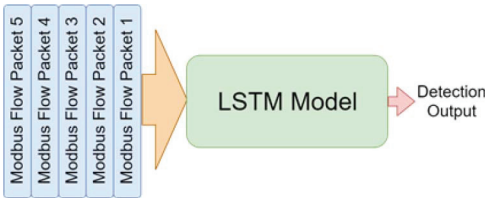

Fig. 4. LSTM for input window size of five packets.

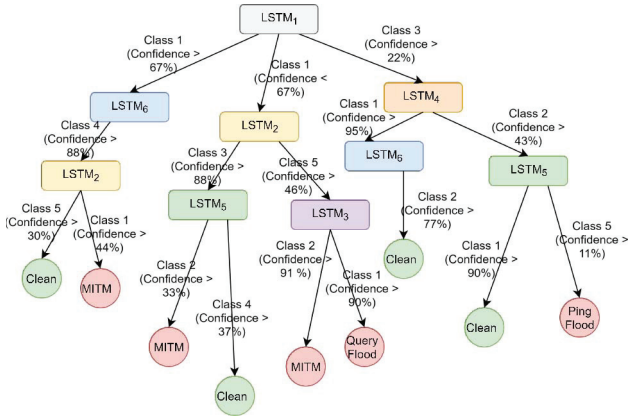

Fig. 5. Schematic view of DT in the proposed method.

B. Ensemble of LSTMs

In order to make an aggregated decision about the output of LSTMs, we integrate a DT component into our proposed method. The DT accepts a collection of confidence rates for each class within the data set and decides about the output. The input of the DT module from LSTMs is as follows:

$$\text{Input of DT} = \left\{\text{LSTM}_{i,c} \text{ where } i \in \{\text{Number of LSTMs}\} \text{ AND } c \in \text{Number of Classes}\right\}. \tag{8}$$

$\text{LSTM}_{i,c}$ refers to the confidence rate of the ith LSTM-trained model for class c. The DT accepts these confidence rates as inputs and hierarchically learns the correlation between the confidence rate of LSTMs and the true label of network traffic. Fig. 5 schematically illustrates how a DT component functions in the proposed method. In other words, DT identifies the manifold of the output space of LSTMs and provides us with an explainable model to decide about the final label.

V. EVALUATIONS AND FINDINGS

In this section, we evaluate the potential of state-of-the-art classification algorithms on the prepared data set (Section V-A). Then, we describe the performance of different LSTMs' in recognizing Modbus cyber attacks (see Section V-B). Findings from Section V-C demonstrate the robustness of the proposed approach in detecting the IoT cyber attack using network traffic. We also discuss the training and inference times of the proposed system.

The experiments are implemented on an Ubuntu 16 system with 128 GB of memory and 32 Core i7 CPUs. All scripts for extracting and preprocessing data as well as learning

TABLE IV
PERFORMANCE OF STATE-OF-THE-ART CLASSIFIERS: A COMPARATIVE SUMMARY

Classifier	Accuracy	Precision	Recall	F-Measure
KNN	85.09%	87.50%	83.33%	85.37%
SVM	86.96%	89.89%	86.96%	88.40%
MLP	88.20%	88.17%	91.11%	89.62%
Random Forest	90.68%	92.22%	91.21%	91.71%

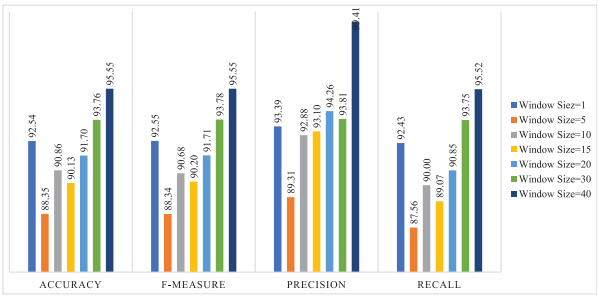

Fig. 6. Performance of LSTM model#1 over Modbus network traffic.

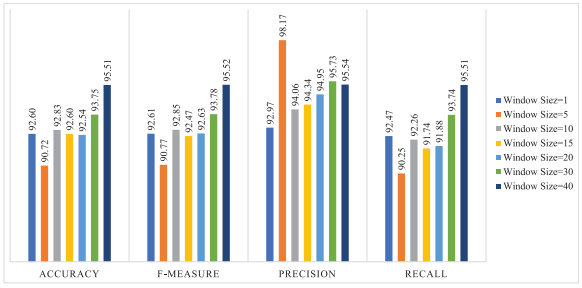

Fig. 7. Performance of LSTM model#2 over Modbus network traffic.

tasks are written in *Python 3.7*. We utilize *Tensorflow*[4] as our DL platform. The experiments are performed for seven different window sizes, namely: {1, 5, 10, 15, 20, 30, 40}, for training the LSTMs. We then apply the tenfold cross-validation technique [35].

A. State-of-the-Art Classifiers

Before evaluating LSTM and the proposed method's outcomes and in order to assess the performance of prevalent classification algorithms, we apply four state-of-the-art classifiers, namely: 1) KNN; 2) multilayer perceptron (MLP); 3) SVM; and 4) RF, on the prepared data set. We use *Scikit-learn*[5] to implement these classification methods. As for KNN, $k = 1$ and for MLP, the size of the hidden layer is set to 200.

Table IV summarizes the results of the experiments.

B. LSTMs

In the first stage of our study, we train a set of LSTMs (see also Section IV-A) and evaluate their performance to identify IoT cyber attacks using network traffic. Specifically, we train six different LSTMs having different structures and on seven different window sizes. The *epoch* for training a deep learner is 200 and batch size sets to 1024. In addition, we utilize the *Adam* optimizer.

During our experiments, we monitor the performance metrics described in Section III-C, in order to analyze the first stage of our proposed method. Figs. 6–11 present the performance of LSTMs in classifying Modbus network traffic for different window sizes. For each metric, the figure includes

[4]https://www.tensorflow.org/
[5]https://scikit-learn.org/

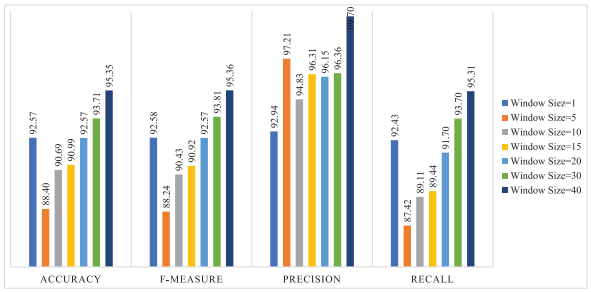

Fig. 8. Performance of LSTM model#3 over Modbus network traffic.

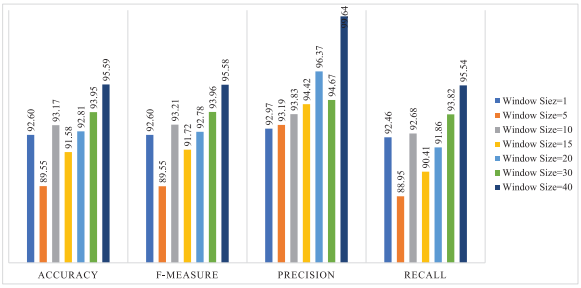

Fig. 9. Performance of LSTM model#4 over Modbus network traffic.

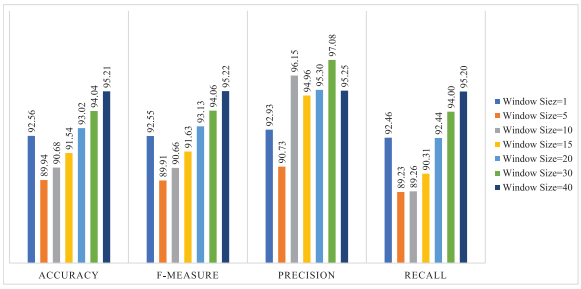

Fig. 10. Performance of LSTM model#5 over Modbus network traffic.

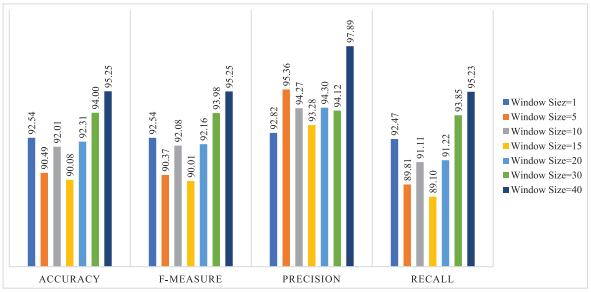

Fig. 11. Performance of LSTM model#6 over Modbus network traffic.

TABLE V
PERFORMANCE OF OUR PROPOSED APPROACH FOR DIFFERENT WINDOW SIZES

Winow size	Accuracy	Recall	Precision	F1-measure
window-size-1	96.459	93.454	94.205	93.812
window-size-5	99.237	99.351	99.265	99.308
window-size-10	99.428	85.211	85.185	85.198
window-size-15	99.370	92.209	99.363	94.594
window-size-20	99.369	98.428	99.185	98.797
window-size-30	99.465	98.579	99.236	98.892
window-size-40	99.620	98.883	99.418	99.142

a group of bar charts that report the metric's highest obtained values for each window size setting.

From the findings, we observe that the LSTM classification approach outperforms the other state-of-the-art classifiers (see Table IV). Also, we observe that the *Precision* of the trained model surpasses other evaluation criteria, and the performance of different LSTMs varies over window sizes (i.e., LSTMs have learned different patterns of network traffic). A general trend is an increased window size that results in increased performance. We also observe that $LSTM_4$ is the most accurate model, with an accuracy rate of 95.59% (for window size $= 40$) and the average accuracy of LSTM models is 92.46%. $LSTM_3$ (for window size $= 40$) is the best model to positively predict samples, with a precision rate of 99.7%. The average precision rate is 95.1%. In terms of the recall metric, $LSTM_4$ for window size $= 40$ achieves 95.54% TP rate and the average is 91.92%. For the mean of precision and recall, $LSTM_4$ for window size $= 40$ has an F-measure of 95.58% and the average is 92.45%.

C. Proposed Approach

Similar to the preceding section, we evaluate the performance of the proposed model for different window sizes while a DT aggregates the output of LSTMs. As shown in Table V, the best accuracy rate obtained is for window size $= 40$, where our proposed approach obtains an accuracy rate of 99.62% and the average accuracy rate is 98.99%. In terms of precision, we achieve 99.41% precision and the average decreases from 95.1% to 96.51%. The proposed method is capable of reaching a detection rate of 99.35% for window size $= 5$, and the average detection rate decreases to 92.32%. In terms of F-measure, our approach obtains 99.30%, and the average decreases from 92.45% to 95.67%. One can observe that the proposed method outperforms single LSTMs and the other state-of-the-art classification algorithms for both maximum and average performance.

D. Time Discussion

In terms of time complexity, an ideal cyber-attack detection should have reasonable and short training and inference times while having acceptable detection performance. During the training time, our proposed method (constructed by a set of LSTMs) achieves acceptable performance within about 25 training epochs that last less than 65 s. Fig. 12 presents the average performance of LSTMs (see also Section IV-A) over training epochs.

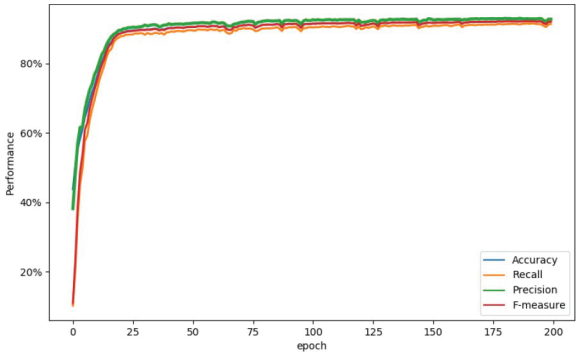

Fig. 12. Average performance of LSTM base models.

In terms of inference time, the proposed method requires only 40 packets in the Modbus network session to detect cyber attacks (see Table V). Based on the data set we used, it requires less than a second for all different attack scenarios to transmit 40 packets over a network session. In addition, the proposed method latency for processing network packets and inferring is approximately 55 ms on average for the experimental workstation.

VI. CONCLUSION

The expanding number of industries utilizing IoT devices partly contributed to an increase in the frequency, size, and severity of cyber attacks against IoT networks; thus, creating an arms race between the cyber defenders and the cyber attackers. In this article, we presented a novel ensemble method to detect IoT cyber attacks over Modbus network traffic. In our approach, we integrated an ensemble of LSTM deep models and aggregated their outputs to achieve enhanced robustness. Findings from our evaluations demonstrated the potential of our approach in an IoT system, where using the DT as an aggregator provides an explainable structure to enhance the transparency of the proposed method [36].

In the future, we will explore the explainability of LSTM models to propose a more transparent DL model for detecting IoT cyber attacks, particularly those in adversarial settings (e.g., battlefields). We also plan to deploy the proposed approach to different IoT protocols and transfer the learned Modbus cyber attacks to other domains.

REFERENCES

[1] K. A. da Costa, J. P. Papa, C. O. Lisboa, R. Munoz, and V. H. C. de Albuquerque, "Internet of Things: A survey on machine learning-based intrusion detection approaches," *Comput. Netw.*, vol. 151, pp. 147–157, Mar. 2019.

[2] N. Moustafa, B. Turnbull, and K.-K. R. Choo, "An ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of Internet of Things," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4815–4830, Jun. 2019.

[3] C. Modi, D. Patel, B. Borisaniya, H. Patel, A. Patel, and M. Rajarajan, "A survey of intrusion detection techniques in cloud," *J. Netw. Comput. Appl.*, vol. 36, pp. 42–57, Jan. 2013.

[4] S. Aljawarneh, M. Aldwairi, and M. B. Yassein, "Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model," *J. Comput. Sci.*, vol. 25, pp. 152–160, Mar. 2018.

[5] E. M. Dovom, A. Azmoodeh, A. Dehghantanha, D. E. Newton, R. M. Parizi, and H. Karimipour, "Fuzzy pattern tree for edge malware detection and categorization in IoT," *J. Syst. Architect.*, vol. 97, pp. 1–7, Aug. 2019.
[6] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, U.K.: MIT Press, 2016.
[7] A. N. Jahromi *et al.*, "An improved two-hidden-layer extreme learning machine for malware hunting," *Comput. Security*, vol. 89, Feb. 2020, Art. no. 101655.
[8] Y. Ye, T. Li, D. Adjeroh, and S. S. Iyengar, "A survey on malware detection using data mining techniques," *ACM Comput. Surveys*, vol. 50, no. 3, p. 41, 2017.
[9] S. Mahdavifar and A. A. Ghorbani, "Application of deep learning to cybersecurity: A survey," *Neurocomputing*, vol. 347, pp. 149–176, Jun. 2019.
[10] H. Karimipour, A. Dehghantanha, R. M. Parizi, K. R. Choo, and H. Leung, "A deep and scalable unsupervised machine learning system for cyber-attack detection in large-scale smart grids," *IEEE Access*, vol. 7, pp. 80778–80788, 2019.
[11] H. HaddadPajouh, R. Khayami, A. Dehghantanha, K.-K. R. Choo, and R. M. Parizi, "AI4SAFE-IoT: An AI-powered secure architecture for edge layer of Internet of Things," *Neural Comput. Appl.*, pp. 1–15, Feb. 2020.
[12] W. Tounsi and H. Rais, "A survey on technical threat intelligence in the age of sophisticated cyber attacks," *Comput. Security*, vol. 72, pp. 212–233, Jan. 2018.
[13] J. Mena, *Machine Learning Forensics for Law Enforcement, Security, and Intelligence*. New York, NY, USA: Auerbach, 2016.
[14] J. Zhao, R. Mortier, J. Crowcroft, and L. Wang, "Privacy-preserving machine learning based data analytics on edge devices," in *Proc. AAAI/ACM Conf. AI Ethics Soc. (AIES'18)*, 2018, pp. 341–346.
[15] N. Goldenberg and A. Wool, "Accurate modeling of modbus/TCP for intrusion detection in SCADA systems," *Int. J. Crit. Infrastruct. Prot.*, vol. 6, no. 2, pp. 63–75, 2013.
[16] H. HaddadPajouh, A. Dehghantanha, R. M. Parizi, M. Aledhari, and H. Karimipour, "A survey on Internet of Things security: Requirements, challenges, and solutions," *Internet Things*, to be published.
[17] D. Oh, D. Kim, and W. Ro, "A malicious pattern detection engine for embedded security systems in the Internet of Things," *Sensors*, vol. 14, no. 12, pp. 24188–24211, 2014.
[18] E. Anthi, L. Williams, M. Słowińska, G. Theodorakopoulos, and P. Burnap, "A supervised intrusion detection system for smart home IoT devices," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 9042–9053, Oct. 2019.
[19] A. Azmoodeh, A. Dehghantanha, and K. R. Choo, "Robust malware detection for Internet of (battlefield) Things devices using deep eigenspace learning," *IEEE Trans. Sustain. Comput.*, vol. 4, no. 1, pp. 88–95, Jan.–Mar. 2019.
[20] M. E. Aminanto, R. Choi, H. C. Tanuwidjaja, P. D. Yoo, and K. Kim, "Deep abstraction and weighted feature selection for Wi-Fi impersonation detection," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 3, pp. 621–636, Mar. 2018.
[21] H. HaddadPajouh, A. Dehghantanha, R. Khayami, and K.-K. R. Choo, "A deep recurrent neural network based approach for Internet of Things malware threat hunting," *Future Gener. Comput. Syst.*, vol. 85, pp. 88–96, Aug. 2018.
[22] A. Azmoodeh, A. Dehghantanha, M. Conti, and K.-K. R. Choo, "Detecting crypto-ransomware in IoT networks based on energy consumption footprint," *J. Ambient Intell. Humanized Comput.*, vol. 9, no. 4, pp. 1141–1152, 2018.
[23] A. Diro and N. Chilamkurti, "Leveraging LSTM networks for attack detection in fog-to-things communications," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 124–130, Sep. 2018.
[24] S. Ghosh and S. Sampalli, "A survey of security in SCADA networks: Current issues and future challenges," *IEEE Access*, vol. 7, pp. 135812–135831, 2019.
[25] B. Zhu and S. Sastry, "SCADA-specific intrusion detection/prevention systems: A survey and taxonomy," in *Proc. 1st Workshop Secure Control Syst. (SCS)*, vol. 11, 2010, p. 7.
[26] S. D. Anton, S. Kanoor, D. Fraunholz, and H. D. Schotten, "Evaluation of machine learning-based anomaly detection algorithms on an industrial modbus/TCP data set," in *Proc. 13th Int. Conf. Availability Rel. Security (ARES)*, 2018, pp. 1–9,
[27] I. Ullah and Q. H. Mahmoud, "A hybrid model for anomaly-based intrusion detection in SCADA networks," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Boston, MA, USA, Dec. 2017, pp. 2160–2167.
[28] D. Peng, H. Zhang, L. Yang, and H. Li, "Design and realization of modbus protocol based on embedded linux system," in *Proc. Int. Conf. Embedded Softw. Syst. Symposia*, Sichuan, China, Jul. 2008, pp. 275–280.
[29] I. Frazão, P. H. Abreu, T. Cruz, H. Araújo, and P. Simões, "Denial of service attacks: Detecting the frailties of machine learning algorithms in the classification process," in *Proc. Int. Conf. Crit. Inf. Infrastruct. Security*, 2018, pp. 230–235.
[30] N. R. Rodofile, K. Radke, and E. Foo, "Framework for SCADA cyber-attack dataset creation," in *Proc. Aust. Comput. Sci. Week Multiconf.*, 2017, p. 69.
[31] A. H. Lashkari, G. D. Gil, M. S. I. Mamun, and A. A. Ghorbani, "Characterization of tor traffic using time based features," in *Proc. 3rd Int. Conf. Inf. Syst. Security Privacy Vol. 1 (ICISSP)*, 2017, pp. 253–262.
[32] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.
[33] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[34] M. Hermans and B. Schrauwen, "Training and analyzing deep recurrent neural networks," in *Proc. 26th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, vol. 1. Lake Tahoe, Nevada, 2013, pp. 190–198.
[35] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-validation," in *Encyclopedia of Database Systems*. Boston, MA, USA: Springer, 2009, pp. 532–538.
[36] D. Gunning, *Explainable Artificial Intelligence (XAI)*. Defense Advanced Research Projects Agency (DARPA), Arlington, VA, USA, 2017.

Mahdis Saharkhizan received the B.Sc. degree in electrical engineering majoring in electronics from Shiraz University, Shiraz, Iran, in fall 2017.

She is mostly interested in application of machine learning in cyber security, IoT, and cloud computing. She has been an Active Member of the IEEE Shiraz University Student Branch and has arranged different seminars and courses to help engineering students enhancing their programming and other necessary skills since 2015.

Amin Azmoodeh received the B.Sc. degree in computer engineering and the M.Sc. degree in artificial intelligence from Shiraz University, Shiraz, Iran, 2006 and 2014, respectively. He is currently pursuing the Ph.D. degree with the University of Guelph, Guelph, ON, Canada.

He is a Microsoft Certified Professional and has several years experience in analyzing and implementing Enterprise Resource Planning software. His main research interests include the theory of machine learning and artificial intelligence and adversarial machine learning. Besides, he is interested in the application of machine learning, especially in cybersecurity and digital forensics.

Ali Dehghantanha (Senior Member, IEEE) received the Ph.D. degree in security in computing from University Putra Malaysia, Seri Kembangan, Malaysia.

He has served as a Senior Lecturer with the University of Sheffield, Sheffield, U.K., and as an EU Marie-Curie International Incoming Fellow with the University of Salford, Greater Manchester, U.K. He is the Director of the Cyber Science Lab, School of Computer Science, University of Guelph, Guelph, ON, Canada. He has served for more than a decade in a variety of industrial and academic positions with leading players in cyber security and artificial intelligence. He has a number of professional certifications, including CISSP and CISM. His main research interests are malware analysis and digital forensics, IoT security, and application of AI in the cyber security.

Kim-Kwang Raymond Choo (Senior Member, IEEE) received the Ph.D. degree in information security from the Queensland University of Technology, Brisbane, QLD, Australia, in 2006.

He currently holds the Cloud Technology Endowed Professorship with the University of Texas at San Antonio, San Antonio, TX, USA.

Dr. Choo was a recipient of the IEEE Technical Committee on Scalable Computing Award for Excellence in Scalable Computing (Middle Career Researcher) in 2019, the UTSA College of Business Col. Jean Piccione and Lt. Col. Philip Piccione Endowed Research Award for Tenured Faculty in 2018, the British Computer Society's Wilkes Award Runner-Up in 2019, the *EURASIP Journal on Wireless Communications and Networking* Best Paper Award in 2019, the Korea Information Processing Society's *Journal of Information Processing Systems* Survey Paper Award (Gold) in 2019, the IEEE Blockchain Outstanding Paper Award in 2019, the Inscrypt Best Student Paper Award in 2019, the IEEE TrustCom Best Paper Award in 2018, the ESORICS Best Research Paper Award in 2015, the Highly Commended Award by the Australia New Zealand Policing Advisory Agency in 2014, the Fulbright Scholarship in 2009, the Australia Day Achievement Medallion in 2008, and the British Computer Society's Wilkes Award in 2008. In 2016, he was named the Cybersecurity Educator of the Year—APAC (Cybersecurity Excellence Awards are produced in cooperation with the Information Security Community on LinkedIn), and he and his team won the Digital Forensics Research Challenge organized by Germany's University of Erlangen–Nuremberg in 2015. He is also a Fellow of the Australian Computer Society and the Co-Chair of the IEEE Multimedia Communications Technical Committee's Digital Rights Management for Multimedia Interest Group.

Reza M. Parizi (Senior Member, IEEE) received the Ph.D. degree in software engineering from University Putra Malaysia, Seri Kembangan, Malaysia, in 2012.

He joined Kennesaw State University, Kennesaw, GA, USA, where he is the Director of Decentralized Science Lab. He was with the New York Institute of Technology, Old Westbury, NY, USA. He is a consummate technologist and software security researcher with an entrepreneurial spirit. His research interests are research and development in decentralized AI, blockchain systems, smart contracts, IoT, and emerging issues in the practice of secure software-run world applications.

Dr. Parizi is a member of the IEEE Blockchain Community and ACM.