ENSEMBLE OF PROBABILISTIC LEARNING NETWORKS FOR IOT EDGE INTRUSION DETECTION

Tony Jan and A.S.M Sajeev

Melbourne Institute of Technology, Australia
E-mails: tjan@mit.edu.au (Tony Jan), asajeev@mit.edu.au (A.S.M Sajeev)

ABSTRACT

This paper proposes an intelligent and compact machine learning model for IoT intrusion detection using an ensemble of semi-parametric models with Ada boost. The proposed model provides an adequate real-time intrusion detection at an affordable computational complexity suitable for the IoT edge networks. The proposed model is evaluated against other comparable models using the benchmark data on IoT-IDS and shows comparable performance with reduced computations as required.

KEYWORDS

adaboosted ensemble learning, IoT edge security, machine learning for IoT

1. INTRODUCTION

The Internet of Things (IoT) with massively interconnected cyber-physical devices (CPD) is expected to carry a significant role in mission-critical industry applications. Many CPDs, originally considered unworthy and unintelligent, have been re-configured for cyber communication for IoT services with no security provision. Consequently, many of the CPDs remain vulnerable to cyber attacks [1].

The recent botnet attacks (e.g. Mirai and its variants) have revealed the vulnerabilities of the IoT devices as millions of weak and small IoT devices were duped to sabotage the victim services with over 700 GBPS aggregated data attacks [2]. In response, there has been a great effort to develop an intelligent Intrusion Detection System (IDS) for the IoT networks. The cloud-based IDS utilizes resource-rich remote servers to offload data processing requirements from the IoT devices, but they can only offer reactive and postmortem responses as a remote decision module, as depicted in Figure 1. On another hand, an edge-based IDS is a resource-poor alternative but it can provide proactive and prompt security responses [3].

There have also been approaches combining cloud-based and edge-based intrusion detection systems. For example, Hosseinpour et al [4] have proposed a three-layered approach including the cloud, the fog, and edge layers. The fog is the intermediate layer between the edge devices of the network and the cloud. Generally, as the number of layers increases, the cost of detection increases.

DOI: 10.5121/ijcnc.2018.10608

For real-time and mission-critical IoT applications, the edge-based IDS is preferred, but the challenge remains on how to satisfy data analytic computing requirements using only the limited pool of computing resources at the edge of IoT networks.

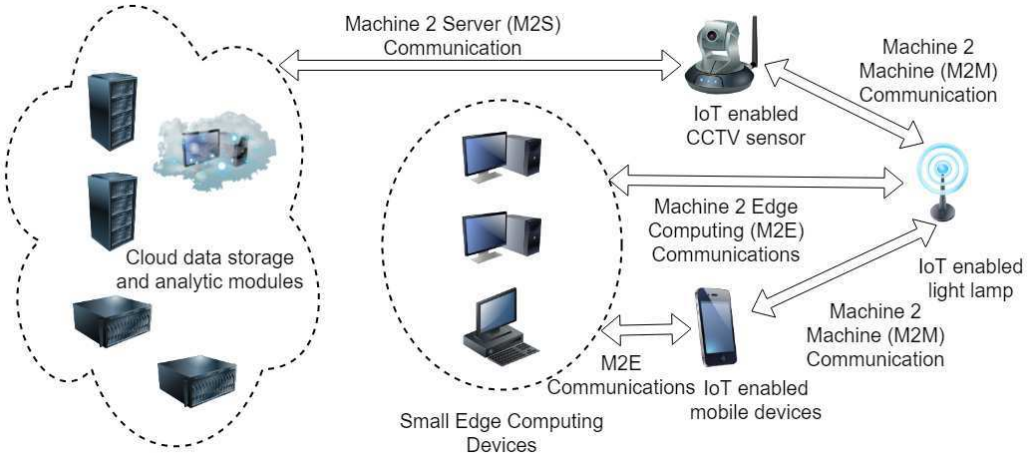

Figure 1. IoT Communication Architecture

Intrusion detection in IoT network, in general, is a major challenge because the data generated by the IoT devices is massive, information-poor, heterogeneous, and dynamic. Detection of any pattern in such a data space is a challenge to most of the simple statistical IDS models, thus it becomes a data mining problem.

In general, a non-parametric or heuristic data mining model is useful in learning and recognizing the underlying patterns from the incomplete and dynamic data sample. They can achieve higher intrusion detection accuracy than the simpler models, but their computational complexity can increase exponentially in modeling massively interconnected IoT networks [5].

The parametric data mining models; on the other hand, are simpler but their detection performance is usually lower and often not acceptable. This phenomenon is well-known as bias and variance dilemma [6]. The challenge is to create a machine learning model that can achieve high accuracy (like a non-parametric model) but maintaining a reduced computational complexity (like a parametric model).

To this aim, we introduce an innovative machine learning model with an ensemble of semi-parametric probabilistic models to approximate a much more complex and powerful non-parametric model (or an ensemble of non-parametric models). The proposed model is to retain the high detection accuracy of a non-parametric model while reducing its computational complexity to the level comparable to the parametric models. A simple analogy is to apply a piece-wise linear regression model to approximate a non-linear regression model under reduced computations.

In this paper, we firstly review the security concerns of IoT networks followed by a literature review of recent machine learning research in IoT-IDS. In the following section(s), our model is introduced and compared to other machine learning models using the benchmark IoT botnet attack data from [7]. The paper concludes with the analysis of the experimental outcomes and the final remarks on IoT security.

2. IoT Security Concerns

In this section, we review the IoT architecture including device, receiver, classifier, and response modules, as shown in Figure 2. We further discuss the security concerns for each part.

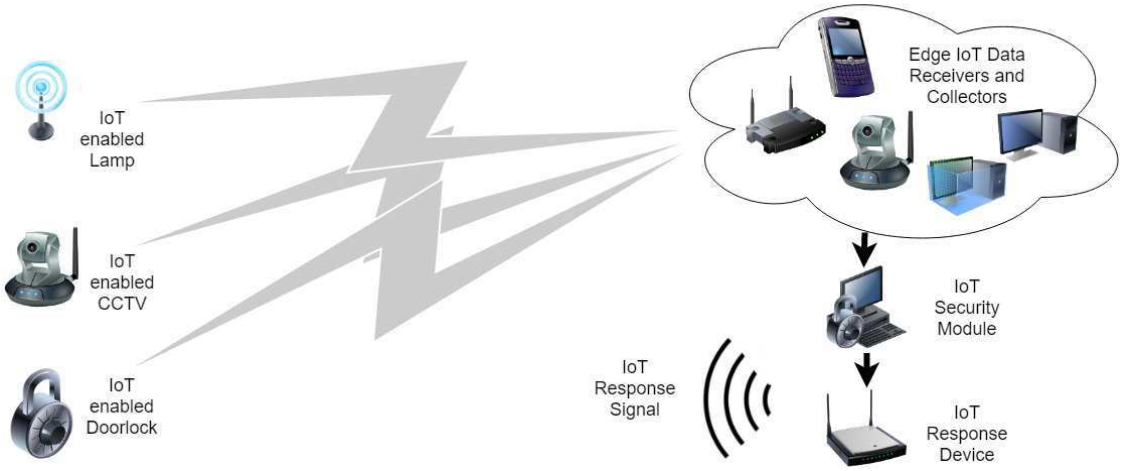

Figure 2. IoT Security Architecture

2.1. IoT Devices

For the sake of ubiquitous computing, many small computing devices have been converted as IoT devices. Some of the devices are resource-enabled with limited memory and processing power, but the majority of others are not given any processing power (in exception of basic Machine to Machine communication capacity). Such devices may include IoT-enabled lamps, doorbells or light switches. The major concern is that these weak and vulnerable resource-depleted IoT devices are still a part of the larger IoT network without any provision of protection.

The small and weak IoT devices are easy targets for binding and privilege escalation attacks. We can harden the IoT devices but its basic Machine to Machine (M2M) messaging [8] will quickly expose the vulnerabilities of these devices [9].

2.2. IoT Data Receivers and Collectors

The IoT communications can be either non-Internet Protocols (IP) based or IP-based. The popular non-IP protocols include ZigBee, WirelessHart and PROFIBUS which are mostly isolated and remain incompatible in heterogeneous IoT networks. The IP protocol is fast becoming the de facto standard in mission-critical IoT applications with innovations in IPSO (IP smart object alliance) and PROFITNET, a real-time Ethernet standard for the IoT networks [10].

For secure IoT communications, we require efficient hardware sensor devices to collect the machine 2 machine or machine 2 server signals and communicate them securely to the edge network controller. The wireless signals can be easily eavesdropped and spoofed, hence we need some access control and integrity checking modules in place.

2.3. IoT Data Analytic Module

The data analytic module on the edge network is to receive and analyze the IoT data in real time as described in Figure 2. The edge computing will have limited computing resources (in comparison to the cloud-based servers). The edge computing cluster is likely to consist of resource-enabled IoT devices, small office computer controller, and personal computing devices in the proximity of the IoT edges. The edge computing can make the use of intelligent distributed processing through either fog computing or other forms of distributed computing [3].

2.4. IoT Responses and Controls

The majority of current system responses are based on disruptive incident response paradigm in which the devices are powered off or restarted. These disruptive responses, in fact, serve the purpose of the attackers. For mission-critical IoT services, we need more sophisticated security responses. The advanced method is to segment the network for efficient security responses [11].

3. LITERATURE REVIEW

Intelligent network intrusion detection is a timeless challenge in machine learning community. In particular, IoT edge computing introduces an even more challenge because of their limited computational resources. The general consensus is to deploy a form of distributed computing and shared memory amongst the small IoT devices and mobile devices in the proximity of the IoT edge as shown in Figure 3. The distributed computing is to offer virtualized server to run data analytic processing for IoT-IDS over the IoT edge devices.

In this section, we review the latest advances in distributed computing over the IoT edge networks followed by some recent works on the use of machine learning models on the IoT edges.

3.1. Distributed Computing

There has been some significant research work on embedding data analytic modeling in the edge IoT devices for proactive critical responses.

The IoT edge computing is mainly used for computational offloading of data storage and analytic processing. The offloading must take into consideration of the dynamic nature of network access requirements, number of edge devices, and available computational resources at the edge devices. We must take into consideration the granularity and hierarchy of edge network topology and how to dynamically partition the application for offloading [12].

In literature, MAUI [13] offers code offloading for adaptive utilization of network resources. COMET [14] offers virtual machine synchronization and shared memory over the IoT edge network, and ThinkAir [15] uses parallel virtualization over the IoT edge devices for distributed data processing. Nishio et al [16] introduces the mobile cloud in which small portions of processing tasks are distributed amongst small mobile devices under the direction of the supervisory control module in awareness of latency and resource optimization.

The underlying idea is to divide the processing task into small modules for distributed processing in awareness of latency and resource optimization amongst the myriad of small mobile devices as shown in Figure 3. The challenge remains on how to segment the data analytic processing tasks into small modules for distributed processing.

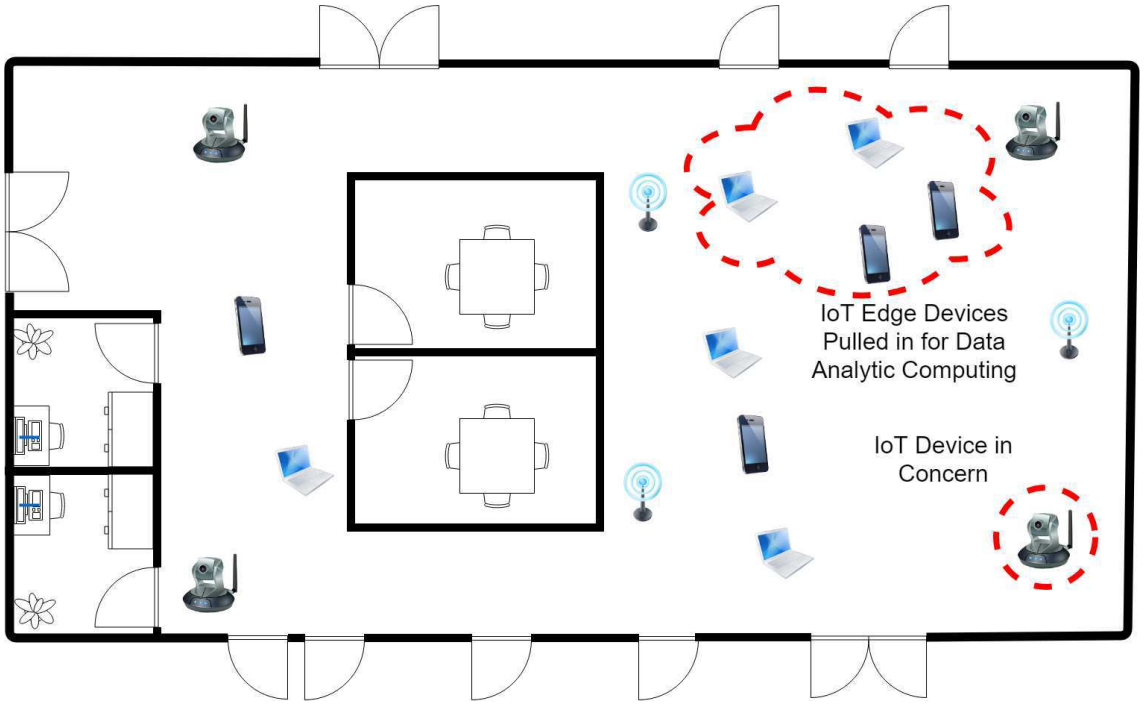

Figure 3. Edge Devices Pooling

3.2. Machine Learning Models

Data analytic computing is inherently intensive, and it may require distributed processing as discussed in this subsection.

A well-known heuristic model such as Multi-Layer Perceptron (MLP) is known for deep learning through multiple layers of interconnected intricate memory modules. MLP is not suitable for distributed processing as the hierarchical segmentation of the network itself is a challenge. The reassembly of the outputs from the segmented components may not reflect the true learning of the global MLP [17].

Another popular model such as Self Organizing Map (SOM) is useful in uncovering data patterns. An advanced variant of SOM has been applied to IoT resource scheduling and sharing with the stringent restriction on energy consumption. This work is appropriate for cloud-server based processing but not suitable for edge-based processing because the sample data points increase exponentially in predictive modeling of IoT network usage requirements, capacities, and availability [18]

Other popular models such as Support Vector Machine (SVM) and its advanced variant such as knn-SVM-PSO [19] are exciting advancement; however, again, they are developed for remote cloud server processing, not for the resource limited IoT edge devices [20].

Given the challenge of the resource-limited IoT edge devices, an intuitive solution is to use an ensemble of simple (data analytic) models. In such case, each data analytic component is already well segmented from the inception; and there are some options to intelligently combine the outcomes from the disjoint learning modules for optimal global learning.

In particular, we take a great interest in Adaboost ensemble of simple semi-parametric probabilistic (kernel-based) models for IoT-IDS on the edges. This model has shown to outperform other complex (non-parametric) models in other dynamic modeling applications [21]. In this paper, we further adjust and modify the model to be suitable for IoT-IDS over the IoT edge networks.

In the following section, we examine the aforementioned model and examine its architecture to make it suitable for IoT edge computing. The proposed model is discussed in detail and compared against other related state-of-the-art models in a simple experiment.

4. PROPOSED MODEL

The proposed model consists of two modules: Adaptive Booster (AB) and Ensemble of Weak Probabilistic Learners (EPL).

4.1. AdaBoost (AB) Module

The weights for each base hypothesis are updated to minimize the model error while maximizing the diversity in each base learner [22]. The procedures are explained below:

The initial weak base learner is constructed based on the available sample data

$$(x_1,\ y_1),\ \ldots,\ (x_n, y_n),\ y_i \in \{-1,\ +1\} \tag{4.a}$$

For the first weak base learner, the weights are simply initialized as

$$W_1(i) = \frac{1}{N} \tag{4.b}$$

The weights of the clusters are iteratively updated (e.g strengthened) based on the model error from the previous set of weights (of the previous weak base learner) as

$$W_{t+1}(i) = \frac{W_t(i) exp(-\alpha_t y_i h_t(x_i))}{Z_t} \tag{4.c}$$

where

$$\alpha_t = \frac{1}{2} \log\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right) \tag{4.d}$$

and

$$h_t = arg \min_{h_j} \varepsilon_j = \sum_{i=1}^{n} W_t(i)\ [[y_i \neq h_j(x_i)]] \tag{4.e}$$

The iteration of weight updates will construct a series of weak base classifiers. The final classifiers is the ensemble of the weak base classifiers. The process is described in Figure 4.

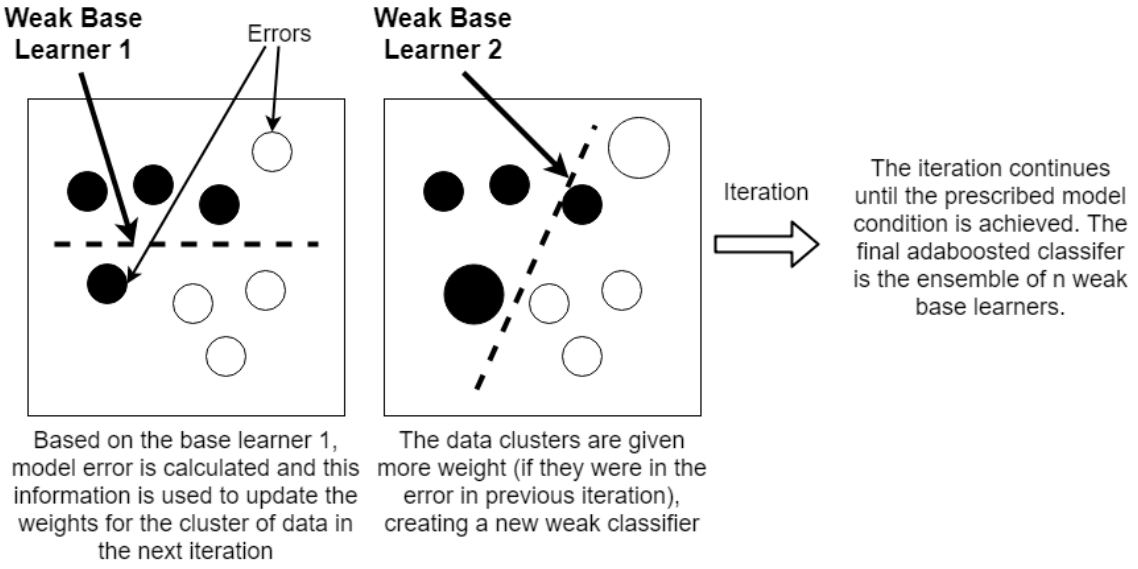

Figure 4. Adaboost Scheme

4.2. The Base Learner

For our base learner, we use kernel-based probabilistic learning network with strong Bayesian statistical framework (instead of linear regression model) [23]. This model can adjust its attributes from parametric to non-parametric modeling by the selection of a single smoothing value. The smoothing value can adjust the granularity of the clusters enabling the approximation of non-parametric model with a simpler semi-parametric model as shown in Figure 5.

This approach can reduce the model complexity whilst maximizing the learning in Adaboost.

The semi-parametric base classifier model is represented as:

$$\hat{y}(\underline{x}) = \frac{\sum_{i=0}^{M} Z_i y_i f_i(\underline{x} - \underline{c}_i, \sigma)}{\sum_{i=0}^{M} Z_i f_i(\underline{x} - \underline{c}_i, \sigma)} \tag{4.f}$$

with semi-parametric approximation shown below:

$$\sum_{i=0}^{Z_k} f_i(\underline{x} - \underline{x}_i, \sigma) \approx Z_k f_k(\underline{x} - \underline{c}_k, \sigma) \tag{4.g}$$

Where c_i is the center vector for class i in the input space, $f_i(x, \sigma)$ is the radial basis function with centre x and the width parameter σ , y_i is the output related to $\underline{c}_i$, Z_i is the number of vectors x_j associated with centre $\underline{c}_i$. $\sum_i Z_i = NV \sum_i Z_i = NV$is the total number of training vectors. Equation 4.g represents semi-parametric approximation of data sample points to simplify the data space as shown Figure 5. The semi-parametric approximation (cluster) is shown with dotted lines in Figure 5.

The proposed base model is a simple semi-parametric model with a set of data points (centres of each clusters) to represent points of influence. The influence is determined by the size of the training data samples in that particular cluster. In our proposed model, the influence on the single point is further adjusted by the weighting factor from the Adaboost as shown in Figure 5.

As mentioned above, the dotted circles present the semi-parametric approximation of the data

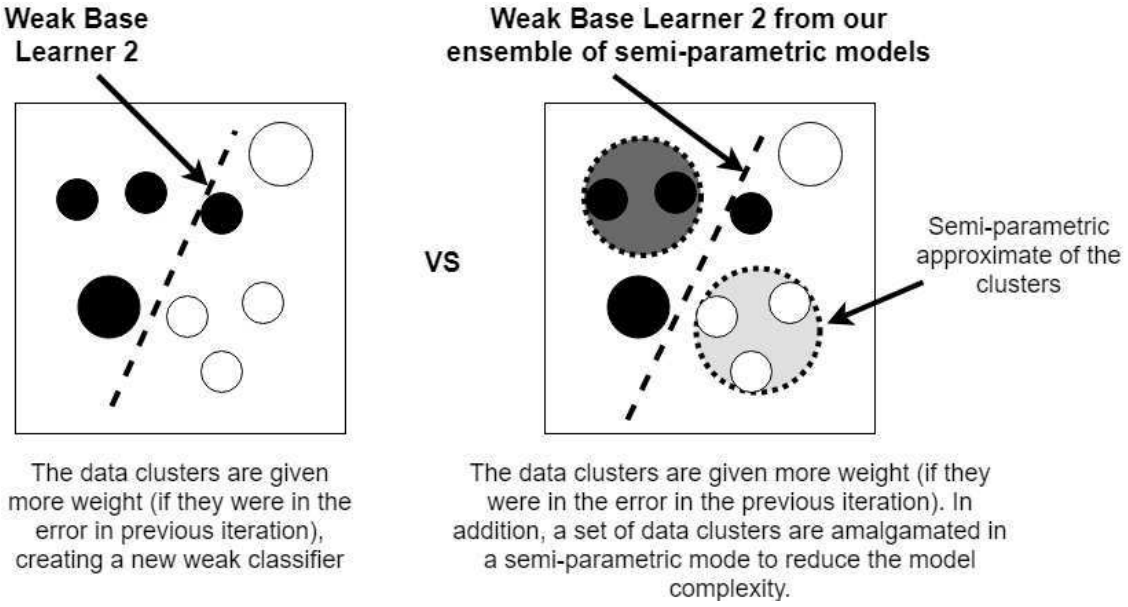

Figure 5. Comparison of Adaboost and Our Model

clusters. In the newly proposed model, the dotted circles are further enlarged by the adjusting weight factor from Adaboot.

4.3. Ensemble Model

This iterative learning is to reduce the model complexity by semi-parametric approximation of the data points (by amalgamation of data points) while to maximize the learning by iterative cluster weight adjust by Adaboot.

The difference(s) between our proposed approach and normal Adaboot is that, in our proposed model, Adaboost weight update is applied to the centre of data cluster instead of each data samples, making the process far more computationally compact.

The final model is an intelligent and compact ensemble of semi-parametric base learners which can provide high detection accuracy at significantly reduced computational cost, as shown in Figure 6.

5. EXPERIMENTAL RESULTS

The proposed model is applied to the detection of botnet attacks on the IoT devices using the benchmark data from [7]. Botnet attack refers to an attack in which a set of host computers or IoT devices are compromised and duped to initiate DDoS (distributed denial of service) attacks on the victim networks or infrastructure.

For botnet attack detection, we use the collection of destination/source IP addresses and their associated port numbers. The abnormal connections within a specified time window is to trigger the alert in the intrusion detection system. The abrupt derivative change in network behavior is

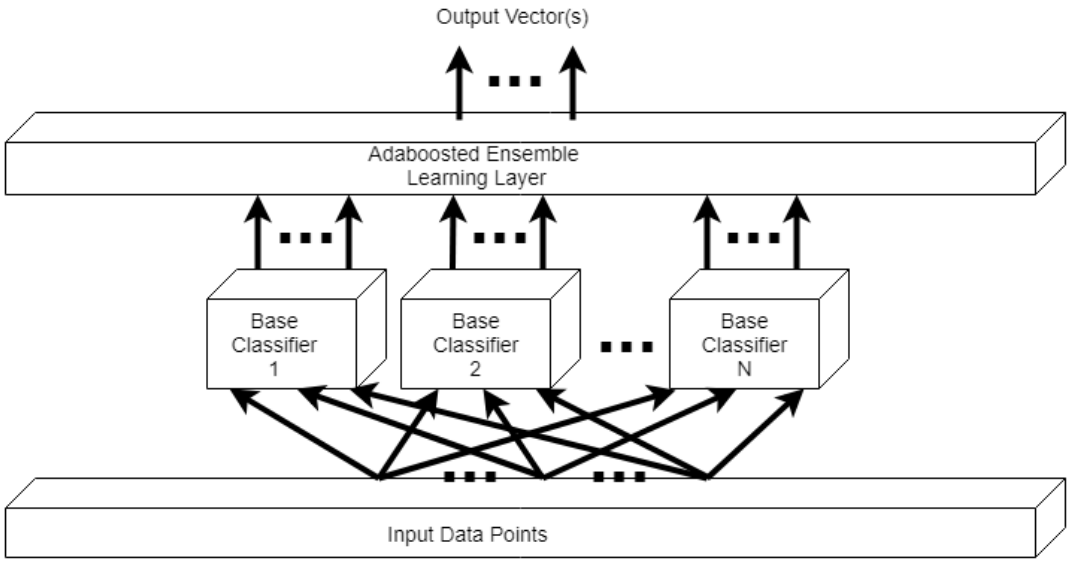

Figure 6. Learning Model Architecture

also to trigger IDS.

For a comprehensive testing of IDS over Botnet attacks, we select to use the benchmark data offered by Meidan et al in [7].

The data attribute information includes:

** Stream aggregation:*

- *H: Stats summarizing the recent traffic from this packet's host (IP)*
- *HH: Stats summarizing the recent traffic going from this packet's host (IP) to the packet's destination host.*
- *HpHp: Stats summarizing the recent traffic going from this packet's host+port (IP) to the packet's destination host+port. Example 192.168.4.2:1242 -> 192.168.4.12:80*
- *HH_jit: Stats summarizing the jitter of the traffic going from this packet's host (IP) to the packet's destination host.*

** Time-frame (The decay factor Lambda used in the damped window): How much recent history of the stream is captured in the statistics L5, L3, L1, ...*

** The statistics extracted from the packet stream: weight: The weight of the stream (can be viewed as the number of items observed in recent history)*

- *mean:*
- *std:*
- *radius: The root squared sum of the two streams' variances*

- *magnitude: The root squared sum of the two streams' means*
- *cov: an approximated covariance between two streams*
- *pcc: an approximated covariance between two streams*

From the benchmark dataset, 2/3 of benign data is selected to train the classifier to learn the normal operating conditions of 9 IoT devices. The other 1/3 of benign data is combined with the attack data to create a set of testing data. There are several types of attacks, but for the simplicity of the experiment, we categorize data as either benign or malicious.

The experimental outcome is a simple confusion matrix. The evaluation was carried out using the Matlab Simulink code exported to run on IoT-edge-hub simulated on the Microsoft Azure platform.

Table 1. Experimental comparison of IoT IDS

Classifier	Detection Accuracy	Computational Time (in ratio)
MLP (multi layer perceptron)	92	1
Hierarchical MLP	98	12+
Self-organising map (SOM)	88	6+
Distributed SOM	93	20+
Online ensemble of parametric model	94	0.37
Our proposed ensemble of semi-parametric models	94	0.71

The experiment compares the performance of the proposed model against the other models in terms of detection accuracy and the computational cost. The computational cost is estimated by the average learning time, given the same computational environment.

The proposed model showed comparable classification performance to the other models including the boosted ensemble of parametric models. The boosted ensemble of parametric models was compact and useful; however, it could not detect intrusion under somewhat more demanding input episodes which required non-parametric modeling. The proposed model, on the other hand, could traverse between the characteristics of parametric model and non-parametric model handing better diverse type of input episodes.

The classification performance was very high for all the testing models as each model was optimized according to the well-known model selection (e.g. sizes of layers and learning units) practices with rigorous testing [24].

6. CONCLUSION

This paper examined the opportunity of utilizing innovative machine learning tools in practical deployment of real-time IoT-IDS.

We introduced an ensemble of semi-parametric probabilistic learning models for intrusion detection at the IoT edge network. The proposed model was applied to IoT-IDS using the benchmark data for comparative analysis against other state-of-the-art IDS (using other machine learning algorithms). The proposed model has shown an improved intrusion detection performance given much constrained computational resources at the edge of IoT networks.

IDS at IoT-edge can greatly improve the responsiveness of the IoT network against real-time attacks, thereby significantly improving the overall IoT security. As the IoT poses to bring a new era of the inter-connected world, such a responsive and prompt security system can be very timely and useful.

REFERENCES

[1] Muhammad Umar Farooq, Muhammad Waseem, Anjum Khairi, and Sadia Mazhar, (2015) "A critical analysis on the security concerns of internet of things (IoT)", *International Journal of Computer Applications*, Vol. 111, No. 7.

[2] Manos Antonakakis, Tim April, Michael Bailey, Matt Bernhard, Elie Bursztein, Jaime Cochran, Zakir Durumeric, J Alex Halderman, Luca Invernizzi, Michalis Kallitsis, et al., (2017) "Understanding the mirai botnet", in *USENIX Security Symposium*.

[3] Antonio Brogi and Stefano Forti, (2017) "QoS-aware deployment of IoT applications through the fog", *IEEE Internet of Things Journal*, Vol. 4, No. 5, pp. 1185–1192.

[4] Farhoud Hosseinpour, Payam Vahdani Amoli, Juha Plosila, Timo Hämäläinen, and Hannu Tenhunen, (2016) "An Intrusion Detection System for Fog Computing and IoT based Logistic Systems using a Smart Data Approach", *International Journal of Digital Content Technology and its Applications*, Vol. 10.

[5] Bernard W Silverman, (2018) *Density estimation for statistics and data analysis*, Routledge.

[6] Ron Kohavi, David H Wolpert, et al., (1996) "Bias plus variance decomposition for zero-one loss functions", in *ICML*, Vol. 96, pp. 275–83.

[7] Yair Meidan, Michael Bohadana, Yael Mathov, Yisroel Mirsky, Dominik Breitenbacher, Asaf Shabtai, and Yuval Elovici, (2018) "N-BaIoT: Network-based Detection of IoT Botnet Attacks Using Deep Autoencoders", *arXiv preprint arXiv:1805.03409*.

[8] Alya Geogiana Buja, Shekh Faisal Abdul-Latip, and Rabiah Ahmad, (2018) "A Security Analysis of IoT Encryption: Side-channel Cube Attack on Simeck32/64", *arXiv preprint arXiv:1808.03557*.

[9] Ryan Williams, Emma McMahon, Sagar Samtani, Mark Patton, and Hsinchun Chen, (2017) "Identifying vulnerabilities of consumer Internet of Things (IoT) devices: A scalable approach", in *Intelligence and Security Informatics (ISI), 2017 IEEE International Conference on*, pp. 179–181.

[10] Sudhi R Sinha and Youngchoon Park, (2017) *Building an Effective IoT Ecosystem for Your Business*, Springer.

[11] Briana Arrington, LiEsa Barnett, Rahmira Rufus, and Albert Esterline, (2016) "Behavioral modeling intrusion detection system (bmids) using internet of things (iot) behavior-based

anomaly detection via immunity-inspired algorithms", in *Computer Communication and Networks (ICCCN), 2016 25th International Conference on*, pp. 1–6.

[12] Otávio Carvalho, Manuel Garcia, Eduardo Roloff, Emmanuell Diaz Carreño, and Philippe OA Navaux, (2017) "IoT Workload Distribution Impact Between Edge and Cloud Computing in a Smart Grid Application", in *Latin American High Performance Computing Conference*, pp. 203–217.

[13] Eduardo Cuervo, Aruna Balasubramanian, Dae ki Cho, Alec Wolman, Stefan Saroiu, Ranveer Chandra, and Paramvir Bahl, (2010) "MAUI: making smartphones last longer with code offload", in *Proceedings of the 8th international conference on Mobile systems, applications, and services*, pp. 49–62.

[14] Mark S Gordon, Davoud Anoushe Jamshidi, Scott A Mahlke, Zhuoqing Morley Mao, and Xu Chen, (2012) "COMET: Code Offload by Migrating Execution Transparently.", in *OSDI*, Vol. 12, pp. 93–106.

[15] Sokol Kosta, Andrius Aucinas, Pan Hui, Richard Mortier, and Xinwen Zhang, (2012) "Thinkair: Dynamic resource allocation and parallel execution in the cloud for mobile code offloading", in *Infocom, 2012 Proceedings IEEE*, pp. 945–953.

[16] Takayuki Nishio, Ryoichi Shinkuma, Tatsuro Takahashi, and Narayan B Mandayam, (2013) "Service-oriented heterogeneous resource sharing for optimizing service latency in mobile cloud", in *Proceedings of the first international workshop on Mobile cloud computing & networking*, pp. 19–26.

[17] Elike Hodo, Xavier Bellekens, Andrew Hamilton, Pierre-Louis Dubouilh, Ephraim Iorkyase, Christos Tachtatzis, and Robert Atkinson, (2016) "Threat analysis of IoT networks using artificial neural network intrusion detection system", in *Networks, Computers and Communications (ISNCC), 2016 International Symposium on*, pp. 1–6.

[18] Nof Abuzainab, Walid Saad, Choong-Seon Hong, and H Vincent Poor, (2017) "Cognitive hierarchy theory for distributed resource allocation in the internet of things", *arXiv preprint arXiv:1703.07418*.

[19] Abdulla Amin Aburomman and Mamun Bin Ibne Reaz, (2016) "A novel SVM-kNN-PSO ensemble method for intrusion detection system", *Applied Soft Computing*, Vol. 38, pp. 360–372.

[20] Wathiq Laftah Al-Yaseen, Zulaiha Ali Othman, and Mohd Zakree Ahmad Nazri, (2017) "Multi-level hybrid support vector machine and extreme learning machine based on modified K-means for intrusion detection system", *Expert Systems with Applications*, Vol. 67, pp. 296–303.

[21] Milad Yousefi, Moslem Yousefi, Ricardo Poley Martins Ferreira, Joong Hoon Kim, and Flavio S Fogliatto, (2018) "Chaotic genetic algorithm and Adaboost ensemble metamodeling approach for optimum resource planning in emergency departments", *Artificial intelligence in medicine*, Vol. 84, pp. 23–33.

[22] YANG Xinwu, M A Zhuang, and YUAN Shun, (2016) "Multi-class Adaboost Algorithm Based on the Adjusted Weak Classifier", *Journal of Electronics & Information Technology*, Vol. 38, No. 2, pp. 373–380.

[23] Anthony Zaknich, (1998) "Introduction to the modified probabilistic neural network for general signal processing applications", *IEEE Transactions on Signal Processing*, Vol. 46, No. 7, pp. 1980–1990.

[24] Elike Hodo, Xavier Bellekens, Andrew Hamilton, Pierre-Louis Dubouilh, Ephraim Iorkyase, Christos Tachtatzis, and Robert Atkinson, (2016) "Threat analysis of IoT networks using artificial neural network intrusion detection system", in *Networks, Computers and Communications (ISNCC), 2016 International Symposium on*, pp. 1–6.