A MODERN APPROACH TOWARDS NETWORK INTRUSION DETECTION USING MACHINE LEARNING ALGORITHMS FOR IOT

Rohini Jogdand[1], Komal Jitakar[2], Chaturthi Bhaskar[3]

[1,2,3]*Department of Electronics and Telecommunication Engineering, Dr. D. Y. Patil Institute of Engineering, Management & Research, Akurdi, Pune-44*

Abstract: *With the continual development of network technology, security problems within the network are emerging one after another, and it's becoming more and harder to ignore. For the present network administrators, the way to successfully prevent malicious network hackers from invading, in order that network systems and computers are at Safe and normal operation is an urgent task. This paper proposes a network intrusion detection method supported deep learning. This method uses big data with deep confidence neural network to extract features of network monitoring data, and uses BP neural network as top level classifier to classify intrusion types with the help of machine learning algorithms. In this research paper, our main objective is on the IoT NIDS deployment via Machine learning algorithms and deep learning which have good success probability in security and privacy. This survey provides a comprehensive review of NIDS's deployment over different aspects of machine learning techniques for Internet of Things, likewise other top surveys focusing on the traditional systems. The results show that the proposed method features a significant improvement over the normal machine learning accuracy.*

Keywords: Intrusion, privacy, security, machine learning, networking, deep learning.

I. Introduction

Security is the major issue in networking. The issue is to prevent and protect against authorized and unauthorized usage in networks. An abnormal status occurs in the network is called an Intrusion. An Intrusion Detection System (IDS) is deployed to identify and remove anomalies over network. Intrusion Detection System protects the network by analyzing traffic and finds attack over the network resources. It checks for anomalies in data. Data mining techniques are often used for extracting huge amount of data from database. Data mining techniques like classification, clustering, association, selection of attributes, visualization are majorly used in Anomaly based IDS [1]. In this, classification technique plays a major role and produce accurate results in finding intrusion in loaded Dataset. Data mining analyzes data in its perspectives and machine language to find information from huge data.

The danger exposed by these Internet-connected Things not only have an effect on the safety of IoT systems, however additionally the whole eco-system as well as web-sites, applications, social networks and servers, via controlled good device as automaton networks (botnet). In alternative words, compromising one part and/or communication channels in IoT-based systems will paralyze the half or complete net network. In 2016, the Dyn cyberattack harvested connected devices put in within smart-homes and conscripted them into "botnets" (also brought up as a "zombie army") via a malware referred to as Mirai. Additionally to IoT systems vulnerabilities, attack vectors area unit evolving in terms of quality and variety. Consequently, additional attention ought to be paid to the analysis of these attacks, their detection similarly because the infection prevention and recovery of systems once the attack.

Since security of the active IoT systems is critical, it is important to identify IoT threats and specify existing defense strategies. This survey starts with IoT threats classification to have a better vision for strategic investigations. For that, we propose a binary classification with: i) IoT layers; and ii) encountered challenges while developing the IoT systems. We believe that IoT networks are different from Wireless

Sensors Networks (WSN) and Cyber Physical Systems (CPS) [2] due to heterogeneous composition of layers in terms of protocols, standards and technologies. Furthermore, variety of challenges encountered during the implementation of various use-cases mentioned in [3] have different context compared to WSN networks.

Traditional defense mechanisms for known attacks have varied use and may be efficient in specific situations; however, they may not be completely secure. Despite the availability of traditional security with encryption, authentication, access control or data confidentiality, IoT networks still have been subject to network attacks necessitating a second line of defense [5], [4]. In such situations, the importance of Intrusion Detection Systems (IDSs) for IoT is relevant. One of the popular strategy deployed among IoT systems is IDSs or Network Intrusion Detection Systems (NIDSs) for connected smart Things. Based on the previous gathered data of the attacks it is evident that the extent of these attacks is vast and their impact on those who are targeted is severe.

II. SCOPE OF THE SURVEY

Even if there is the availability of traditional security with encryption, authentication, access control or data

confidentiality, IoT networks still have been subject to network attacks necessitating a second line of defense to the potential attacks. In such situations, the importance of Intrusion Detection Systems (IDSs) for IoT is relevant. One of the popular strategy deployed among IoT systems is IDSs or Network Intrusion Detection Systems (NIDSs) for connected smart Things. NIDSs have been subject to scrutiny to achieve secure traditional computer science systems since the 1980s [6]. Thus, NIDS is a mature scientific field. Unfortunately, traditional NIDS techniques may be less efficient and/or inadequate for IoT systems due to characteristic changes like constrained resources, limited power, heterogeneity and connectivity [7], [8]. Traditional systems have usually master nodes which are powerful in terms of computation resource and storage/memory space. These nodes monitor inbound and outbound flows with no major resource or network bandwidth constraints. However, IoT systems are distributed and composed of a large number of devices whose computing capacity, storage/memory space and battery life are mainly limited in resources. IoT is also limited by its network bandwidth capacity. Moreover, IoT allows interaction between the virtual and physical environment which is unpredictable. Every node has an IP address to guarantee its communication with Internet. This causes trust problems and particular vulnerabilities. In addition to these limitations, IoT is based on heterogeneity in terms of communication protocols and co-existing technologies. The protocols and technologies are either not employed in traditional networks such as IEEE 802.15.4, 6LoWPAN1 and CoAP; or at least not at the same time within a single system [9]. Finally, IoT environment generate tremendous critical data that must be protected. Consequently, it is observed that NIDSs are more challenging and restrictive in IoT networks compared to NIDSs in traditional computing systems. Many IoT NIDS have been developed using attacks rules/signatures or normal behavior specification. Unfortunately, these NIDS have; i) high false positive and/or false negative attack recognition (false alarms); ii) inability to detect unknown/zero day attacks. Hence, researchers explored artificial intelligence (AI) and machine learning (ML) with an emphasis on deep learning (DL) algorithms to improve systems security [10], [11], [12]. In fact, learning techniques have a significant impact in fraud detection, image recognition and text classification. The effectiveness of machine learning has encouraged the researchers to deploy learning algorithms among IDS to improve detection of cyber attacks, anomaly detection and identify abnormal behaviors among the IoTs. Therefore, this paper surveys and evaluates notable machine learning contributions for IoT NIDSs. In the recent past, academia and industry have shifted their focus towards developing ML based NIDSs. They accomplish interesting results; from 86.53% [13] to over 99% [14] in detection accuracy and a reduction in false positive (FP) from about 4% [15] to 0.01% [16]

III. NETWORK INTRUSION DETECTION SYSTEM

A Network Intrusion Detection System (NIDS) is used to keep track of and provide analysis of internet traffic on the subnet.

A NIDS reads all incoming data and looks out for suspicious behavior. The system reacts to such behavior based on the seriousness of the threat [17]..

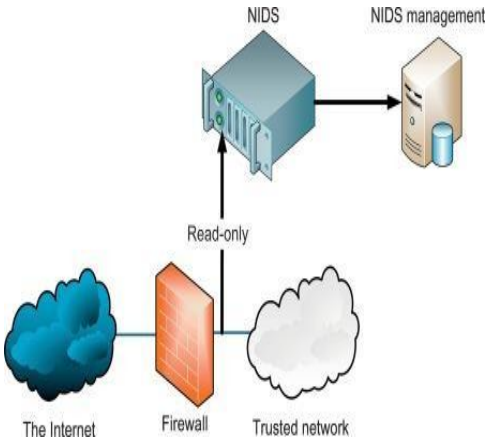


Fig1.Block diagram of NIDS

IV. MACHINE LEARNING

Machine learning is a class of algorithms that allows software applications to become more precise in estimating outcomes without being explicitly programmed [18].The algorithm applied to any data is jointly called a model. A machine learning algorithm learns from experience 'E' concerning some class of tasks 'T' and performance measure 'P' if its Performance at task 'T' enhances with experience 'E' [19]. A generic framework of Machine Learning process is shown in Fig. 2 [20].

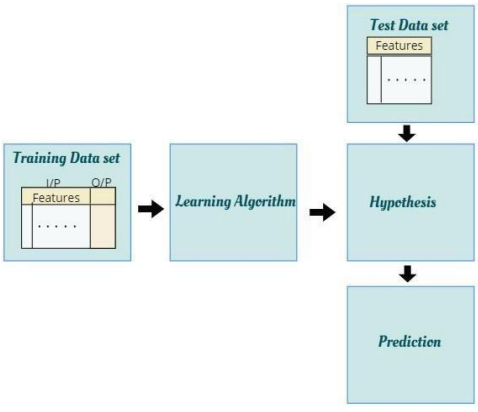


Fig2. Block Diagram of machine learning

A machine learning problem can be classified into:

A. Supervised learning

Supervised Learning is training a model with a dataset which also contains the correct answer for a prediction called as a label.

B. Unsupervised Learning

Unsupervised learning is training a model without using labels.

C. Semi-supervised Learning

Along with the above mentioned two categories, there is yet another field called Semi-Supervised learning, which contains datasets with a few labeled data points in addition to predominantly unlabelled data. Machine Learning is used for Network Intrusion Detection to make the process dynamic as opposed to the current static detection techniques being used.

V. REVIEWED MACHINE LEARNING METHODOLOGIES FOR NETWORK INTRUSION DETECTION

A few Machine Learning methodologies that are currently being used for Network Intrusion Detection are:

A. Support Vector Machine

Support Vector Machine (SVM) is an algorithm that rests upon the notion of decision planes known as decision boundaries. These decision planes assist SVM to classify data into their respective categories. Finding the optimal decision boundary is the main objective of SVM. These boundaries are constructed in the multidimensional space to achieve the optimal result in the case of non-linear data. This is a major advantage of SVM over a simple linear classifier. Margin holds the key for the correctness of classification of a new data point. The margin is the distance between nearest data point, also called as 'Support Vector', and the decision boundary. Mathematical representation of SVM [21] [22] is given as follows [23]:

B. Algorithm proposed (Md Nasimuzzaman Chowdhury et. al.)

The method proposed by Md Nasimuzzaman Chowdhury and Ken Ferens, Mike Ferens [24] begins with an arbitrary selection of 3 features at a time is done in the training samples. This combination is then fed to the SVM. This gives SVM the power to detect any odd activity from internet traffic data. The total number of features (N) from the dataset were identified and arbitrary combination of 'n' features was done (n belongs to N). SVM was applied to these training samples. Total S data samples were selected. SVM parameters such as Gamma, coefficient theta, nu etc. were selected. The T-train dataset is the training dataset; it contains n*S data samples. Testing dataset(T-test) was created using n*M data samples. Ttrain is used to train SVM. Testing the performance of SVM is done by using the T-test data set. Detection accuracy, FPR, FNR and total time taken by system defines the performance. Steps 2 and 3 were repeated until the highest detection accuracy and lowest FPR and FNR was achieved.

C. Min-Max K-means clustering

Another study by Mohsen et. al. have put forth the MinMax K-means clustering [25] for intrusion detection. The suggested algorithm attempts to minimize the maximum internal variance of clusters instead of minimizing the sum of internal variance as that of the K-means algorithm. Every cluster has some weight and higher weights are assigned to the cluster with larger internal variance. Experimentation shows that Min-Max K-means is used to solve the initialization problem of K-means algorithm, as compared to clustering algorithms such as K-means++ [26] and pifs K-means. Min-Max K-means displayed 81% detection rate as compared to75% obtained by the K-means algorithm. False Positive Rate is improved from 14% to 9% for the Min-Max K-means algorithm. It is concluded that the Min-Max K-means clustering has a higher detection rate than the K-means clustering algorithm.

D. Intelligent Intrusion Detection System

The 'Intelligent Intrusion Detection Process' proposed by Jiaqi Li, Zhifeng Zhao and Rongpeng Li [27] consists of two phases. The first phase consists of using a Random Forest algorithm to obtain a subset of features by weighing their importance. The second phase includes a'Hybrid Clustering-Based Adaboost' which acts as a classifier based on the subset of features as the input. The 'Hybrid Clustering-Based Adaboost' is performed in two stages. The first stage consists of using the unsupervised clustering algorithm 'k-means++' [28] to create two preliminary clusters of malicious and benign activities. The 'k-means++' algorithm is preferred over the regular k-means algorithm in order to choose preliminary clustering centers which are as far away from each other as possible. The clusters are further classified into four types of anomaly clusters using AdaBoost [29]. AdaBoost is an ensemble classifier which consists of multiple smaller classifiers trained on the same data. The weight of every data point is the same at the beginning; however, if the example is misclassified in the previous classifier then the weight is increased, and conversely, if the example is correctly classified, then the weight is decreased, the same is followed in successive iterations.

A. Artificial Neural Network

Artificial Neural Networks (ANN) is a machine learning methodology inspired by the human nervous system. A single processing unit of an ANN is known as a perceptron. A perceptron receives weighted input along with a fixed bias value and generates an output. The mathematical

representation of a perceptron is as follows: where *w* is the weight vector, *x* is the input and *b* is bias value. A typical neural network consists of three types of layers - An input layer which gets real values from data points (Network dump files), hidden layers to process inputs and an output layer which provides an actual prediction. The design of the Artificial Neural Network for the Network Intrusion Detection System proposed by Alex Shenfield, David Day and Aladdin Ayesh [30] is as shown in Fig. 3.

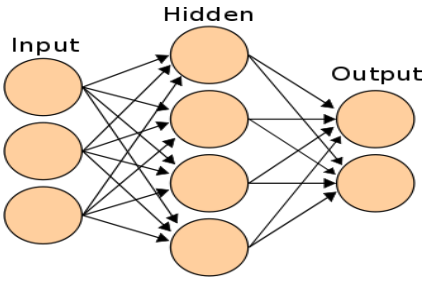


Fig3. proposed Diagram of ANN

VI. IOT & SECURITY

Diversity and heterogeneity makes IoT systems security more crucial. IoT systems differs from traditional systems security due to following reasons:

- IoT systems are constrained in terms of computational capability, memory capacity, battery life and network bandwidth. Hence, it is not possible to deploy existing traditional security solutions which are often resource intensive.
- IoT systems are heavily distributed and heterogeneous systems. Thus, centralized traditional solution may not be suitable. Moreover, the distributed aspect of IoT add more difficulties and constraints in their protection.
- IoT systems are deployed in a physical environment which is unpredictable. Thus, physical attacks have joined the list of traditional security threats.
- IoT systems are connected to Internet since each device has access with its IP address. Hence, there is one more panel of threats related to Internet.
- IoT systems are composed of a large number of constrained objects that generate huge amount of data. So it is easy to flood and attack these small devices on the one side, and the limited bandwidth of the networks on the other side.
- IoT systems cover a large number of heterogeneous protocols and technologies in the same system. Hence, the proposed IoT security solution must take into consideration the large panel of these protocols and technologies in the same proposal.

Consequently, IoT systems threats classification is discussed; then, traditional defense mechanisms employed against such threats are introduced.

VII. SECURITY THRETS ANAYLYSIS AND SECURITY REQUIREMENTS OF IOT DEVICES

IoT devices are exposed to a variety of security threats and vulnerabilities. Hackers who launch attacks using these vulnerabilities exhibit malicious behaviors. Typical security threats and vulnerabilities of IoT devices include unauthorized access, loss or theft, physical destruction, information leakage, illegal data modification, and denial of service attacks.

The openness of IoT platform accelerates inter-working between heterogeneous devices, and the variety of security threats is increasing. In addition, three elements of information security which consist of confidentiality, integrity, and availability are increasing the possibility of infringement. From these threats, we present the security requirements to keep the IoT devices safe.

A. Confidentiality

• [Transmitted message encryption] Messages transmitted between IoT devices are to be transmitted in encrypted format to prevent illegal sniffing or eavesdropping.

• [Malware response] IoT devices should provide the ability to detect and defend against malware infections and external hacker attacks, such as worms and viruses to prevent information leakage.

• [Data encryption] IoT devices should encrypt sensitive data such as private information and cryptographic key, and securely process and store these data to prevent information leakage.

• [Tamper resistance] IoT devices should provide tamper resistance function to ensure the safety and reliability from physical attacks.

• [Device ID management] IoT device should have unique device identification information and safely handled so as not to leak outside or to change illegally.

B. Integrity

• [Data integrity] IoT device should provide data integrity verification function to prevent forgery of data.

• [Platform integrity] IoT devices should provide platform integrity verification function of system level such as firmware and operating system.

• [Secure booting] When power is first introduced to the device, IoT devices should provide secure booting

function to ensure the reliability of the device through authenticity and integrity of the software on the device.

C. *Availability*

• [Logging] IoT device should provide the appropriate log function for the user, the system, the security event.

• [State Information Transmission] IoT device should provide a periodic keep-alive message or device state information transmission function for prevention from physical removal/destruction and abnormal installation attempt.

• [External attack response] IoT device should provide the capability to respond to external attacks, such as denial of service attacks and persistent connection attempt attack.

• [Security monitoring/management] IoT devices should provide security monitoring and management capabilities to respond adequately if lost or stolen, installation and disposal, etc.

• [Security patch] IoT device should provide a safe and secure software update and patch function.

• [Security policy setting] IoT device should provide the capability to securely set an appropriate security policy on the various types of devices.

• [Software safety] IoT devices should ensure software safety, with features such as appropriate module separation or removal, and access restrictions, despite a software failure or malfunction due to malware infections.

D. *Authentication/Authorization*

• [User authentication] IoT device should provide a user authentication function to block the access of unauthorized users.

• [Device authentication] IoT device should provide a device authentication function in order to block the access of illegal device.

• [Password management] IoT device sets the secure and robust password, and should provide the periodic update feature.

• [Mutual authentication] IoT device should provide a mutual authentication between the devices to establish secure, autonomous communication environment.

• [Authority control] IoT device should provide the authority control functions, such as ownership control for preventing information leakage and privacy protection.

• [Access control] IoT device should provide a access control function to block the access of unauthorized users and devices.

• [Identification information verification] IoT device should provide the unique device identification information verification function for preventing device replication, alteration, and appropriation.

VIII. TRADITIONAL DEFENSE MECHANISMS

After detailing and classifying IoT threats, in the following we discuss attack mitigation techniques which protects existing IoT systems and networks. Over time, conventional IT security solutions have covered servers, networks and cloud storage. Most of these solutions can be deployed for security of IoT systems. Defense mechanisms can be separate, or combined depending on the treated threats [31]. In this Section, traditional defense mechanisms that can be used to protect IoT devices that are described.

First, filter packets [32], with firewalls and proxies for example, represents an important defense against IP spoofing attacks (and consequently DDoS attacks). Two types of filtering are possible: i) ingress filtering; and ii) egress filtering. Ingress filtering on incoming packets is about Blocking the data packets from outside the network with a source address inside the network to guard against outside spoofing attacks. However, egress filtering on outgoing packets is about blocking packets within the network with a source address that is not inside to stop an indoor hacker from attacking external machines.

Second, adopt encryption with cryptographic protocols, data storage encryption or virtual private networks (VPNs). Using cryptographic network protocols (i.e., Transport Layer Security (TLS), Secure Shell (SSH), HTTP Secure (HTTPS),etc.) leads to the encryption of data/code/updates before sending and authenticating them. The defense is based on digital signatures/certificates (pair of public and private keys) to ensure, in one hand, that data/code/update was sent by the legitimate device/service and never modified. On the other hand, it guarantees that data/code/updates are encrypted and cannot be read or used by unauthorized individual. Cryptographic network protocols can be used to protect things against IP spoofing, tampering, repudiation, MITM, user privacy compromising attacks and node cloning. Moreover, encrypting data storage helps prevent information disclosure and maintains user privacy. Concerning VPN (Virtual Private Network), it is a secure communication tunnel between two or more devices. It encrypts the communication by creating a virtual private link over the existing insecure network. Encryption is a good solution to preserve confidentiality and privacy. However, IoT networks are vulnerable since the resource limits the devices. Therefore, the use of light cryptographic solutions proposal from Al-Turjman et al. [33] is an

interesting approach. They propose a confidential cloud assisted WSN based framework maintaining confidentiality, integrity and access privileges (CIA). The proposed agile framework ensures integrity of collected sensor data with elliptic curve cryptography.

Third, employ robust password authentication schemes. Moreover, limit data access by assigning the resources with appropriate privileges. The use of One-Time Password (OTP) can be an interesting solution. spoofing, tampering, information disclosure, elevation of privileges and MITM can be avoided by the above mechanisms.For IoT networks, authentication strategies need to be lightweight such as in Al-Turjman et al. solutions. In [70] authors propose a light weight framework to strengthen the safety of IoT networks.They introduce a cloud supported mobile-sink authentication,an elliptic-curve based seamless secure authentication and key agreement (S-SAKA). However in [34], authors propose a “Hash” and “Global Assertion value” based authentication scheme for the evolving 5G technology. Their proposal considers context-sensitive seamless identity provisioning(CSIP) framework for futuristic Industrial Internet of Things(IIoT).

Fourth, audit and log activities on web servers, database servers, and application servers. Due to these traces, outliers can be detected. More specifically, log key events suchas transaction, login/logout, access to filing system or failed resource access attempt(s) can detect anomalous behavior.A good practice to protect these files is to back up them, regularly analyze them for detection of suspicious activity and relocate system log files from their default locations.Further, secure the log files by using restricted ACLs (Access Control List: an inventory of permissions attached to an object) and ecrypt the transaction log. These techniques prevent IoT systems from repudiation and privilege elevation attacks.

Fifth, detect intrusions using IDS (Intrusion Detection System). An IDS [35] is a combination of software and hardware which monitors network or systems to identify malicious activities and gives immediate alerts. They have been adopted [36] since 1970 [37]. IDSs are generally categorized according to i) deployment; and ii) detection methodology.

IDS deployment is categorized as i) HIDSs; and ii) NIDSs. Host-based Intrusion Detection Systems (HIDSs) are installed on a number machine (i.e., a tool or a Thing). They monitor and analyze activities related to system application files and operation system. HIDSs are preferred against insider intrusion deterrence and prevention. Network-based Intrusion Detection Systems capture and analyze packet flow in the network. In other words, they're scanning sniffed packets.NIDSs are strong against external intrusion attacks. Since our interest is towards security of resourcSystem design is the process of designing the elements of a system such as the architecture, modules and components, the different interfaces of those components and the data that goes through that system. The purpose of the System Design process is to provide sufficient detailed data and information about the system and its system elements to enable the implementation consistent with architectural entities as defined in models and views of the system architecture.e constrained IoT systems, the rest of the paper will focus on NIDSs solutions.

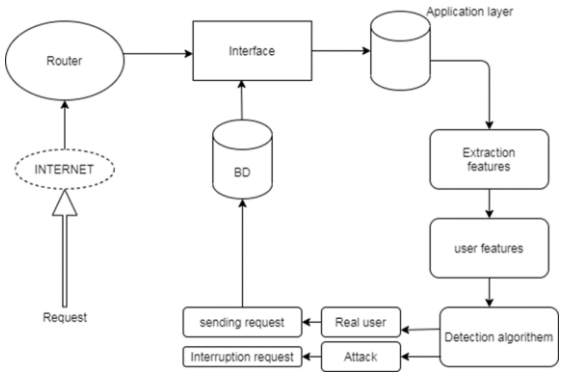


Fig4. Proposed System Architecture

IX. TYPES OF INTRUSION DETECTION:

In the following, we discuss scenario the system is under attack. A good detection system is the one which identifies the compromised situation and minimizes the loss by quickly identifying the attack(s). There are a variety of IDSs. In [38], detection methodologies are classified as i) misuse detection; ii) anomaly detection, iii) specification detection; and iv) hybrid detection.

• Misuse detection or signature detection (knowledge based) is a set of predefined rules (such as bytes sequence in network traffic or known malicious instructions sequence used by a malware) that are loaded and matched with events. When a suspicious event is detected, an alert is triggered. This type of IDS is efficient for known attacks; unfortunately it cannot detect zero-day [39] / unknown / unseen attacks [40] due to lack of signatures. Cyber security solutions prefer signature based detection as it is simple to implement and effective for identifying known attacks (high detection rate with low false alarm rate).

• Anomaly detection (behavior based) compares a normal recorded behavior with current input. Initially, normal network and system behavior are modeled. In case of deviation from normal behavior, the detector considers it an attack. Anomaly is identified with statistical data analysis, mining and algorithmic learning approaches. Anomaly detector is successful in preventing unknown attacks. However, they tend to generate high false positive rate since previously unseen (yet legitimate) behaviors may be categorized anomalous. Another advantage is that the normal profile activities are customized for every system, every application and every network, which makes things

difficult for the attacker. It is difficult to know exactly which activities can be undetected.

• Specification detection has the same logic as anomaly detection. It defines anomaly as deviation from normal behavior. This approach is based on manually developed input specifications to capture legitimate (rather than those previously seen) behavior and its deviations. However, specifications require the user to give input. This method reduces high false alarm rate as compared to anomaly detectors.

• Hybrid detection is a combination of previous methods, especially signature and anomaly based detection. Hybrid detector improves accuracy by reducing false positive events. Most of the existing anomaly detection systems are in reality hybrid one. They start with an anomaly detection, then try to relate it with the correspond signature. Sixth, **prevent intrusions with IPS (Intrusion Prevention System)**. An **IPS** is an IDS which respond to a potential threat by attempting to prevent it from succeeding. An IPS responds immediately and stop malicious traffic to pass before it responds by either dropping sessions, resetting sessions, blocking packets, or proxying traffic. However, an IDS responds after detecting passed attacks. There are many types of IPS [41] mainly in-line detection, layer seven switches, deceptive systems, application firewalls, and hybrid switches. To get more details about IPS types, please refer to Patil et al. paper [42].

The above presented mechanisms can be used to protect IoT systems. Some of them like encryption and authentication are insufficient [43] to protect IoT, therefore; IDS are necessary and are more suitable for this case of systems. They can be considered as the last line of defense when other tools are broken. Another advantage of IDS is that they are diverse and adaptable depending on needs. They can be doted with learning logic such as machine learning and artificial intelligence techniques in addition to other advanced technologies. This subject will be discussed in the next section.

From the different types and categories of IDSs, this survey concentrates on Anomaly and Hybrid Network IDSs (ANIDSs - HNIDSs) for IoT systems. This choice was made due to the power and the ability of anomaly and hybrid IDSs to detect unknown attacks. Moreover, the paper focuses on the network deployment since it offers more freedom in solution development unlike host deployments in IoT which necessitate lowpower consumption and are resource constrained. IoT systems are heterogeneous and too big in term of number of devices. Therefore, having a single/multiple system(s) monitoring the entire network rather than analyzing each host separately (i.e.; the approach of HIDS is per-device security) is more suitable for the case of IoT networks security. After all, IoT is by definition about the inter-connection of heterogeneous Things (devices).

X. WORKING AND IMPLEMENTATION:

NIDSs analyze network traffic to detect malicious behaviors. To build a NIDS, these are the needed basic steps :

1) Collect the traffic data from the network.

2) Analyze the collected data.

3) Identify relevant security events.

4) Detect and report malicious events.

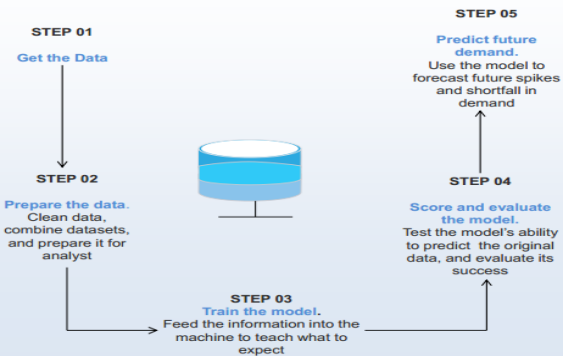
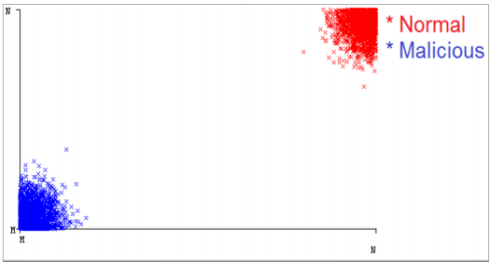


Fig5. Dataflow diagram of NIDS

SVM algorithm:

- The purpose of machine learning is to distinguish between **normal traffic** and **malicious traffic**
- Support Vector Machine (SVM) is technique based on supervised machine learning algorithm.
- mainly used to **classified data** into different classes



XI. FUTURE ENHANCEMENT

With the explosion of IoT, two new paradigms come out: **edge computing and fog computing**. Both of them tend to push intelligence and processing logic employment down near to data sources (which means as close as possible to sensors and actuators) to reduce the network bandwidth needed to communicate data from the perception layer to data-centers where analytics are usually processed. The main difference between edge and

fog architecture lies in the place where the intelligent processing and the computing power are located. Edge computing pushes them to the extremes of the network such as edge gateways and devices (e.g. Programmable Automation Controllers PACs). However, fog computing tends to place them in the local area network level of the network architecture which means in hubs, routers or gateways (fog nodes). These two concepts should be deeply explored and exploited for future IoT IDS architecture. They enable the intrusion detection process to be distributed. Consequently, this strategy should enable intrusion detection with less resource needs which is suitable for IoT. **Big Data** [44] is a solution to remedy problems related to the big volume of network traffic generated by IoT networks. So as future works in IDS architecture deployment, edge and fog computing as well as Big Data methods should be deeply explored for IoT NIDS with paying more attention on protecting IDSs themselves in case of IoT system fall. Furthermore, IoT NIDS needs a **real-world IoT-dedicated dataset**. A common real-world dedicated dataset would help with a real, efficient comparison between the different researches. A dataset benchmark enables training, validating and evaluating studies with different ML algorithms. Besides, according to Sommer and Paxson [45], IDS based on learning techniques suffer from "a semantic gap between results and their operational interpretation". Unfortunately, IDS based on learning techniques are usually evaluated with rates such as accuracy, false positive and false negative. We believe that presenting just these metrics is not sufficient. Researchers should interpret the results and understand semantics of then features choice and the detection process. Semantic would also help to differentiate between abnormal and malicious behaviors. Therefore, **semantic relation between detection and learning process** seems to be an interesting track to explore. Moreover, **features choices** as well as **features reconstruction and features dimension reduction** can be more inspected for IoT NIDS based on learning techniques. Such techniques can help overcoming IoT resource constraints challenges. **Deep learning techniques** used alone or combined should be also more experienced since algorithms like auto-encoders are efficient in features reconstruction and dimension reduction.

In addition, techniques like **software accelerator**, for low powering the learning algorithms on tiny devices, could be experienced in IoT security environment like in [46]. Nicholas D. Lane en al. designed and implemented DeepX, a software accelerator for deep learning execution DeepX that lowers significantly the device resources (viz. memory, computation, energy) required by deep learning. Security researchers can get inspired from works such as Ravi et al. in [47] where they presented an optimization approach to enable the use of realtime deep learning in low-power devices. The authors used a spectrogram representation of the inertial input data to provide invariance against changes in sensor placement, amplitude, or sampling rate, thus allowing a more compact method design. The same authors proposed in [48] a combination of shallow learned features from a deep learning approach to enable accurate and real-time activity classification. Such a proposal overcomes some limitations for deep learning when on-node computation is needed. Last but not least, for the future, more efforts need to be made to detect **unknown and zero-day attacks** in IoT networks and develop IDSs that can automatically update the list of the considered attacks when new ones appear. IoT NIDS need to be experienced with ML algorithms and big data [49], [50] strategies to update their training model in **real-time**, in a **streaming detection**. For example, **Incremental ML** field in the intrusion detection should be experienced. Incremental learning is about retraining the model on both previously seen and unseen data to construct new models. It aims to ensure continuity in the learning process through regular model update based only on the new available batch of data. This idea joins the approach of making IDSs more intelligent and human-independent in decision making. Finally, IoT is being deployed increasingly in industrial systems, military operations, health-care environment and many other sensitive areas that are cognitive based human-centric IoT. Sensitive data and private information are exchanged between the travelling objects in a context that puts people's lives at risk on the one side and where human behaviors affect the IoT systems in the other side. Hence, more security attention needs to be paid to these IoT human-based systems.

X. CONCLUSION

Connected Things (IoTs) have become pervasive for every individual. In fact, the IoT benefits has human life evolving with the Things. IoT is in smart cities (e.g. smart parking), smart environment (e.g. for air pollution), in smart metering (e.g. smart grid), in industrial control (e.g. for vehicle auto-diagnosis), etc. They are in every domain even in critical ones like military, health care and buildings security. Unfortunately, industries are focusing on innovating and developing more connected products without verifying too much their quality and security. At this stage, we remark that IoT is a double edged weapon. This army of connected devices can be hacked and used against humanity. One compromised node can affect the whole IoT network. A malicious user becomes able to break down home automation systems and so steals them or can remotely control vehicles to hit innocent people in roads. One of the powerful mechanisms to ensure IoT network security are NIDSs. They help detecting intrusions in systems. To enhance their efficiency, they are becoming provided with learning techniques. To the best of our knowledge, our survey is the first proposal with comprehensive discussion of learning based NIDSs for IoT systems. In this paper the field of IoT security has been introduced with a comparison between previous surveys. Moreover, IoT threats and detection techniques over traditional defense mechanisms have been classified. Then, a comprehensive evaluation of NIDS implementation tools

have been presented; starting with free network datasets, to free and open-source network sniffers, to open-source NIDS that can be used by researchers and industrious to implement and evaluate their own sophisticated NIDS solution. Furthermore, an overview about NIDS in IoT systems has been given with a focus on their architecture, deployments, detection methodologies and treated threats. The pros and cons of each proposal is thoroughly evaluated. Last but not least, we continued with learning NIDSs for IoT eco-system where, learning terminologies have been introduced and the working mechanism of IoT learning NIDS has been detailed. Each work has been summarized separately; then, adopted strategies have been compared to come up with strengths and tactics ideal for ML and non-ML NIDS. The State-of-art shows interesting results; up to 99% detection accuracy and 0.01% false positive. Finally, top IoT NIDS proposals have been compared with a focus on ML algorithms and future research directions have been detailed.

In the coming time, IoT based solutions will explode. We believe that one of the most important needs to deal with is the validation strategy improvement; more specifically, the development of a public benchmark dataset for network exchanges of IoT systems. It should include different IoT protocols with the different IoT threats. This dataset would enable a clear, practical and convenient comparison of the different developed NIDS. Furthermore, it is also important to concentrate on developing IoT NIDS that detect known and unknown attacks without being protocol-dependent. To conclude, a combination of edge and fog computing approaches could be more and more explored for IoT NIDS architectures. These approaches enable IoT intrusion detection with less resource consumption, thus, with respect to IoT challenges.

References

[1] G.Eason, B Noble, and I.N.Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955.

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.

[3] I.S. Jacobs and C.P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.

[4] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[5] J.P.Anderson, \Computer security threat monitoring and surveillance, & quot; James P. Anderson Company, Technical Report, 1980.

[6] L. W. S. Raza and T. Voigt, \Svelte: Real-time intrusion detection in the internet of things ,& quot; Ad Hoc Networks, vol. 11, no. 8, pp. 2661{2674, Nov. 2013.

[7] B. M. N. K. O. A. T. I. Z. M.Fadlullah, F. Tang and K. Mizutani, \State-of-the-art deep learning: Evolving machine intelligence toward tomorrow’s intelligent network traffic con-Trol systems, & quot; IEEE Communications Surveys Tutorials, vol. 19, no. 4, pp. 2432{2455 Nov. 2017.

[8] R.Sommer and V. Paxson, \Outside the closed world: On using machine learning for network intrusion detection," in 2010 IEEE Symposium on Security and Privacy, vol. 19, no. 4, pp. 305{{316, May 2010.

[9] J. C. R. Z. Y. Z. Liu, C.; Yang, \Research on immunity-based intrusion detection technology For the internet of things,& quot; In Proceedings of the 2011 Seventh International Conference On Natural, Shanghai, China, 2011.

[10] J. L. J. Roman, R.; Zhou, \On the features and challenges of security and privacy in Distributed internet of things." Comput. Netw., vol. 57, pp. 2266{{2279, 2013.

[11] P. K. Zegzhda, \Host-based intrusion detection system: Model and design features,& quot; In Proceedings of the International Conference on Mathematical Methods, Models, and Architectures for Computer Network Security, St. Peters burg, Russia, vol. 57, pp.340{345, 2007.

[12] J. Schmidhuber, \Deep learning in neural networks: An overview." INeural Netw., vol. 61, PP. 85{117, 2015.

[13] A. S. B. A. S. K. K. J. M. B. Saleh, A.J.; Karim, \F.d. an intelligent spam detection model based on Artificial immune system." Information, vol. 10, p. 209, 2019.

[14] Zilberstein, \S. book review: \multiagent systems: A modern approach to distributed Artificial intelligence, " Gerhard Weiss. Int. J. Comput. Intell. Appl., vol. 1, pp. 331{334, 2001.

[15] E. Bertino and N. Islam, \Botnets and internet of things security," Gerhard Weiss. Int. J.Comput. Intell. Appl., vol. 50, no. 2, pp. 76|-79, 2017.

[16] .G.-B. Huang, H. Zhou, X. Ding and R. Zhang, "Extreme learning machine for regression and multiclass classification", IEEE Transactions on Systems Man and Cybernetics, vol. 42, no. 2, pp. 513-529, 2012.

[17] Mehrnaz Mazini, Babak Shirazi and Iraj Mahdavi, "Anomaly network-based intrusion detection system using a reliable hybrid artificial bee colony and AdaBoost algorithms", Journal of King Saud University-Computer and Information Sciences, 2018.

[18] Hudan Studiawan, Christian Payne and Ferdous Sohel, "Graph Clustering and Anomaly Detection of Access Control log for Forensic Purposes", ELSEVIER Digital Investigation, vol. 21, pp. 76-87, June 2017.

[19] R.M. Elbasiony, E.A. Sallam, T.E. Eltobely and M.M. Fahmy, "A hybrid network intrusion detection framework based on random forests and weighted k-means", Ain Shams Eng. J., vol. 4, no. 4, pp. 753-762, 2013.

[20] Karen A. Garcia, Raul Monroy, Luis A. Trejo, Carlos Mex-Perera and Eduardo Aguirre, "Analyzing Log Files for Postmortem Intrusion Detection", IEEE Transactions on Systems Man and Cybernetics part C(Application and Reviews), vol. 42, no. 6, pp. 1690-1704, 2012.

[21] Vajiheh Hajisalem and Shahram Babaie, "A hybrid intrusion detection system based on ABC-AFS algorithm for misuse and anomaly detection", ELSEVIER Department of Computer Engineering, vol. 136, pp. 37-50, May 2018.

[22] Buse Gul Atli, Yoan Miche, Aapo Kalliola, Ian Oliver, Silke Holtmanns and Amaury Lendasse, "Anomaly-Based Intrusion Detection Using Extreme Learning Machine and Aggregation of Network Traffic Statistics in Probability Space", SPRINGER Cognitive Computation, pp. 1-16, June 2018

[23] Wathiq Laftah Al-Yaseen, Zulaiha Ali Othman and Mohd Zakree Ahmad Nazri, "Multi-Level Hybrid Support Vector Machine and Extreme Learning Machine Based on Modified K-means for Intrusion Detection System", ELSEVIER Expert System with Applications, vol. 66, pp. 296-303, Jan 2017.

[24] Setareh Roshan, Yoan Miche, Anton Akusok and Amaury Lendasse, "Adaptive and Online Network Intrusion Detection System using Clustering and Extreme Learning Machines", ELSEVIER Journal of the Franklin Institute, vol. 355, no. 4, pp. 1752-1779, March 2018.

[25] H. Wang, J. Gu and S. Wang, "An effective intrusion detection framework based on SVM with feature augmentation", Knowl.-Based Syst., vol. 136, pp. 130-139, Nov. 2017.

[26] Pinjia He, Jieming Zhu, Shilin He, Jian Li and Michael R. Lyu, "A Feature Reduced Intrusion Detection System Using ANN Classifier", ELSEVIER Expert Systems with Applications, vol. 88, pp. 249-247, December 2017.

[27] Iftikhar Ahmad, Mohammad Basheri, Muhammad Javed Iqbal and Aneel Raheem, "Performance Comparison of Support Vector Machine Random Forest and Extreme Learning Machine for Intrusion Detection", IEEE ACCESS Survivability Strategies for Emerging Wireless Networks, vol. 6, pp. 33789-33795, May 2018.

[28] .J. Ryan, M. Lin and R. Miikkulainen, "Intrusion Detection with Neural Networks", AI Approaches to Fraud Detection and Risk Management: 1997 AAAI Workshop, pp. 72-79, 1997.

[29] D. Joo, "The neural network models for IDS based on the asymmetric costs of false negative errors and false positive errors", Expert Systems with Applications, vol. 25, no. 1, pp. 69-75, October 2003.

[30] H. Debar, M. Becker and D. Siboni, "A neural network component for an intrusion detection system", Proceedings of 1992 IEEE Computer Society Symposium on Research in Security and Privacy, pp. 240-250, 1992.

[31] A Alfantookh, "DoS Attacks Intelligent Detection using Neural Networks", Journal of King Saud University Computer and Information Sciences, vol. 18, 2005.

[32] M. Moradi and M. Zulkernine, "A Neural Network Based System for Intrusion Detection and Classification of Attacks", Proc. of the 2004 IEEE International Conference on Advances in Intelligent Systems - Theory and Applications, vol. 148, pp. 1-6, November 2004.

[33] Vesely Arnost, "Neural networks in intrusion detection systems", Zemedelska ekonomika, vol. 50, pp. 35-39, 2004.

[34] H Debar, M Becker and Ulis Les, "A Neural Network Component for an Intrusion Detection System", Proceedings IEEE Computer Society Symposium, 1992.

[35] V Theuns and H Ray, "Intrusion Detection and Approahes", Journal of Computer Communications, vol. 25, pp. 1356-1365, 2002.

[36] J Cannady, "Artificial Neural Networks for misuse detection", National Information Systems Security Conference, pp. 368-81, 1998.

[37] P. A. Porras, "STAT: A State Transition Analysis Tools for Intrusion Detection", pp. 15-20, 1992.

[38] .Dorothy E. Denning, "An Intrusion-Detection Model", IEEE Transactions on Software Engineering, vol. SE-13, no. 2, pp. 222-232, February 1987.

[39] Sang-Jun Han and Sung-Bae Cho, "Evolutionary Neural Networks for Anomaly Detection Based on the

Behavior of a Program", IEEE Transactions on Systems, vol. 36, no. 3, June 2006.

[40] S. Mukkamala, "Intrusion detection using an ensemble of intelligent paradigms", Journal of Network and Computer Applications YJNCA, pp. 1-15, 2004.

[41] Khattab M. Ali, Venus W. Samawi and Mamoun Suleiman AL Rababaa, "The Affect of Fuzzification on Neural Networks Intrusion Detection System", Proc. ICIEA, 2009.

[42] CHEN Hong, WAN Guangxue and XIAO Zhenjiu, "Intrusion detection method of deep belief network model based on optimization of data processing", Journal of Computer Applications, vol. 37, no. 6, pp. 1636-1643, 2017.

[43] 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", Computer Networks, vol. 151, pp. 147-157, 2019.

[44] K Yang, J Ren, Y Zhu and W Zhang, "Active Learning for Wireless IoT Intrusion Detection", IEEE Wireless Communications, vol. 25, no. 6, pp. 19-25, 2018.

[45] Nour Moustafa and Jill Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)", 2015.

[46] Amrita Ghosal and Subir Halder, "A survey on energy efficient intrusion detection in wireless sensor networks", Journal of Ambient Intelligence and Smart Environments, vol. 9, pp. 239-261, 2017.

[47] V Jaiganesh, P Sumathi and S. Mangayarkarasi, "An analysis of intrusion detection system using back propagation neural network", 2013 International Conference on Information Communication and Embedded Systems ICICES 2013, pp. 232-236, 2013.

[48] Alex Shenfield, David Day and Aladdin Ayesh, "Intelligent intrusion detection systems using artificial neural networks", ICT Express, vol. 4, 2018.

[49] I. Obeidat, N. Hamadneh, M. Alkasassbeh, M. Almseidin and AlZubi, Intensive Pre-Processing of KDD Cup 99 for Network Intrusion Classification Using Machine Learning Techniques, 2019.

[50] Zuech et al., Intrusion detection and Big Heterogeneous Data: a Survey, 2015.

[51] .L Dhanabal and S P Shantharajah, "A Study on NSL-KDD Dataset for Intrusion Detection System Based on Classification Algorithms", International Journal of Advanced Research in Computer and Communication Engineering, 2015.

[52] Javaid et al., A Deep Learning Approach for Network Intrusion Detection, 2016.

[53] Steven R. Snapp, James Brentano, Gihan V. Dias, Terrance L. Goan, L. Todd Heberlein, Che-Lin Ho, et al., DIDS (Distributed Intrusion Detection System) – – Motivation Architecture and An Early Prototype, 2017.

[54] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection", IEEE Communications Surveys & Tutorials, vol. 18, no. 2, pp. 1153-1176, 2016.

[55] Mohsen Eslamnezhad and A. Varjani, "Intrusion detection based on MinMax K-means clustering", 2014 7th International Symposium on Telecommunications IST 2014, pp. 804-808, 2014.

[56] Terzi et al., "Big data analytics for network anomaly detection from netflow data", International Conference on Computer Science and Engineering (UBMK), 2017.

[57] Yogita B. Bhavsar and Kalyani C. Waghmare, "Intrusion detection system using data mining technique: Support vector machine", International Journal of Emerging Technology and Advanced Engineering, vol. 3, no. 3, pp. 581-586, 2013.

[58] S. Revathi and A. Malathi, "A Detailed Analysis on NSL-KDD Dataset Using Various Machine Learning Techniques for Intrusion Detection" in International Journal of Engineering Research and Technology, ESRSA Publications, vol. 2, no. 12, December 2013.

[59] Poonam Dabas and Rashmi Chaudhary, "Survey of Network Intrusion Detection Using K-Mean Algorithm", International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, no. 3, pp. 507-511, 2013.

[60] M. Crosbie and G. Spafford. Defending a computer system using autonomous agents. Technical Report 95-1022, Dept. of Computer Sciences, Purdue University, Mar 1996.

[61] G. Helmer, J. Wong, V. Honavar, and L. Miller. Intelligent agents for intrusion detection. In IEEE Information Technology Conference, pages 121-124, September 1998.

[62] J. S. Balasubramaniyan, J. O. Garcia-Fernandez, D. Isacoff, E. Spafford, and D. Zamboni. An architecture for intrusion detection using autonomous agents. Technical Report 98/05, Purdue University, 1998.

[63] S. Cheung, R. Crawford, M. Dilger, J. Frank, J. Hoagland, K. Levitt, J. Rowe, S. Staniford-Chen, R. Yip, and D. Zerkle. The design of grids: A graph-based intrusion

detection system. Technical Report CSE-99-102, U.C. Davis Computer Science Department, January 1999.

[64] Judith Hochberg, Kathleen Jackson, Cathy Stallings, J. F. McClary, David DuBois, and Josephine Ford. Nadir: An automated system for detecting network intrusion and misuse. Computers & Security, 12(3):235-248, 1993.

[65] F. Chen et al., "Data mining for the Internet of Things: Literature review and challenges", Int. J. Distrib. Sensor Netw., vol. 11, no. 8, Aug. 2015.

[66] F. Hosseinpour, P. V. Amoli, J. Plosila, T. Hämäläinen and H. Tenhunen, "An intrusion detection system for fog computing and IoT based logistic systems using a smart data approach", Int. J. Digit. Content Technol. Appl., vol. 10, no. 5, pp. 34-46, Dec. 2016.

[67] S. Rathore and J. H. Park, "Semi-supervised learning based distributed attack detection framework for IoT", Appl. Soft Comput., vol. 72, pp. 79-89, Nov. 2018.

[68] Z. M. Fadlullah et al., "State-of-the-art deep learning: Evolving machine intelligence toward tomorrow's intelligent network traffic control systems", IEEE Commun. Surveys Tuts., vol. 19, no. 4, pp. 2432-2455, 4th Quart. 2017.

[69] P. Kasinathan, C. Pastrone, M. A. Spirito and M. Vinkovits, "Denial-of-service detection in 6LoWPAN based Internet of Things", Proc. IEEE 9th Int. Conf. Wireless Mobile Comput. Netw. Commun., pp. 600-607, 2013.

[70] S. Raza, L. Wallgren and T. Voigt, "SVELTE: Real-time intrusion detection in the Internet of Things", Ad Hoc Netw., vol. 11, no. 8, pp. 2661-2674, Nov. 2013.

[71] E. Benkhelifa, T. Welsh and W. Hamouda, "A critical review of practices and challenges in intrusion detection systems for IoT: Towards universal and resilient systems", IEEE Commun. Surveys Tuts., vol. 20, no. 4, pp. 3496-3509, 4th Quart. 2018.

[72] R. Mitchell and I.-R. Chen, "A survey of intrusion detection techniques for cyber-physical systems", ACM Comput. Surveys, vol. 46, no. 4, pp. 1-55, Mar. 2014.

[73] O. Vermesan and P. Friess, Internet of Things Applications—From Research and Innovation to Market Deployment, Aalborg, Denmark:River, Jun. 2014.

[74] D. Singh, G. Tripathi and A. J. Jara, "A survey of Internet-of-Things: Future vision architecture challenges and services", Proc. IEEE World Forum Internet Things (WF-IoT), pp. 287-292, Mar. 2014.