Orchestrating SDN Control Plane towards Enhanced IoT Security

Hasan, Tooba; Akhunzada, Adnan; Giannetsos, Athanasios; Malik, Jahanzaib

Published in:
Proceedings of 2020 IEEE Conference on Network Softwarization

Link to article, DOI:
10.1109/NetSoft48620.2020.9165424

Publication date:
2020

Document Version
Peer reviewed version

Orchestrating SDN Control Plane towards Enhanced IoT Security

Tooba Hasan*, Akhunzada Adnan†, Thanassis Giannetsos†, Jahanzaib Malik∓
* Department of Information Security, COMSATS University, Pakistan
‡ DTU Compute, Department of Applied Mathematics and Computer Science, Technical University of Denmark
† National Cyber Security Lab, NUST University, Pakistan
Email: {Toobahasan.int@gmail.com, adnak@dtu.dk, atgi@dtu.dk, Jahanzaibm.ncsael@mcs.edu.pk}

***Abstract*—The Internet of Things (IoT) is rapidly evolving, while introducing several new challenges regarding security, resilience and operational assurance. In the face of an increasing attack landscape, it is necessary to cater for the provision of efficient mechanisms to collectively detect sophisticated malware resulting in undesirable (run-time) device and network modifications. This is not an easy task considering the dynamic and heterogeneous nature of IoT environments; i.e., different operating systems, varied connected networks and a wide gamut of underlying protocols and devices. Malicious IoT nodes or gateways can potentially lead to the compromise of the whole IoT network infrastructure. On the other hand, the SDN control plane has the capability to be orchestrated towards providing enhanced security services to all layers of the IoT networking stack. In this paper, we propose an SDN-enabled control plane based orchestration that leverages emerging Long Short-Term Memory (LSTM) classification models; a Deep Learning (DL) based architecture to combat malicious IoT nodes. It is a first step towards a new line of security mechanisms that enables the provision of scalable AI-based intrusion detection focusing on the operational assurance of only those specific, critical infrastructure components,thus, allowing for a much more efficient security solution. The proposed mechanism has been evaluated with current state of the art datasets (i.e., N_BaIoT 2018) using standard performance evaluation metrics. Our preliminary results show an outstanding detection accuracy (i.e., 99.9%) which significantly outperforms state-of-the-art approaches. Based on our findings, we posit open issues and challenges, and discuss possible ways to address them, so that security does not hinder the deployment of intelligent IoT-based computing systems.**

***Index Terms*—Deep learning, IoT botnet, LSTM, Network security, Software Defined Network.**

I. Introduction

During the last few years, the area of Internet-of-Things (IoT) has met a great development and has the potential to offer a new understanding of our environment that will lead to innovative applications with tangible positive impact on user's experience. This new paradigm leverages the proliferation of modern sensing-capable devices to build a wide-scale information collection network that can provide insights for practically, *anything*, from *anywhere* and at *anytime*. In order to provide concrete implementations for such complex environments, many challenges have to be overcome with *security* and *privacy* being critical pillars [1]: especially in the context of safety applications where critical decisions are based on information collected by users regarding their status or surrounding events [2]. For instance, in the civil protection space, this human-as-a-sensor paradigm is used to harvest information that enhances situational awareness [3].

In the past, extensive research has been conducted towards "*protecting the users from the system*": building secure and accountable IoT architectures that can safe-guard user privacy while supporting user incentive mechanisms. Plethora of research efforts [4], [5] have leveraged advanced cryptographic primitives (e.g., pseudonyms, group signatures, etc.) for protecting users' data from unauthorized access and preventing potential leak of personal identifiable information. However, the question, of how to "*protect the system from the users*" in assessing the trustworthiness of contributed data so that strong guarantees can be provided towards the accuracy and correctness of the system output remains still open [6].

In the machine learning community, security problems have already been addressed in the form of adversarial machine learning [7]. For instance, novelty detection has been addressed to detect anomaly in acoustic data [8]. Game theory has also been exploited in the design of convolutional neural networks to detect image tampering [9]. Concept drift, which is a common phenomenon in IoT data, has also been considered in security problems such as in feature extraction and fraud detection [10]. Yet, most security/adversarial machine learning is based on the assumption that training data are readily available. As such, there are only a few works on unsupervised learning for IoT data such as anomaly detection [11]. Our work addresses this shortcoming in the context of trustworthiness within IoT.

The trustworthiness of collected information is typically studied in relation to the trustworthiness of the human sensors which raises important concerns on the *content integrity*. Data are not necessarily originated from trustworthy sources (e.g., sensors deployed and managed by authorities) but from contributions of any user volunteers that possess a sensing-capable device. This desired openness of IoT systems, such as Mobile Crowd-Sensing (MCS) [1], renders them vulnerable to malicious users that can *pollute* the data collection process, thus, manipulating the system output [12]. A major challenge in these settings is the timely analysis of large amounts of data to produce highly reliable and accurate insights and decisions on the correctness of incoming user reports. Unfortunately, this is not straightforward especially in the presence of intelligent

978-1-7281-5684-2/20/$31.00 ©2020 IEEE

and colluding adversaries trying to manipulate the system's perception of the phenomenon. Data mining and other artificial intelligence methods are among the top methods to gain hidden insights from IoT data, albeit with many challenges [13].

In this context, attackers can use diverse exploits such phishing, DoS, DDoS, Man-in-the-Middle, a variety of malware attacks and Botnets to compromise IoT devices and network sensitive information in an attempt to compromise the functionality of the overall system. Among these, Botnet is considered as one of the most devastating types of attacks capable of paralyzing the entire IoT network stack. A Botnet is a network of collaborating and compromised connected devices whose aim is to conduct online crimes by exploiting vulnerabilities and are remotely controlled by out-of-band command and control channels [14]. Such a channel is controlled by an attacker for stealing users data by sending commands to the target compromised systems [15]. In order to secure the IoT infrastructure, from such sophisticated infiltration techniques, there is a pressing need to formulate comprehensive security mechanisms that, however, do not place any extra burden on the underlying IoT constrained devices and do not undermine system flexibility and scalability [16].

To cope with these pressing security, trust and operational assurance issues, adopting Deep Learning mechanisms that can be orchestrated through the SDN control plane, seems an ideal solution: It can leverage the capabilities of the IoT edge devices, without requiring any additional processing requirements, while providing the necessary scalability envisioned in such environments [17]. SDN centralized control intelligence [18] can provide a flexible, easily configurable and thorough IoT management due to the separation of the fronthaul/backhaul control and data planes [19].

Contributions: In this paper, we propose an SDN-enabled DL framework leveraging LSTM classification models for an early and efficient detection of a wide range of sophisticated attacks in IoT environments. The proposed mechanism is highly scalable and can support any commercial SDN controller. Additionally, the control plane orchestration does not add any extra burden on the underlying IoT resource constraint devices. We have thoroughly tested and evaluated the newly designed architecture using state of the art IoT datasets (i.e., N_BaIoT). Detailed experimentation and analysis show outstanding performance results with (99.9%) detection accuracy that significantly outperforms state-of-the-art approaches. Our proposed solution is scalable and decentralized, removing the need for federated trust of the infrastructure entities in cloud-based environments [20], [21]. This is clearly a viable approach for remedying the limitations of existing detection techniques, nonetheless, there is a need to still overcome a number of open issues towards a holistic end-to-end approach.

II. Towards AI-based Intrusion Detection

Towards the detection of evolving IoT cyber threats and attacks, Recurrent Neural Networks (RNNs) play an important role, especially for network traffic classification, as their inherent characteristics enable them to memorize long term

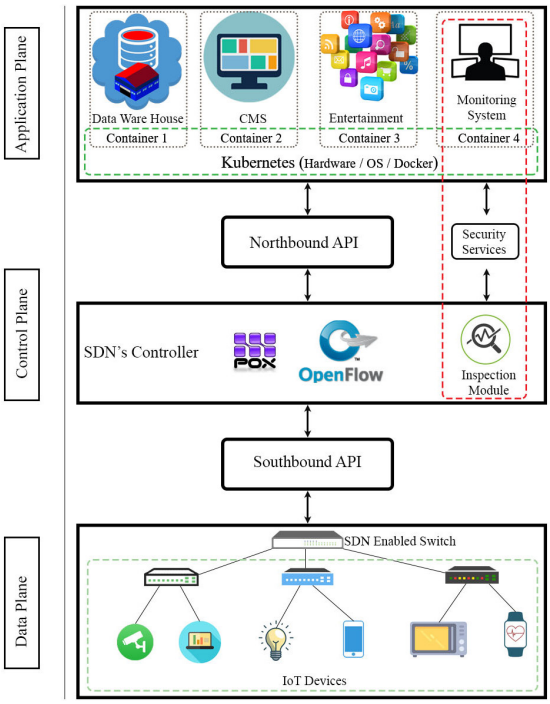

Fig. 1. System Model

sequences for better prediction. In this context, Long Short-Term Memory (LSTM), a variant of RNNs, is capable of malicious traffic classification with high accuracy and minimum false alarm rate. In [18], the authors employ such LSTM classifiers for enhanced detection of IoT threats. The leveraged dataset, for evaluation, comprises network traffic captured from CVUT University (for practical experimentation) demonstrating a detection accuracy of 99.90%. However, the proposed scheme lacks proper evaluation against current state-of-the-art detection mechanisms. In the same line of research, the authors in [22], present a novel approach on malicious packet-level detection by implementing a Deep Learning approach leveraging Bidirectional LSTMs. In this paper, self-generated dataset Mirai botnet and normal IoT traffic are used for training the classifier resulting in a detection accuracy of 96%. In [23], a security solution based on the combined used of Convolutional Neural Networks (CNNS) and Recurrent Neural Networks (RNNs) is presented for inspecting the statistically-based network flow features. Two datasets are employed for evaluation: CTU-13 and ISOT with a mixture of real and synthetic data. Different algorithms are also used (i.e., SVM, KNN, Random Forest, Logistic Regression, and LSTM), while the integration of LSTM shows an accuracy of 99.3%.

The authors in [24] achieve an accuracy of 98% by using Deep Neural Networks based on an underlying (contextual) LSTM architecture for exploiting the content and metadata towards the efficient detection of botnet attacks. They also propose the synthetic minority oversampling for generating a large labeled dataset. The mechanism presented in [25] focuses on detecting and classifying the domain name that does not rely on statistical information. Towards this direction, Deep Learning techniques are used based on LSTMs, RNNs,

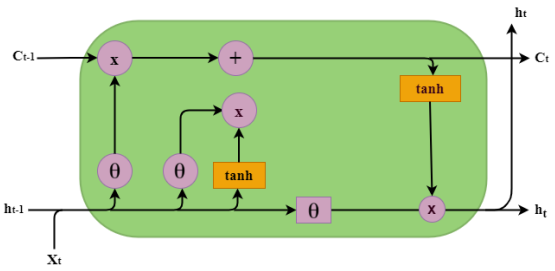

Fig. 2. Architecture of LSTM

CNNs, and CNN-LSTM mechanisms. The evaluation dataset comprises of one billion instances of benign collected records, from Alexa and open-DNS, and malicious data from 17-DGA. The employed approach provides a detection rate of 90%.

Additionally, [26] discuss the importance of Deep Learning as a key enabler for multi-level abstraction of data and for drastically improving speech, objection recognition and detection mechanisms. For classification of images, audio, videos and text files, CNNs and RNNs are executed. Furthermore, this paper also discusses the major progress of Artificial Intelligence (AI) and future impact on various fields. In [27], the authors proposed a fog assisted Intrusion Detection and Prevention System (IDPS) that adds protection at the network edge towards the detection of a variety of threats while providing adequate levels of scalability. The study in [28] provides a secure intrusion detection framework based on Deep Learning-enabled restricted Boltzmann machines (RBM) while providing 95% of detection accuracy.

Overall, the existing literature either lacks a detailed evaluation analysis, against state-of-the-art IoT datasets, or less numbers of instances are utilized both for classification training and testing. Furthermore, not enough attention is given on how to orchestrate such decentralized AI-based detection agents, thus, limiting their applicability and scalability. In contrast, our proposed SDN-enabled mechanism is a first step towards enhancing the security landscape of IoT environments while allowing enhanced scalability and performance.

III. An Architectural Blueprint of LSTM-enabled Intrusion Detection

As described in the previous sections, our envisioned architecture relies on two core pillars: **the employment of advanced LSTM classification models** and the **orchestration of such decentralized LSTM-enabled detection agents through the SDN control plane**. Such a prominent IoT malware detection mechanism is depicted in Figures 1 and 3.

The IoT devices are connected to the backend SDN, through an intelligent Data Plane, while monitoring and security services are orchestrated in the Control Plane. The SDN controller is responsible for orchestrating the various security strategies to control the entire system. For a more detailed description, a thorough and comprehensive overview of such an SDN architecture can be found in [29]–[31].

The proposed mechanism is highly scalable and can be easily customized for integration as an extended module on any commercial SDN controller such as Floodlight, POX, OpenDaylight, etc. It is based on RNNs, a Deep Learning-based technique for detecting sophisticated attack vectors and malicious nodes in the underlying IoT environment. More specifically, we have employed Long Short-Term Memory (LSTM) models, a specific set of RNNs for enriched prediction. The evaluation dataset comprises both normal and malicious data (i.e, attack signatures). The data is split into training and testing sets for preparation and system analysis, respectively. The training data (80% of the original dataset) is fed into the learning algorithm for the configuration and calibration of the classification model while the remaining 20% of data is used for our analysis (Section V).

TABLE I
Description of Proposed Long-Short Term Memory (LSTM)

Algo Family	Layer	No of Layer	Neuron
RNN	*LSTM*	*5*	100,350,300,125,50
	Dense	*4*	250,125,16,2
Activation function	*Relu, Softmax*		
Loss function	categorical cross-entropy		
Optimizer	*Adam*		
Batch-size	256		
Epochs	10		

A. LSTM Models

This section briefly explains the LSTM-based classification models used in our experiments. Long Short-Term Memory [32] is an architecture composed of three integral components (input, output, forget) and one cell (Figure 2). Three gates modulate the information in and out of the cell. The cell structure is designed for remembering the input values at different time intervals, thus, making LSTMs a perfect candidate for enhanced classification, processing and prediction. Table I specifies the configuration parameters of our proposed LSTM model; i.e., layers, neurons, optimizer, batch size, epochs, activation and loss function.

TABLE II
N_BaIoT 2018 Dataset Description

Sr. No	Name	Benign	Attack
1	Ecobee Thermostat	13,000	4,77,565
2	Provision PT 737E Security Camera	32,499	4,77,567
3	Provision PT 838E Security Camera	41,449	4,77,650
4	Samsung SNH 1011 Web Cam	32,450	3,22919
5	Simple Home XCS7 1002 WHT Security Camera	23,150	5,32,130
6	Simple Home XCS7 1003 WHT Security Camera	19,500	4,77,566

TABLE III
THE FEATURE LIST OF N_BAIOT2018 DATASET.

Sr. No	Features	Sr. No	Features	Sr. No	Features
01	MI_dir_L5_weight	40	HH_L3_std	79	HH_L0.1_std
02	MI_dir_L5_mean	41	HH_L3_magnitude	80	HH_L0.1_magnitude
03	MI_dir_L5_variance	42	HH_L3_radius	81	HpHp_L5_weight
04	MI_dir_L3_weight	43	HH_L3_covariance	82	HpHp_L5_mean
05	MI_dir_L3_mean	44	HH_L3_pcc	83	HpHp_L5_std
06	MI_dir_L3_variance	45	HH_L1_weight	84	HpHp_L5_magnitude
07	MI_dir_L1_weight	46	HH_L1_mean	85	HpHp_L5_radius
08	MI_dir_L1_mean	47	HH_L1_std	86	HpHp_L5_covariance
09	MI_dir_L1_variance	48	HH_L1_magnitude	87	HpHp_L5_pcc
10	MI_dir_L0.1_weight	49	HH_L1_radius	88	HpHp_L3_weight
11	MI_dir_L0.1_mean	50	HH_L1_covariance	89	HpHp_L3_mean
12	MI_dir_L0.1_variance	51	HH_L1_pcc	90	HpHp_L3_mean
13	MI_dir_L0.01_weight	52	HH_L0.1_weight	91	HpHp_L3_std
14	MI_dir_L0.01_mean	53	HH_L5_magnitude	92	HpHp_L3_magnitude
15	MI_dir_L0.01_variance	54	HH_L5_radius	93	HpHp_L3_radius
16	H_L5_weight	55	HH_L5_covariance	94	HpHp_L3_covariance
17	H_L5_mean	56	HH_L5_pcc	95	HpHp_L3_pcc
18	H_L5_variance	57	HH_L3_weight	96	HpHp_L1_weight
19	H_L3_weight	58	HH_L3_mean	97	HpHp_L1_mean
20	H_L3_mean	59	HH_L3_std	98	HpHp_L1_std
21	H_L3_variance	60	HH_L3_magnitude	99	HpHp_L1_magnitude
22	H_L1_weight	61	HH_L3_radius	100	HpHp_L1_radius
23	H_L1_mean	62	HH_L3_covariance	101	HpHp_L1_covariance
24	H_L1_variance	63	HH_L3_pcc	102	HpHp_L1_pcc
25	H_L0.1_weight	64	HH_L1_weight	103	HpHp_L0.1_weight
26	H_L0.1_mean	65	HH_L1_mean	104	HpHp_L0.1_mean
27	H_L0.1_variance	66	HH_L1_std	105	HpHp_L0.1_std
28	H_L0.01_weight	67	HH_L1_magnitude	106	HpHp_L0.1_magnitude
29	H_L0.01_mean	68	HH_L1_radius	107	HpHp_L0.1_radius
30	H_L0.01_variance	69	HH_L1_covariance	108	HpHp_L0.1_covariance
31	HH_L5_weight	70	HH_jit_L3_mean	109	HpHp_L0.1_pcc
32	HH_L5_mean	71	HH_jit_L3_variance	110	HpHp_L0.01_weight
33	HH_L5_std	72	HH_jit_L1_weight	111	HpHp_L0.01_mean
34	HH_L5_magnitude	73	HH_jit_L1_mean	112	HpHp_L0.01_std
35	HH_L5_radius	74	HH_jit_L1_variance	113	HpHp_L0.01_magnitude
36	HH_L5_covariance	75	HH_jit_L0.1_weight	114	HpHp_L0.01_radius
37	HH_L5_pcc	76	HH_L1_pcc	115	HpHp_L0.01_covariance
38	HH_L3_weight	77	HH_L0.1_weight	116	HpHp_L0.01_pcc
39	HH_L3_mean	78	HH_L0.1_mean	117	Label

IV. EXPERIMENTAL SETUP

***Datasets*:** We have evaluated our approach under various scenarios by employing real-world datasets. *Real-world* datasets provide us with a good understanding of a classifier's performance in real case scenarios, i.e., deployed IoT sensors measuring noisy phenomena. This data has originated from the N_BaIoT 2018 dataset [33]. It contains 117 attributes which include 116 network features and a tag. The benign and attack samples of various IoT devices in the dataset are 292,044 and 429,8092 respectively (Table III).

This real-world dataset considers two types of Botnet attacks, i.e., Mirari and Gafgyt. Both these types of adversarial behaviours are used to launch Distributed Denial of Service (DDoS) attacks. The data features considered, with all the input values, are shown in Table II.

***Adversarial Behaviour*:** The overall goal of the framework is to detect malicious attacks (e.g., Botnets) in the presence of adversarial users [7] who generate malicious traffic to set the system perception to a faked value. We assume that adversaries collaborate to attack the data collection process. The collaboration is achieved by having injected malicious traffic drawn from the same normal distributions, which are also different to those of the adversary-free datasets.

In our experiments, we measure the performance of the LSTM classifier in terms of its resistance to adversarial data infection. The adversarial choices determine the distortion adversaries try to impose on data trustworthiness. The chosen values for (*mean value*, *standard deviation*) determine the similarity (overlap) between the legitimate and adversarial distributions. Intuitively, adversarial detection decreases with this similarity because the malicious reports introduce values that are very near to the legitimate ones. To increase the probability to detect adversarial reports, which are entered to the machine learning model, we consider the following cases:

Case I: Adversaries may cause significant distortion of the input values by increasing the distance between the mean values of the adversarial and legitimate distributions;

Case II: Adversaries may maximize the system uncertainty by choosing a normal distribution with large SD;

Case III: Adversaries may increase the system uncertainty on the correctness of the input values by selecting an adversarial distribution with equal mean value to the one of the legitimate distribution but with significantly smaller SD.

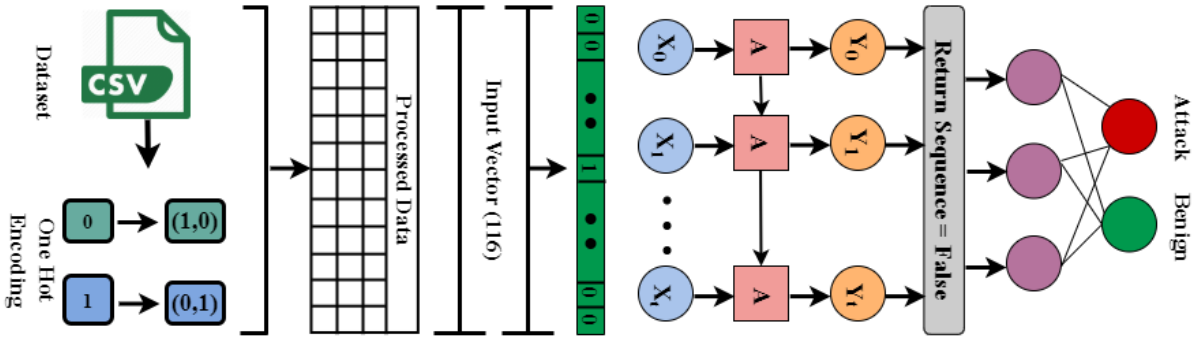

Fig. 3. Architecture of our LSTM-enabled Framework

V. Results & Analysis

The performance metrics of interest used to evaluate and interpret the classifiers' results are the following: (i) *confusion matrix* that contains information about the classifications' results, (ii) *true positive rate* that depicts the percentage of correct predictions, (iii) *false positive rate* the reflects the proportion of instances classified in class x, but belong to a different class, along with all the instances that are not in class x, (iv) *recall* that depicts the proportion of instances that are correctly predicted as positive, and (v) *precision* that estimates the probability that a positive prediction is correct.

In what follows, we discuss the results from the experimentation of our LSTM-enabled intrusion detection architecture. The detection accuracy, precision, recall, and F1-score properties are depicted in Figure 4.

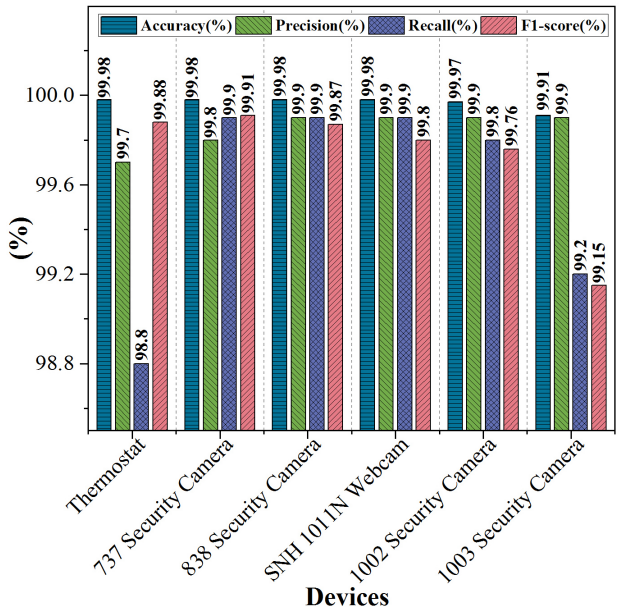

Fig. 4. Accuracy, Precision, Recall and F1-Score of LSTM

Furthermore, for better evaluation of our proposed mechanism, we have also calculated the value of the True Negative Rate (TNR), Negative Predictive Value (NPV) and Matthews Correlation Coefficient (MCC). These are depicted in Figure 5.

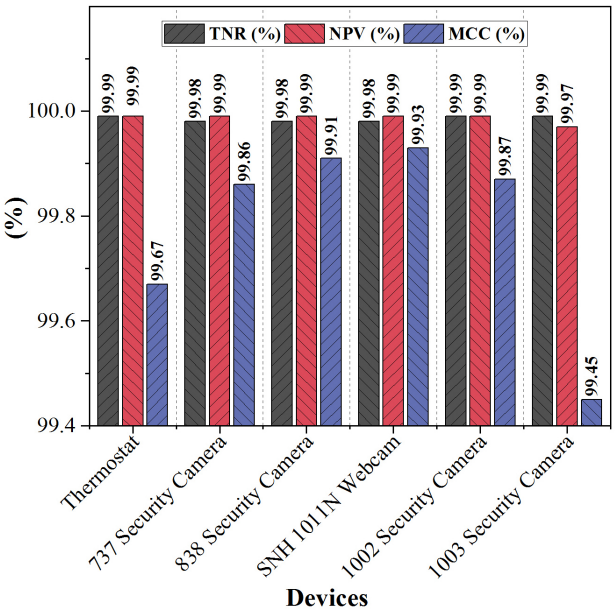

Fig. 5. TNR, NPV, MCC of LSTM

TNR is the ratio of negatives that are perfectly classified, which means the greater the value, the better the performance of the system. NPV is the ratio of positive and negative classification results which basically reflects the ratio between TN and TP values. MCC is an interrelationship between true and predicted instances in binary classification, which means that larger values (between -1 and +1) yield better performance in terms of prediction results.

To evaluate our proposed architecture even further, other properties are also analyzed such as False Negative Rate (FNR), False Positive Rate (FPR), False Discovery Rate (FDR), and False Omission Rate (FOR) are shown in Figure 6. The False Negative Rate (FNR) is the proportion of positive samples that were incorrectly classified. False Positive Rate (FPR), which is also called False Alarm Rate (FAR), represents the ratio between the incorrectly classified negative samples to the total number of negative samples. False Discovery Rate (FDR) and False Omission Rate (FOR) measures complement the PPV and NPV, respectively. The value of FNR, FPR, FDR, FOR are in the range of 0 and 0.13 which is appropriate for

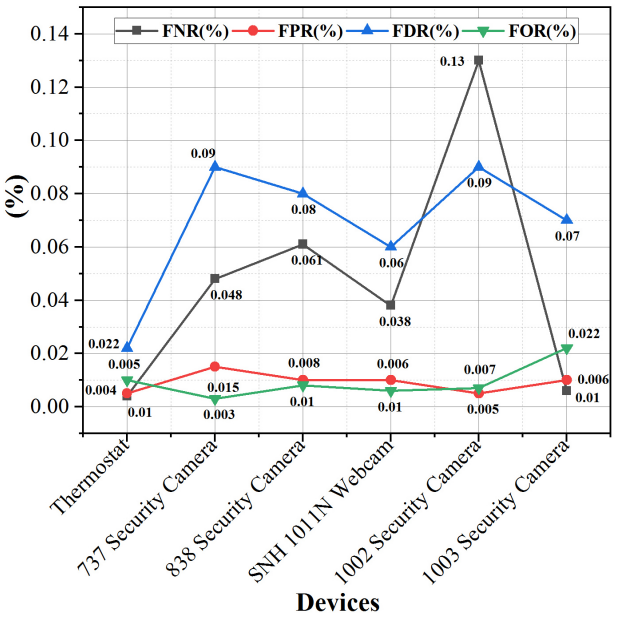

Fig. 6. FNR, FPR, FDR, FOR of LSTM

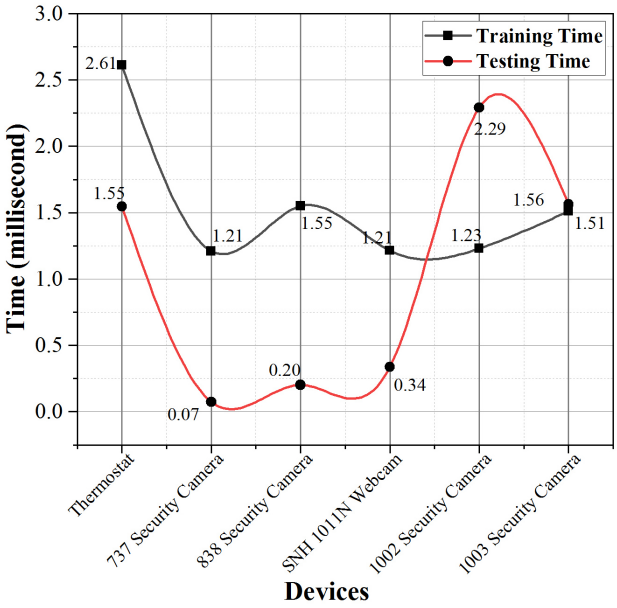

Fig. 7. Training and Testing Time of LSTM

detection of Botnets in various IoT devices.

To assess the time complexity, we have also calculated the training and testing times of our LSTM algorithm while considering the input taken from every IoT device. The results are shown in Figure 7. As we can see, the training time of all IoT devices is almost the same except from the case of one specific sensor (i.e.,Thermostat) that takes more time due to the inherent structure of its value format. Various IoT devices have different time complexity such as *security camera 737*, *security camera 838* and *SNH 1011N webcam* that took less time for completing the testing phase.

As aforementioned, the AU-ROC represents the connection between True Positive (TP) and False Positive (FP) rates. As depicted in Figure 8, the line portrayal of each class close to the x axis shows its high performance.

Finally, Table V provides the comparison of our work with other current-state-of-the-art detection schemes that clearly showcases the scalability and efficiency of our LSTM-enabled architecture compared to these other existing solutions.

VI. Road-Map & Future Prospects

As it is commonly the case for any relatively young research area, the landscape of IoT applications domains is fragmented into various families based on the emerging research challenges. Undoubtedly, data trustworthiness is a prominent challenge with unprecedented number of consequences, should it is not addressed appropriately. We consider this paper as the first step towards the development of a holistic framework, which will improve data trustworthiness in IoT environment that utilize machine learning capabilities. Although, standard classification algorithms are not designed with such requirement in mind, in this paper we assessed their accuracy in presence of data infected with adversarial samples. We strongly believe that this work can be the basis of future research that will attempt to address two main challenges within the IoT security and privacy field: (a) *accuracy*, in the context of concept drift, of IoT data and how this can be balanced with *computational complexity*; and (b) *near real-time performance* of any proposed data trustworthiness framework in the presence of vast volume of IoT data processed; e.g., in the cloud in the form of big data or at the edge of a network

Speaking about improving data verification, future work in the field can be geared towards proposing a combination of machine learning techniques to enhance classification accuracy within the investigated model with other advanced data verification approaches [35]. One shall use *ensemble learning* to utilize multiple classifiers so that they can leverage their advantages and enhance the overall accuracy of IoT data verification [36]. However, ensemble learning will introduce high computational complexity. This generates by default an interesting challenge of investigating trade-offs between accuracy and complexity to determine optimal choices. We envisage that ensemble learning can alleviate the effects of concept drift, which refers to changes in the data distribution over time. Another direction to improve IoT data trustworthiness is the application of deep learning by using autoencoders to train the deep neural network [37]. Due to this technique being computationally expensive, training must be done offline so that classification can be done online.

In the presence of vast volume of IoT data, it is often the case that limited computational resources (e.g., memory, time) and the requirement to make near real-time predictions affect the efficacy of various IoT applications: one of them being Mobile Crowd-Sensing (MCS). Furthermore, vast amount data may be collected so quickly that labeling all items may be delayed or even not possible. To address these issues we envisage the use of game theory, which can determine optimal defense strategies, in the form of thresholds that determine

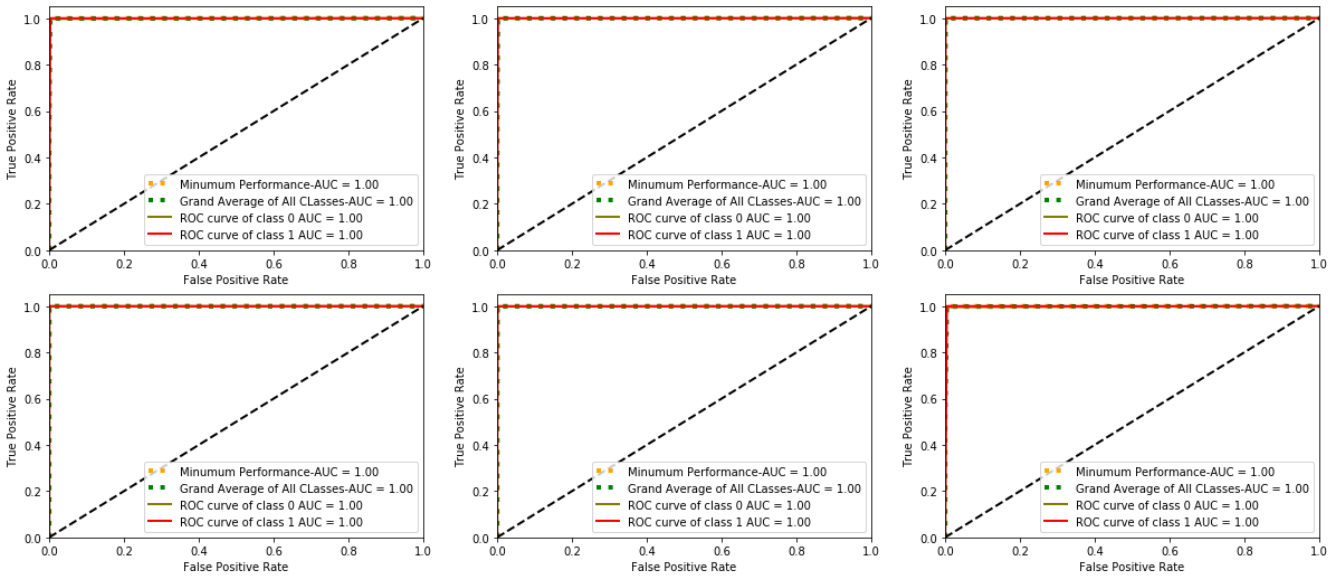

Fig. 8. ROC of Thermostat, 737 Security camera, 838 Security Camera, SNH 1011N Webcam, 1002 Security Camera, 1003 Security Camera

TABLE IV
TABLE FOR COMPARISON OF OUR RESULTS WITH OTHERS

Parameters	**Liu [34]**	**Pekta [23]**	**Shafi [27]**	**Our Work**
Dataset	ISCX(2012)	CTU(2013),ISOT(2010)	UNSW-NB(2015)	N_BaIoT(2018)
Algorithm	CNN	CNN and RNN	RNN, MLP, ADT	LSTM
Binary_class	–	✓	–	✓
Multi_class	✓	–	✓	–
Average Accuracy	99.57	99.3	98.12	99.96
Precision	99.02	90.25	97.29	99.93
Recall	99.26	91.46	91.25	99.88
F1-score	99.10	98.2	96.73	99.88
Testing Time	✓	–	–	✓
FPR	0.11	–	1.02	0.05
Evaluation Metrices(*others*)	✓	–	✓	✓

Others = *TNR, FNR, FDR, FOR, MCC, NPV.*

when the system shall conduct certain required actions, such as re-clustering. Game theory can also support decisions of the defender, i.e., the IoT ecosystem itself, in presence of strategic attackers. These strategies will aim to maximize data trustworthiness in the presence of advanced colluding adversaries who shall utilize sophisticated adversarial strategies targeting data distortion by taking into account more parameters than our current model, such as: geography, users' density and number of submissions per second. A potential approach will seek optimal allocation of defending resources in a similar fashion to our previous work [38].

VII. Conclusions

SDN-enabled DL-based architectures can be a promising solution towards securing IoT infrastructures. This paper presented a first step towards a new line of security mechanisms that enable the provision of scalable AI-based intrusion detection focusing on the operational assurance of only those specific, critical infrastructure components, thus, allowing for a much more efficient security solution. The proposed architecture is based on the use of LSTM classification models, orchestrated in the control plane, for enhanced Botnet detection. Such a control-plane-based orchestration does not place any additional burden on the underlying IoT constrained (edge) devices. Moreover, the proposed framework is highly effective and shows promising results in terms of detection accuracy (99.97%). To the best of our knowledge, this is the first attempt at analyzing such data mining techniques in the context of secure and privacy-preserving IoT, where data trustworthiness is of paramount importance.

After this preliminary analysis, our future plans include exploiting more advanced techniques so as to be able to propose a holistic framework that will improve IoT security and privacy using the right combination of machine learning models. This is a particularly challenging space due to the uncertainty of the classifiers with regards to the real nature of the reports submitted to a reporting station as legitimate.

VIII. Acknowledgment

This work was supported by the European Commission, under the ASTRID and FutureTPM projects; Grant Agreements no. 786922 and 779391, respectively.

REFERENCES

[1] T. Giannetsos, S. Gisdakis, and P. Papadimitratos, "Trustworthy people-centric sensing: Privacy, security and user incentives road-map," in *13th Annual Mediterranean Ad Hoc Networking Workshop*, 2014, pp. 39–46.

[2] J. Whitefield, L. Chen, T. Giannetsos, S. Schneider, and H. Treharne, "Privacy-enhanced capabilities for vanets using direct anonymous attestation," in *2017 IEEE Vehicular Networking Conference (VNC)*, 2017, pp. 123–130.

[3] J. Ballesteros, M. Rahman, B. Carbunar, and N. Rishe, "Safe cities. a participatory sensing approach," in *37th IEEE Conf. on Local Computer Networks*, 2012, pp. 626–634.

[4] M. Shin, C. Cornelius, D. Peebles, A. Kapadia, D. Kotz, and N. Triandopoulos, "Anonysense: A system for anonymous opportunistic sensing," *Pervasive and Mobile Computing*, vol. 7, no. 1, pp. 16–30, 2011.

[5] S. Gisdakis, T. Giannetsos, and P. Papadimitratos, "SPPEAR: security & privacy-preserving architecture for participatory-sensing applications," in *7th ACM Conf. on Security & Privacy in Wireless and Mobile Networks*, 2014, pp. 39–50.

[6] A. Majeed, "Internet of things (iot): A verification framework," in *2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 2017, pp. 1–3.

[7] N. Pitropakis, E. Panaousis, T. Giannetsos, E. Anastasiadis, and G. Loukas, "A taxonomy and survey of attacks against machine learning," *Computer Science Review*, vol. 34, p. 100199, 2019.

[8] E. Principi, F. Vesperini, S. Squartini, and F. Piazza, "Acoustic novelty detection with adversarial autoencoders," in *International Joint Conference on Neural Networks*, 2017, pp. 3324–3330.

[9] A. S. Chivukula and W. Liu, "Adversarial learning games with deep learning models," in *Int. Joint Conference on Neural Networks*, 2017, pp. 2758–2767.

[10] A. D. Pozzolo, G. Boracchi, and O. Caelen, "Credit card fraud detection and concept-drift adaptation with delayed supervised information," in *IEEE Joint Conference on Neural Networks*, 2015.

[11] S. Ahmad, A. Lavin, S. Purdy, and Z. Agha, "Unsupervised real-time anomaly detection for streaming data," *Neurocomputing*, vol. 262, pp. 134–147, 2017.

[12] S. S. Rahman, R. Heartfield, W. Oliff, G. Loukas, and A. Filippoupolitis, "Assessing the cyber-trustworthiness of human-as-a-sensor reports from mobile devices," in *Software Engineering Research, Management and Applications*, 2017.

[13] N. Banerjee, T. Giannetsos, E. Panaousis, and C. C. Took, "Unsupervised learning for trustworthy iot," in *2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2018, pp. 1–8.

[14] S. S. Silva, R. M. Silva, R. C. Pinto, and R. M. Salles, "Botnets: A survey," *Computer Networks*, vol. 57, no. 2, pp. 378–403, 2013.

[15] R. A. Rodríguez-Gómez, G. Maciá-Fernández, and P. García-Teodoro, "Survey and taxonomy of botnet research through life-cycle," *ACM Computing Surveys (CSUR)*, vol. 45, no. 4, pp. 1–33, 2013.

[16] F. Tang, B. Mao, Z. M. Fadlullah, and N. Kato, "On a novel deep-learning-based intelligent partially overlapping channel assignment in sdn-iot," *IEEE Communications Magazine*, vol. 56, no. 9, pp. 80–86, 2018.

[17] P. Krishnan, J. S. Najeem, and K. Achuthan, "Sdn framework for securing iot networks," in *International Conference on Ubiquitous Communications and Network Computing*. Springer, 2017, pp. 116–129.

[18] P. Torres, C. Catania, S. Garcia, and C. G. Garino, "An analysis of recurrent neural networks for botnet detection behavior," in *2016 IEEE biennial congress of Argentina (ARGENCON)*. IEEE, 2016, pp. 1–6.

[19] S. K. Tayyaba, M. A. Shah, O. A. Khan, and A. W. Ahmed, "Software defined network (sdn) based internet of things (iot) a road ahead," in *Proceedings of the International Conference on Future Networks and Distributed Systems*, 2017, pp. 1–8.

[20] A. Carrega, M. Repetto, F. Risso, S. Covaci, A. Zafeiropoulos, T. Giannetsos, and O. Toscano, "Situational awareness in virtual networks: The astrid approach," in *2018 IEEE 7th International Conference on Cloud Networking (CloudNet)*, 2018, pp. 1–6.

[21] A. Michalas and T. Giannetsos, "The data of things: Strategies, patterns and practice of cloud-based participatory sensing," in *International Conference on Innovations in InfoBusiness and Technology (ICIIT)*, 2016, pp. 1–6.

[22] C. D. McDermott, F. Majdani, and A. V. Petrovski, "Botnet detection in the internet of things using deep learning approaches," in *2018 international joint conference on neural networks (IJCNN)*. IEEE, 2018, pp. 1–8.

[23] A. Pektaş and T. Acarman, "Deep learning to detect botnet via network flow summaries," *Neural Computing and Applications*, vol. 31, no. 11, pp. 8021–8033, 2019.

[24] S. Kudugunta and E. Ferrara, "Deep neural networks for bot detection," *Information Sciences*, vol. 467, pp. 312–322, 2018.

[25] R. Vinayakumar, K. Soman, P. Poornachandran, and S. Sachin Kumar, "Evaluating deep learning approaches to characterize and classify the dgas at scale," *Journal of Intelligent & Fuzzy Systems*, vol. 34, no. 3, pp. 1265–1276, 2018.

[26] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[27] Q. Shafi, A. Basit, S. Qaisar, A. Koay, and I. Welch, "Fog-assisted sdn controlled framework for enduring anomaly detection in an iot network," *IEEE Access*, vol. 6, pp. 73 713–73 723, 2018.

[28] A. Dawoud, S. Shahristani, and C. Raun, "Deep learning and software-defined networks: Towards secure iot architecture," *Internet of Things*, vol. 3, pp. 82–89, 2018.

[29] A. Akhunzada, E. Ahmed, A. Gani, M. K. Khan, M. Imran, and S. Guizani, "Securing software defined networks: taxonomy, requirements, and open issues," *IEEE Communications Magazine*, vol. 53, no. 4, pp. 36–44, 2015.

[30] A. Akhunzada, A. Gani, N. B. Anuar, A. Abdelaziz, M. K. Khan, A. Hayat, and S. U. Khan, "Secure and dependable software defined networks," *Journal of Network and Computer Applications*, vol. 61, pp. 199–221, 2016.

[31] A. Akhunzada and M. K. Khan, "Toward secure software defined vehicular networks: Taxonomy, requirements, and open issues," *IEEE Communications Magazine*, vol. 55, no. 7, pp. 110–118, 2017.

[32] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[33] Y. Meidan, M. Bohadana, Y. Mathov, Y. Mirsky, A. Shabtai, D. Breitenbacher, and Y. Elovici, "N-baiot—network-based detection of iot botnet attacks using deep autoencoders," *IEEE Pervasive Computing*, vol. 17, no. 3, pp. 12–22, 2018.

[34] J. Liu, S. Liu, and S. Zhang, "Detection of iot botnet based on deep learning," in *2019 Chinese Control Conference (CCC)*. IEEE, 2019, pp. 8381–8385.

[35] N. Koutroumpouchos, C. Ntantogian, S. Menesidou, K. Liang, P. Gouvas, C. Xenakis, and T. Giannetsos, "Secure edge computing with lightweight control-flow property-based attestation," in *5th IEEE Conference on Network Softwarization, NetSoft 2019, Paris, France, June 24-28, 2019*. IEEE, 2019, pp. 84–92.

[36] M. Woźniak, M. Graña, and E. Corchado, "A survey of multiple classifier systems as hybrid systems," *Information Fusion*, vol. 16, pp. 3–17, 2014.

[37] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, "Multimodal deep learning," in *28th Int. Conf. on machine learning*, 2011, pp. 689–696.

[38] G. Rontidis, E. Panaousis, A. Laszka, T. Dagiuklas, P. Malacaria, and T. Alpcan, "A game-theoretic approach for minimizing security risks in the internet-of-things," in *IEEE Int. Conf. on Communication Workshop*, 2015, pp. 2639–2644.