Performance Enhancement for Intrusion Detection Systems

Abdullah Baz, Samah Abuayeid, Hosam Alhakami, Tahani Alsubait
aobaz01@uqu.edu.sa, s44180807@st.uqu.edu.sa, hhhakam@uqu.edu.sa, tmsubait@uqu.edu.sa
College of Computer and Information Systems, Umm Al-Qura University, Makkah, Saudi Arabia

Summary

Due to the rapid improvements in the networking and communication area, the internet becomes the primary connection and influence in people's life. Besides, many organizations store, manipulate, and transfer their secure data via the internet. However, this increases the system's vulnerabilities making it prone to different kinds of security threats. An efficient information system must achieve the goal of a security triangle by protecting system confidentiality, integrity, and availability. A particular practice to meet the security requirements in the modern organization's information systems is to establish an intrusion detection system (IDS). IDS is considered an effective network technology to monitor and detect security attacks. Recently, IDS has addressed many problems related to detection accuracy, such as false-positive and false-negative alarm. In this paper, we introduce the primary concerns and challenges encountered continuously by IDS with a review of the current studies and research in the IDS area that solve and enhance these issues. Moreover, we propose a unified framework that utilizes a combination of IDS and machine learning techniques to address any potential impact on IDS performance.

Key words:
Intrusion Detection System; Information security; Network; Security attacks; Malware.

1. Introduction

The rise of dependency on the internet by the society and the modern organizations and the technology evolution cause an increase in the rate of cybersecurity attacks and threats. Many countries have been affected by many kinds of attacks, which cause significant impacts on these countries' businesses, e.g., the cybersecurity attack that affected the Ukrainian power gird [1] in 2015 and continuously repeated in USA and Russia in the next years. This significant attack affected not only the main power gird but also substations have been impacted too. Additionally, the electricity had stopped in many areas, and the residents were left without power for hours until the problem was solved. Moreover, in 2017, the USA and Australia statistics have addressed a considerable amount of Zero-day-attack [2]. As a consequence of all this technology and advancement, the security criminals and hackers have improved their abilities and skills in threatening the systems. A significant example in this situation is bank credentials and account information stealing [2]. Security attacks can be classified into two categories depending on the attack behaviors. They are 'Active attack' and 'Passive attack.' Additionally, Fig 1 shows the general categories of security attacks classification as some security studies proposed [3].

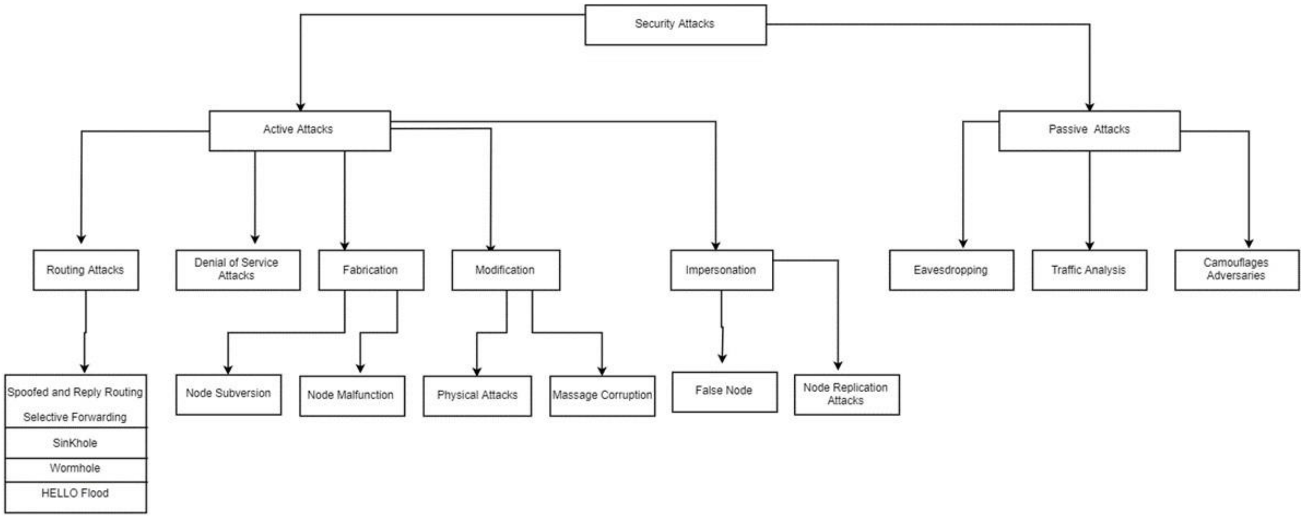

Fig. 1 Security Attacks Classifications.

Manuscript received April 5, 2020
Manuscript revised April 20, 2020

However, the Center of Internet Security (CIS) statistics had presented the most significant security attacks in the years from 2012 to 2019, shown in Table 1.

Table 1: A comparison between machine learning papers results.

Algorithm	*Paper*	*Dataset*	*Precision*	*Recall*	*F-Measure*
Naive-Bayes9	[16]-[12]	KDDCUP99—NSL-KDD	86.1 — 0.964	90.6 — 0.963	87.2 — 0.963
J48	[16]-[12]	KDDCUP99—NSL-KDD	96.2 — 0.992	86.1 — 0.992	90.3 — 0.992
Random-Forest	[16]-[12]	KDDCUP99—NSL-KDD	97.7 — 0.997	87.5 — 0.997	91.8 — 0.997
Random-Tree	[16]	KDDCUP99	97.4	88.8	92.3
multi-algorithms	[24]	KDDCUP99	99.83	1.00	-

On the other hand, the papers [4] [5] [6] has discussed the most threaten malware's, which are EmoTet, DorkBot, WannaCry [7] and Gh0st. Nowadays, the internet of things (IoT) introduced new challenges [8] in the world of information security due to the heterogeneous systems and multiple numbers of devices connected by a network, such as smart cities and smart homes. However, this kind of system may need special IDS that shall be convenient with its heterogeneous nature. On the other hand, the demand for developing and establishing a robust, accurate, and responsiveness IDS has increased. However, an IDS may be classified based on the detection approach or based on the implementation environment. Based on the detection approach, there are two different types of IDS, which are the Signature intrusion detection system (SIDS) and the Anomaly intrusion detection system (AIDS). The two types of IDS have been discussed in [9] [10] in which SIDS is described as depending on a public database that contains information about different attacks and malware signatures. Therefore, SIDS cannot explore unfamiliar unauthorized access on the system unless the attack signatures have been stored in the database. On the other hand, AIDS [9] learns the attacks behavior when it starts to threaten the system and uses this information in attack detection. Secondly, IDS can be classified based on the implementation environment [11], in this case IDS can be classified as [12] Host-Intrusion Detection System (HIDS) [12], which is usually installed as an application on the host device and Network Intrusion Detection System (NIDS) [12], which is established as a hardware in the organization's network system. In this paper, we introduce the primary defects and challenges encountered continuously by IDS with the current studies and research in the IDS area that solve and enhance these defects.

This paper is structured as follows: Section 2 introduces a detailed description of the IDS. Section 3 discusses the system's heterogeneity. Section 4, presents emergent studies in enhancing the performance of IDS. The last section concludes the paper.

2. Intrusion Detection Systems

In this section, we describe in brief the typical information about IDS. IDSs observe network and system activities to find any malicious activities or access policy breaking, which could be difficult to discover by a traditional firewall application, and then reports are produced for the system administrator. Moreover, IDS can be either software or hardware, and it also could be network-based or host-based. Additionally, IDS can be categorized depending on its implementation and detection methodologies into two categories SIDS and AIDS [13] [11]. Moreover, SIDS is usually known as the knowledge-based detection system or misuse detection system; these nomenclatures came from the method of implementing SIDS. The main idea behind SIDS is to examine each transferred network packet and compare the packet content with the database of intrusion signatures when a match appears; then, the SIDS warns the system administrator that there is a possibility for security intrusion as illustrated in Fig 2.

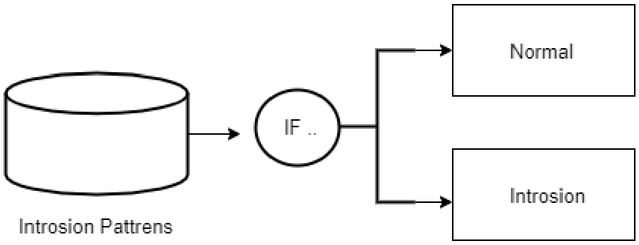

Fig. 2 Signature Intrusion Detection System [8].

Additionally, the dependability of the SIDS on the knowledge in a database is known as the Knowledge- Based Detection system. On the other hand, the accuracy of detection for this kind of detection system will be minimized to a low average if the current attack or malware does not have a well-known signature, such as, zero-day attack and a new attack, which could be considered as SIDS drawback. On the other hand, AIDS's main attribute shown in Fig 3 is the definition of the network behavior; The detection system particularly implemented with an associative build-in database that stores the accepted behaviors.

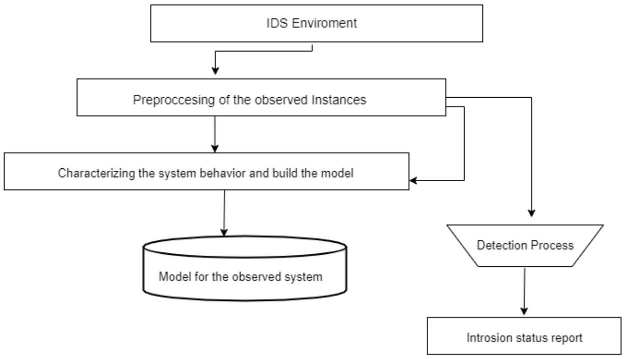
IDS Enviroment
Preproccesing of the observed Instances
Characterizing the system behavior and build the model
Model for the observed system
Detection Process
Introsion status report

Fig. 3 Anomaly Intrusion Detection System [14].

If the network behavior matched with the predefined AIDS behaviors, then the behavior is accepted by the AIDS; otherwise, the AIDS will trigger an alarm to the system administrator for the possibility of intrusion or security attack. However, the critical phase in specifying the network behavior is the capability of AIDS to deal with different types of protocols at all system layers because the system will understand the aim of each protocol. Although this step is costly and it will never be counted corresponding to its benefit in reducing the numbers of False- Positive alarms. Moreover, the main advantage of AIDS is its ability to detect and solve the problem of Novel Attack [14] that is particularly one of the top drawbacks of SIDS. On the other hand, intrusion on the information system is an active (modification-destroying-forwarding) or passive (obtain) effort by the attackers on the system resources, and it could happen by sending a network packet that associated with a kind of malicious code, and it can be on the server or at client side. Detecting an intrusion or unauthorized access in a network system needs technology with a high ability to explore, analyze and classify any malicious attack depending on these malicious signatures and to notify the system as quickly as possible. On the other hand, the high rate in increasing numbers and types of security attacks and malware, decreasing the accuracy, speed, and reliability presented the network intrusion detection success triangle factors [15]. Additionally, many kinds of research had mentioned the problem of high false-positive and high false-negative [12]. The false-positive alarm is the number of intrusion connections alarm, which incorrectly classified as normal access [16]. The second type is a false-negative alarm, which is the number of normal connections that are incorrectly classified as intrusion access by the intrusion detection system [16]. On the other hand, the true-positive alarm is the number of connections that correctly classified as an intrusion [16]. These defects reduce the dependability and security of the IDS. Additionally, choosing the right type of intrusion, the dataset will involve a huge effort when implementing or examining an IDS approach. However, there are two categories [8] of intrusion dataset for commercial products, which usually used in the analysis of network packet, but it's not available for public use; an example of intrusion dataset as in [17] [8] KDD-CUP99, CAIDA, NSL-KDD, and ISCX. However, Table 2 shows a comparison between these datasets. On the other hand, to prevent the attack from detection by IDS, some modifications and techniques shall be used, which known as evasions attacks [2]. However, in practical evasions, attacks are not easy for exploitation, as it appears theoretically; thus, it requires more knowledge in the network to be attacked. However, the well-known evasions attacks are discussed in [2] [13], which are Obfuscation, fragmentation, encryption, and denial of Service.

Table 2: Comparing different types of Intrusion dataset

Dataset	*Creation year*	*Count*	*Format*	*Network Type*	*Duration*	*Normal Traffic*	*Attack*
KDD-CUP99	1998	5M points	–	small network	–	Yes	Yes
CAIDA	2012	32M flows	uni.flow	small network	28 days	Yes	Yes
NSL-KDD	1998	150M points	other	small network	–	Yes	Yes
ISCX	2012	2M flows	bi.flow	small network	7 days	Yes	Yes

3. Systems Heterogeneity

Information systems heterogeneity is one of the significant challenges faced by researchers in all areas of information security. However, systems heterogeneity [18], where different users could work on different types of distributed information systems that could use and produce various heterogeneity data resources. However, managing and protecting these complex systems need out-of-the-box ideas and consuming more effort and time because of the critical data that is manipulating in these systems. On the other hand, the rate of complexity and security vulnerability in a stand-alone system or interactive transaction-based systems can not be compared with the challenging complexity and weaknesses in the system of systems (SOS) or data collection system (DCS). However, the concept of SOS refers to a set of systems or system components and services that interact to accomplish a unique capability that none of the foundation systems can achieve individually. On the other hand, DCS is the system that collects data from their technical environment by using a set of sensors and then send the collected data to other systems for processing and manipulating. Additionally, the systems could have to react with sensors and often are deploying in a hostile environment, such as inside an engine or a distant location. Also, the system could use some of the ideas of Big data analysis and cloud computing to transfer out statistical analysis data besides fetching relationships and intersections between the collected data.

4. Emergent Studies in IDS

This section will present a comparative review for current studies that use different technologies and algorithms to enhance the ability of IDS concerning system heterogeneity ideas and information security triangle goals.

4.1 IDS based on Fuzzy Logic

Fuzzy logic is an approach for data processing depending on the degrees of truth instead of the Boolean true or false values. In particular, fuzzy logic is a powerful tool for executing under uncertainty [19], which is one of the features in analyzing. The security intrusions fuzzy logic has become a suitable area for many studies in enhancing the IDS detection accuracy.

Recently, multiple studies have developed an IDS framework proper with MANET network, which is a collection of wireless nodes, where each node could change its geographical position consequently and serve as a router, which in charge of forwarding network packets. Since the MANET network depends on wireless communication, a few sorts of security attacks can threaten it, [20] such as black hole attack, warm hole attack, and gray hole attack. One of the most surprising studies [21] has established a new MANET IDS by using the advantage of the biometric system. Another researcher [20] has proposed an IDS that uses a node blocking approach with fuzzy logic to establish a secure communication link between different nodes in the network and protect the MANET from black hole attack and gray hole attack. The proposed system consisted of three parts:

- **Attack categorization**: The attacks have classified into two-classes: internal attacks and external attacks, where the internal attacks performed by adjustment of the nodes that belong to the same network. On the other hand, external attacks are performed by the outside source, which replays false data forwarding or old data routing to increase the network overhead.
- **Fuzzy implementation**: As a result of the three attacks proposed in [20], three fuzzy measurements calculated in the study and then threshold values estimated from the measurements values.
- **Fuzzy estimation**: The value of membership matrix had been calculated depending on the result of previous steps, and the results sent to the additional part which use these values to trace the nodes, so if the node considered as a malicious node, then packet path change.

The experiment in [20] has been implemented using a software called (network simulator-2), and in the end, the system has a powerful ability to detect not only the attack but also the range and extension of the attack. Moreover, the work [22] has also discussed the types of MANET attacks, and they proposed a detection system that could be particular only for detecting the black hole attack without any consideration for warm hole attack or gray hole attack that was eliminated by the proposed IDS in paper [22]. However, the authors [22] have introduced a new detection system depending on a novel method, which consists of a combination of adaptive Neuro-Fuzzy inference system associative with Particle Swarm Optimization to detect the black hole attack on MANET network. The experiment was run on Linux Ubuntu using three types of simulation software (Network Simulator-2) for MANET simulation, and MatLab for simulate Neuro-Fuzzy Inference System and Particle Swarm Optimization and the third simulator was QTFUZZYLITE for encoding fuzzy interface system on C++ programing language to integrate it with the IDS. Additionally, the simulation had done on numbers of parameters, which are 50 mobile nodes within a square of 800800 m. As a result, the authors decided that the proposed approach addressed a good detection rate against a black hole attack on the MANET network, but then it could increase the overhead of routing normalization. The researcher [19] has established a fuzzy intrusion detection system to detect the Neptune attack, which is a type of TCP SYN flooding attack and compared the performance of the proposed IDS with the decision tree classifier by using the NSL-KDD as an intrusion dataset and the system implementation has done on Matlab. Additionally, the membership function calculated using three membership values (High-Low-medium) depending on three attributes (Normal-Attack-Mixed) and the fuzzy rule implemented using the form of IF-THEN statements. The result of the fuzzy classification rule compared to the decision tree classifier on the NSL-KDD dataset and the results were shocking, where the accuracy of the proposed system in detecting Neptune attack was lower than the performance of decision tree. Consequently, [19], another group of researchers [23], had enhanced the previous ideas by establishing a layered classifier for detecting many types of attacks including Neptune attack using entropy feature selection algorithm associated with the layered fuzzy control language to generate the fuzzy rules. The experiment was done on the KDD-Cup1999 dataset using an open-source java environment called fuzzy logic. On the other hand, the proposed fuzzy layer consists of three layers fuzzification, fuzzy inference system that used in [19], and defuzzification and three different membership matrixes used in testing the system performance. At the end of the testing phase, the system has addressed a better detection accuracy than the proposed system in [19], and that could be the result of using efficient features selection method. Table 3 shows a comparison between fuzzy logic studies.

Table 3: A comparison between the Fuzzy Logic proposed systems.

System	***Paper***	***Dataset***	***Attack type***	***Simulation software***	***Detect***
Fuzzy logic IDS for MANET	[20]	NO	Black hole-Gray hole	NS-2	attack, range, extension
ANFIS and PSO for MANET	[22]	NO	Black hole	NS-2- MATLAB	black hole attack
fuzzy logic IDS	[19]	NSLKDD	neptune	R Studi	Low accurate
layered fuzzy control classifier	[23]	KDD-Cup	Many network attacks	jFuzzyLogic	best detection rate

4.2 IDS based on Machine Learning and Data Mining

Recently many studies in IDS have used different machine learning techniques for enhancing the detection accuracy of attacks. Either by implementing different test cases for comparing the machine learning classification algorithms on well-known intrusion datasets or by improving and developing a new classification framework for IDS. Moreover, machine learning has introduced some classification algorithms used by many researchers in many fields, and the well-known of them are Naive-Bayes, J48, Random-Forest, and Random-Tree. One clear example is the study that was done by some researchers [16] on the KDD-CUP99 intrusion dataset. They classify the attacks class using the four machine learning classification algorithms (Naive-Bayes, J48, Random-Forest, and Random- Tree), and the implementation was done on (WEKA 3.8) software, which provides an environment for knowledge analysis with Best-First-search method. Additionally, the final step of the experiment is to obtain the most accurate classification algorithm that could enhance the attack detection accuracy and minimize the false alarm problems. As a result of the previous study [16], the authors proposed that according to the level of precision value and F-measure value, Random-Forest and Random-Tree could entail the best intrusion classification algorithms comparing to Naive-Bayes and J48. On the other hand, another study [12] has followed the same strategies of [16] in obtaining the best classification algorithms. However, in the second paper [12] the comparison has done only on three classifications algorithms, namely Naive-Bayes, J48 and Random-Forest using the NSL-KDD intrusion dataset and they also used the same analysis software (WEKA 3.8) as in the case of the previous paper [16]. Additionally, the study [12] had explored 20 percentage of the NSL-KDD dataset for executing the experiment procedure, which follows two types of equations for measuring the performance of each classifier and seven different phases. As a result, the Random Forest classifier addressed the highest outcome and better performance compared with Naive-Bayes and J48. Clearly, according to the results that explored from the two studies concerning the type of datasets, we can estimate that Random Forest could be the best attack classification algorithm for IDS. Table 3 shows a comparison between the previous studies where Precision: represent the total number of intrusions, F-Measure: is a value that evaluates the correctness of the test case and Recall: is the percentage of correctly detected intrusions. On the other hand, multiple studies have worked on new frameworks that aggregate the machine learning knowledge to build an enhanced IDS. The first framework in 2017 had presented in the paper [25] used the advantage of recurrent neural network (RNN), which considered as a class of artificial neural networks that could remember the previous output and predict the next output depending on the previously recalled result. RNN includes three main units: input unit, output unit, and the most important unit, which is the hidden unit. However, the value of hidden unit appears in its ability to remember the previous result and feedback, so the researcher in [25] proposed to consider the hidden unit as a storage for the whole network traffic and to build a bi-directional flow RNN-IDS model that follows two steps forward propagation and backpropagation, where forward propagation is in charge of calculating the output values. In contrast, the backpropagation passes the remaining values that heaped up to update the weights. Additionally, the study had tested the accuracy of the proposed model using the NSLKDD dataset and compared the performance of RNN-IDS with the other machine learning classifications such as J48, naive Bayesian, and random forest. Moreover, they concluded that the RNN-IDS model is not only a strong model for intrusion detection but also includes a high accuracy as binary and multiple class classification. On the other hand, another study [26] in the same area had proposed a similar framework for the anomaly intrusion detection system that could extend to be appropriate to use on the internet of things (IoT) systems by using the novel nonparametric Bayesian which is one of the machine learning models. However, one of the advantages of nonparametric machine learning models is that it constitutes an approach to model selection and adaptation, and the sizes of the models could increase depending on the data size. Additionally, in nonparametric Bayesian, parameter space is typically chosen as the set of all possible solutions for a given learning problem. In the case of the paper [26], the researcher had proposed a new Bayesian approach for the innate bounded generalized gaussian mixture model with features selection algorithms that could enhance the detection accuracy of AIDS the proposed model had to be evaluated in multiple stages using different intrusion datasets KDD-Cup99 and ISCX on a simulation application for anomaly intrusion detection. As a result of the previous study [26], the effectiveness of the proposed model has addressed especially for IoT systems, and the researcher attempted to build a specific dataset for IoT real environments. Furthermore, the study [24] has proposed a multi-level intrusion classifier; in the first step, the random forest algorithm used to extract features from a subset of the KDD-Cup99 dataset to reduce the data redundancy. The output data trained and used multiple machine learning classification algorithms such as a k-nearest neighbor, decision tree and Naïve Bayes the output from this step, which is the new training dataset, will enter the final stage, which uses the neural network as data mining techniques to produce the new training dataset. The proposed intrusions classifier has proved its efficiency in detecting DOS attack, Probe, and U2R. Besides, machine learning algorithms had played an important role in enhancing IDS performance, especially in alter management processes, which has discussed in detail in [10]. The demand for altering management increased with the rise in the IDS wrong alters (FP and FN) rate, which reduces the amount of detection accuracy and leads to low performance with an expensive intrusion detection process. However, alter management had two different processing techniques, as it mentioned in [10], i.e., the low-level processing or high-level processing. Each level is associated with specific functionality and objectives. Implementing machine learning on alter clustering and management shall follow a special framework discussed in the paper [10].

- **Pre-processing step**: A step in which all nonnumeric alter attributes such as IP addresses shall be transferred to alternative numeric values.
- **The features selection step**: It used to increase the detection accuracy, and we need to study the relationship with different alters.
- **Clustering of intrusion alerts**: This is the most challenging step in the framework, and it depends mostly on the efficiency of the chosen machine learning technique.

4.3 IDS based on Biometrics System

Biometrics and Biometrics systems have played a significant role in users' authentication and authorization in many information systems. However, Biometrics can be defined [27] as the statistical measurement of human biological characteristics such as iris or fingerprint scanning or voice recognition or behavioral characteristics, such as keystroke and signature. For many years Biomatrices have become limited to the processes of user authentication and authorization. The main reasons for this limitation have discussed by researchers in [28]. Firstly, Biometrics implementing systems need individual hardware devices to capture users' data; then, most of these systems require active actions from the users. However, recently some studies have tried to connect the abilities of the biometrics system with the IDS to enhance IDS accuracy besides decreasing false positive and negative alarms. Most of these researchers have selected mouse actions or mouse strokes as a behavioral biometric that could not need a particular device to implement a powerful IDS. Besides, the different studies have emphasized that the false detection alarm can happen if and only if the user changed his behavior or executed an unexpected action, which will not be concerned with the situation if the IDS used the Biometrics. Additionally, in 2016, the researchers have built a unique dataset called Balabit Mouse Challenge [29], which contains users' mouse dynamic characteristics for intrusion detection purposes. The experimental version of the dataset includes two sessions: user training session and test session. Besides, the process of creating a final user model follows three main steps, as proposed in [29]. The data divided into segments that could involve zero to many mouse movements between two points of the screen x and y that could be executed by the user before any mouse point click action PC or drag and drop action. Then segments shall be isolated using a threshold to calculate the r-time that presented the required time from the beginning of the IDS recording session. The next step features extraction was done by using individual mouse action instead of using a mouse stroke that has used in the same area, and this regard experiment was also done by another researcher [28]. However, actually, mouse action or mouse stroke had addressed the same result in [29]; the researcher has already defined the mouse actions, which have previously described in the Balabit Mouse dataset. Then that was discussed by the research paper [29]. Based on the previous dataset experiment, multiple studies have to appear in this area, such as AIDS proposed by [28] to detect the user level attacks such as masquerade attacks using a statistical user profile for mouse strokes associated with keyboard stroke. However, the researcher proposed that this AIDS will offer real-time and dynamic system monitoring. Moreover, the proposed AIDS in Fig 4.

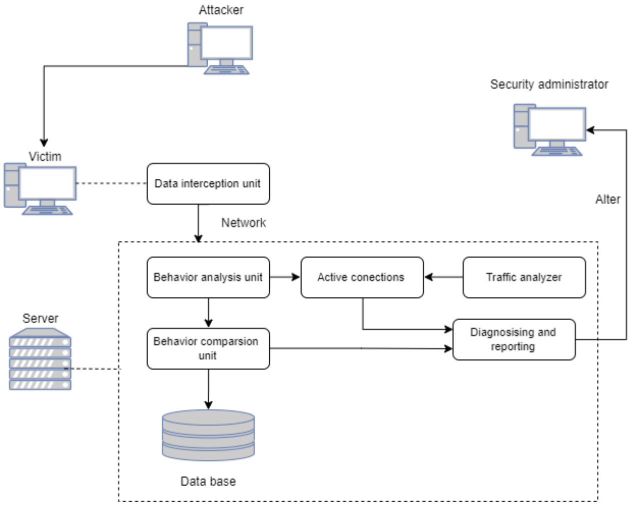

Fig. 4 AIDS based on biometrics system.

It has to deployed as a client and server software. Whereas the client-side of the IDS run as real-time auditing that is responsible for collecting the keyboard stroke and mouse stroke, and it sends them to the server-side that runs on a remote machine and responsible for analyzing and computing the biometrics data profile. Additionally, the AIDS expected to follow the main steps of any other biometrics system containing enrollment and identification or a verification linked with a database for user profile which consists of all stored mouse strokes and keyboard strokes in the identification and verification stages multiple comparison algorithms can use if there is no matching between the database and the behavior considered as an intrusion The same idea was proven by another study [21] on the MANET network, which is a collection of wireless nodes, where each node could change its geographical position consequently and serve as a router that is in charge of forwarding network packets. On the other hand, MANET could be useful in conferences and crowd controls, but it's vulnerable for different types of internal and external attacks. However, they [21] also use the mouse strokes as [28] to establish a hierarchical IDS that implements the idea of stretch and shrink method depending on the threshold value. Moreover, this method could also work in some way as the segmentation stage [29] used a threshold value in creating a user model for the Balabit Mouse dataset. On the other hand, some researchers in the same area [30] have proved that the behavioral characteristics of biometrics such as mouse dynamic and keystroke are less efficient than the biological characteristics of biometrics in the term of information security since some users do not use mouse or keyboard regularly and it is easy for the attacker to memorize the keystroke of the target user and get the access to the system. Also, behavioral biometrics may consume more time in the authenticating process. For this issue, researchers had motivated to use physiological biometrics in establishing IDS. To illustrate this, a study [31] had used the fiscal biometric to build a facial intrusion detection system deployed in an embedded system that also takes the benefits of machine learning approaches. However, the proposed system integrated with two smart cameras, where each one has an 8-megapixel sensor with a video camera placed in a different place due to get a clear image for the entrance door and persons. Additionally, the system had training on VGG-Face-2 dataset that contains faces pictures for 44 different authorized persons, who allowed to access the monitoring area and the pictures had classified using the KNN classifier as a machine learning algorithm that always calculates the property of the similarity between the dataset pictures and the run-time pictures by counting the distance between the two pictures. The more the distance value increases, the more the property of attack increases. In the end, the facial IDs [31] addressed a false negative value of 2.7% and 2.5%; also, it obtains an intrusion accuracy of 96.82% and 97.02% and high rate in the value of false-positive by 0.1. Moreover, fingerprint, as the most popular biometrics, had been used by another researcher [30] to develop a framework for a new intrusion detection system. The proposed system had followed the same steps of the facial IDS [31], but instead of fiscal recognition, they tend to use fingerprint recognition. As a result, all the previous studies and other in the same area had proved the efficiency of biometrics system to enhance the ability and accuracy of small IDS, but there is no proof if this theory could implementing an IDS that protect a more complex system such as IoT or SOS, so this may considered as an area for future research.

4.4 Intrusion Detection System based on Genetic Algorithm

The artificial intelligence advancing algorithms have proved an increasing role in the detection of the network intrusions, one of the most known AI algorithms is a genetic algorithm which is a robust technique that has used recently in some IDS research for enhancing IDS accuracy and prevention network policies. The genetic algorithm is a programming technique that simulates biological evolution as a strategy for complex problem-solving. Additionally, [32] this algorithm has developed using the idea of Darwinian's evolution and survival principle. The main three factors that shall considered when using the genetic algorithm are fitness function, genetic algorithm parameters, and individual representation. A clear example of this is the proposed IDS framework discussed in the paper [32], where the author had used the idea of a genetic algorithm with the KDD99 dataset in the system implementation by following two stages. Firstly, the pre-calculation stage creates a set of a chromosome using training data for comparison purpose. Secondly, in the detection stage, the population is calculated for the test data and different evaluation procedures executed on the test data. Also, the set of a chromosome that has obtained in the previous stage used in this stage for finding out the fitness function for each population chromosome. At the end of the experiment, they obtain the confusion metrics for the most intrusion classes for the proposed system [32], and the system had proved a good performance in detecting the DOS attack, user-to-root, and probe. Another study [33] in the same area had followed the same strategy, and the system implementation also follows two stages in the first stage, which is the training stage. A classification rule set created using the network monitoring data and the genetic algorithm in an offline circumference. Additionally, the fitness function for the classification rules has obtained using the following equations:

$$support = \|A \,\&\, B\| \,/\, N \tag{1}$$

$$confidence = \|A \,\&\, B\| \,/\, \|A\| \tag{2}$$

$$fitness = w1 * support + w2 * confidence \tag{3}$$

Unlike the paper [32] that only assumes offline training, the system presented in [33] includes two training models offline and online training. Besides, the system achieved the primary goal of detection accuracy but with some limitations, which the author has promised to resolve as future work. On the other hand, the paper [34] has proposed an IDS similar to [32] and [33] with additional functionalities in which the system could work independently without any outside support. The proposed framework could counter any new risks or attacks without the need for continuous database updating, and it contains the facilities of sending an SMS alert message to alarm the administrator when an attack is detected. Additionally, the system methodology follows 11 different implementation stages start with coding the software by C# programming language using the Microsoft visual studio, which ends with network packets sniffing. At the end of the experiment, the system worked correctly, but the author assumes that the system needs some future improvement as what had assumed by the researcher in [33].

4.5 Internet of Things (IOT) Intrusion Detection Systems

Recently the adoption of IoT systems has faced a significantly increased rate in different life fields such as smart cities, smart grid, smart home, and many other examples for IoT systems. However, these systems have a high priority for threatening by different types of security attacks, especially the DDOS attack in which IoT systems consist of multiple devices connected by network using the theory of big data and cloud computing. Additionally, IoT systems are the best representation of systems heterogeneity that had addressed as one of the significant challenges for IDS. Consequently, many research efforts have made on this area one good example is the study [35] that proposed a hidden Markov model for intrusion detection on the network-level sensor for the smart home by using the idea of big data with AIDS and a new testing dataset with 780 records created by the researcher. To illustrate this, smart home usually consists of many advanced automation subsystems such as voice assistants, thermostats, lighting, cameras, and doorbells associated with network sensors that collect the device's behavioral data and control the smart home devices and systems with transferring signals. Additionally, the proposed system [35] has been used for learning the common behaviors in the smart home with only two sensors Google-mini and phone. Also, the experiment had generated a particular vector equation

$$\gamma = (A, B, \pi) \tag{4}$$

by using the Baum-welch algorithm for increasing the experiment rate. In the end, the hidden Markov intrusion detection model had addressed 97% on detection smart home attacks. On the other hand, adversarial attacks had been an area of study by many computer vision researchers for years. However, a few security researchers had discussed its impact on IoT systems, so unlike the previous paper [35] a group of researchers has started arguing about the problems of adversarial attacks on IoT security network [36] they proposed to use the idea of Neural Network in which it widely associated with machine learning approaches to detect and classify IoT intrusions, so they are [36] developed a Self-normalizing neural network (SNN) as a deep learning IOT intrusion detection system.

Additionally, the study has used the IoT dataset BoT-IoT to extract an adversarial simple using the Fast-Gradient-Sign Method. However, the testing and evaluation step has been repeated many times on multiple adversarial samples. In the first round, the proposed IDS model [36] had addressed 95.1% initial detection accuracy, but in the second time, FNN-IDS(feedforward Neural Networks) detection accuracy had reduced to 24% in the last step the performance of the proposed model SNN-IDS compared to the performance of FNN-IDS based on various performance metrics. AS a result of the experiment, the self-normalizing model makes it more flexible to flair based adversarial samples that had extract in this experiment, and that shows that the proposed model is convenient as an IDS for IoT systems. Multiple security defects addressed by the smart factory that may cause the stoppage in the manufacturing process, trigger malfunction and unavailability of important management information these threats occurred due to the increasing complexity of smart factory system; thus the paper [37] discussed these security attacks and proposed an IDS depending on machine learning and context-aware approaches. Additionally, the main constrains on the proposed IDS framework [37] are to be flexible and more responsive to the complicated and the shift in attack patterns. The proposed framework consists of three main steps:

- **Data capturing**: The data collection phase consists of collecting data from sensors, networks, and system resources. This step work with the concept of big data analyzing.
- **Model build**: In this step, the model is created through model learning and repeated learning by applying Clustering and Autoencoder approaches.
- **Threat conception**: Results can be scored and used as a joint score. However, the resulted score applied in a simulation graph that represents the rate of security threat, and a warning message could present.

According to Smart factory IDS [37] had addressed from 33% to 1.33 process achievement and above 29% to 1.29 intrusion detection rate. Although the three previous studies [37] [36] [35] different in their IoT fields, they had proved that IOT network security needs flexible, expert, self-predictable, and more precision Intrusion detection systems. Moreover, significant demand to solve the traditional power gird problems and the demand for achieving the customer's electricity needs during many hours of the days are the reasons to implement a new generation of electronic power gird called smart gird. Recently, a high rate of security attacks such as false data injection attacks has motivated many security researchers to study these security issues from different perspectives. To illustrate this, the supervisory control and data acquisition SCADA is a real-time system for gathering, analyzing, and controlling environmental data such as water flow control, energy, and gas purification. SCADA considered one example of smart gird that is suffering from security attacks the papers [38] [39] [40] had discussed the modules of deploying a multilevel AIDS convenience to SCADA smart gird [39]. The first module is data collection in which the data gathering from different system environment resources which could define as a big data module. The next module is the Feature selection module, which consists of multiple processing and testing approaches for the system training data. In the last module, which is the anomaly detection module, several machine learning algorithms implemented to classify the attacks besides increasing the detection rate. Finally, there is a response module where the output of AIDS sent to the security administrator. On the other hand, an additional SCADA Protection Scheme had proposed integration with AIDS in [38]. The Protection Scheme is a remedial action that performs corrective system actions during system self-maintenance stages to increase system dependability, reliability, and stability.

5. Conclusion

In this paper, we introduced the primary issues and challenges encountered continuously by IDS with the current studies and research in the IDS area that solve and enhance these problems. To conclude, we admitted that there are no universal IDS due to system heterogeneity and the increasing rate of security attacks. However, the combination of AIDS and machine learning techniques could address a potential impact on IDS performance.

References

[1] T. Shekari, C. Bayens, M. Cohen, L. Graber, and R. Beyah, "Rfdids: Radio frequency-based distributed intrusion detection system for the power grid." in NDSS, 2019.

[2] Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, "Survey of intrusion detection systems: techniques, datasets and challenges," Cybersecurity, vol. 2, no. 1, p. 20, 2019.

[3] D. G. Padmavathi, M. Shanmugapriya et al., "A survey of attacks, security mechanisms and challenges in wireless sensor networks," arXiv preprint arXiv:0909.0576, 2009.

[4] S. Sokolov, T. Iliev, and I. Stoyanov, "Analysis of cybersecurity threats in cloud applications using deep learning techniques," in 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). IEEE, 2019, pp. 441–446.

[5] S. Mohurle and M. Patil, "A brief study of wannacry threat: Ransomware attack 2017," International Journal of Advanced Research in Computer Science, vol. 8, no. 5, 2017.

[6] V. Boinapally, G. Hsieh, and K. S. Nauer, "Building a gh0st malware experimentation environment," in Proceedings of the International Conference on Security and Management (SAM). The Steering Committee of The World Congress in Computer Science, Computer . . . , 2017, pp. 89–95.

[7] G. Martin, S. Ghafur, J. Kinross, C. Hankin, and A. Darzi, "Wannacry—a year on," 2018.

[8] Khraisat, I. Gondal, P. Vamplew, J. Kamruzzaman, and A. Alazab, “A novel ensemble of hybrid intrusion detection system for detecting internet of things attacks,” Electronics, vol. 8, no. 11, p. 1210, 2019.

[9] S. Aljawarneh, M. Aldwairi, and M. B. Yassein, “Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model,” Journal of Computational Science, vol. 25, pp. 152–160, 2018.

[10] W. Alhakami, “Alerts clustering for intrusion detection systems: Overview and machine learning perspectives.”

[11] K. KR and A. Indra, “Intrusion detection tools and techniques–a survey,” International Journal of Computer Theory and Engineering, vol. 2, no. 6, pp. 1793–8201, 2010.

[12] N. Ashraf, W. Ahmad, and R. Ashraf, “A comparative study of data mining algorithms for high detection rate in intrusion detection system,” Annals of Emerging Technologies in Computing (AETiC), vol. 2, no. 1, 2018.

[13] J. A. Marpaung, M. Sain, and H.-J. Lee, “Survey on malware evasion techniques: State of the art and challenges,” in 2012 14th International Conference on Advanced Communication Technology (ICACT). IEEE, 2012, pp. 744–749.

[14] V. Jyothsna, V. R. Prasad, and K. M. Prasad, “A review of anomaly based intrusion detection systems,” International Journal of Computer Applications, vol. 28, no. 7, pp. 26–35, 2011.

[15] N. Sultana, N. Chilamkurti, W. Peng, and R. Alhadad, “Survey on sdn based network intrusion detection system using machine learning approaches,” Peer-to-Peer Networking and Applications, vol. 12, no. 2, pp. 493–501, 2019.

[16] C. J. Ugochukwu and E. Bennett, “An intrusion detection system using machine learning algorithm,” International Journal of Computer Science and Mathematical Theory, vol. 4, no. 1, pp. 39–47, 2018.

[17] M. Ring, S. Wunderlich, D. Scheuring, D. Landes, and A. Hotho, “A survey of network-based intrusion detection data sets,” Computers & Security, 2019.

[18] M. Saleh, V. Miss´eri, and M.-H. Abel, “Managing heterogeneous information in a system of information systems,” 2016.

[19] N. N. P. Mkuzangwe and F. V. Nelwamondo, “A fuzzy logic based network intrusion detection system for predicting the tcp syn flooding attack,” in Asian conference on intelligent information and database systems. Springer, 2017, pp. 14–22.

[20] E. V. Balan, M. Priyan, C. Gokulnath, and G. U. Devi, “Fuzzy based intrusion detection systems in manet,” Procedia Computer Science, vol. 50, pp. 109–114, 2015.

[21] M. Antal and E. Egyed-Zsigmond, “Intrusion detection using mouse dynamics,” IET Biometrics, 2019.

[22] H. Moudni, M. Er-Rouidi, H. Mouncif, and B. El Hadadi, “Fuzzy logic based intrusion detection system against black hole attack in mobile ad hoc networks,” International Journal of Communication Networks and Information Security, vol. 10, no. 2, pp. 366–373, 2018.

[23] S. Ramakrishnan and S. Devaraju, “Attack’s feature selection-based network intrusion detection system using fuzzy control language,” International journal of fuzzy systems, vol. 19, no. 2, pp. 316–328, 2017.

[24] J. Ling and C. Wu, “Feature selection and deep learning based approach for network intrusion detection,” in 3rd International Conference on Mechatronics Engineering and Information Technology (ICMEIT 2019). Atlantis Press, 2019.

[25] C. Yin, Y. Zhu, J. Fei, and X. He, “A deep learning approach for intrusion detection using recurrent neural networks,” Ieee Access, vol. 5, pp. 21 954–21 961, 2017.

[26] W. Alhakami, A. ALharbi, S. Bourouis, R. Alroobaea, and N. Bouguila, “Network anomaly intrusion detection using a nonparametric bayesian approach and feature selection,” IEEE Access, vol. 7, pp. 52 181–52 190, 2019.

[27] Z. Akhtar, A. Hadid, M. Nixon, M. Tistarelli, J.-L. Dugelay, and S. Marcel, “Biometrics: In search of identity and security (q & a),” IEEE MultiMedia, 2017.

[28] E. Ahmed and I. Traore, “Anomaly intrusion detection based on biometrics,” in Proceedings from the Sixth Annual IEEE SMC Information Assurance Workshop. IEEE, 2005, pp. 452–453.

[29] P. Prabhusundhar, B. Srinivasan, and M. Ramalingam, “Stretch and shrink method for security in manet using biometric and intrusion detection system,” 2018.

[30] S. S. Mudholkar, P. M. Shende, and M. V. Sarode, “Biometrics authentication technique for intrusion detection systems using fingerprint recognition,” International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), vol. 2, no. 1, pp. 57–65, 2012.

[31] G. Amato, F. Carrara, F. Falchi, C. Gennaro, and C. Vairo, “Facial-based intrusion detection system with deep learning in embedded devices,” in Proceedings of the 2018 International Conference on Sensors, Signal and Image Processing. ACM, 2018, pp. 64–68.

[32] M. S. Hoque, M. Mukit, M. Bikas, A. Naser et al., “An implementation of intrusion detection system using genetic algorithm,” arXiv preprint arXiv:1204.1336, 2012.

[33] R. H. Gong, M. Zulkernine, and P. Abolmaesumi, “A software implementation of a genetic algorithm based approach to network intrusion detection,” in Sixth International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing and First ACIS International Workshop on Self-Assembling Wireless Network. IEEE, 2005, pp. 246–253.

[34] J. P. Mehta and D. M. Rathod, “Detection of cyber attack using artificial intelligence based genetic algorithm with feedback ingestion,” 2019.

[35] S. Ramapatruni, S. N. Narayanan, S. Mittal, A. Joshi, and K. Joshi, “Anomaly detection models for smart home security,” in 2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing,(HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS). IEEE, 2019, pp. 19–24.

[36] Ibitoye, O. Shafiq, and A. Matrawy, “Analyzing adversarial attacks against deep learning for intrusion detection in iot networks,” arXiv preprint arXiv:1905.05137, 2019.

[37] S.-T. Park, G. Li, and J.-C. Hong, “A study on smart factory-based ambient intelligence context-aware intrusion detection system using machine learning,” Journal of Ambient Intelligence and Humanized Computing, pp. 1–8, 2018.

[38] V. K. Singh, H. Ebrahem, and M. Govindarasu, “Security evaluation of two intrusion detection systems in smart grid scada environment,” in 2018 North American Power Symposium (NAPS). IEEE, 2018, pp. 1–6.

[39] G. Efstathopoulos, P. R. Grammatikis, P. Sarigiannidis, V. Argyriou, A. Sarigiannidis, K. Stamatakis, M. K. Angelopoulos, and S. K. Athanasopoulos, “Operational data based intrusion detection system for smart grid,” in 2019 IEEE 24th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD). IEEE, 2019, pp. 1–6.

[40] Babay, J. Schultz, T. Tantillo, S. Beckley, E. Jordan, K. Ruddell, K. Jordan, and Y. Amir, “Deploying intrusion-tolerant scada for the power grid,” in 2019 49th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN). IEEE, 2019, pp. 328–335.

Abdullah Baz received the B.Sc. degree in electrical and computer engineering from UQU, in 2002, the M.Sc. degree in electrical and computer engineering from KAU, in 2007, and the M.Sc. degree in communication and signal processing and the Ph.D. degree in computer system design from Newcastle University, in 2009 and 2014, respectively. He was a Vice-Dean, and then the Dean of the Deanship of Scientific Research with UQU, from 2014 to 2020. He is currently an Assistant Professor with the Computer Engineering Department, a Vice-Dean of DFMEA, the General Director of the Decision Support Center, and the Consultant of the University Vice Chancellor with UQU. His research interests include VLSI design, EDA/CAD tools, coding and modulation schemes, image and vision computing, computer system and architecture, and digital signal processing. Since 2015, he has been served as a Review Committee Member of the IEEE International Symposium on Circuits and Systems (ISCAS) and a member of the Technical Committee of the IEEE VLSI Systems and Applications. In 2017, IEEE has elevated him to the grade of IEEE Senior Member. He served as a Reviewer in a number of journals, including the IEEE Internet of Things, the IET Computer Vision, the Artificial Intelligence Review, and the IET Circuits, Devices and Systems.

Samah Abuayeid received her B.Sc. degree from King Abdulaziz university in 2009. Currently, she is a Master of Science candidate at Umm Al-Qura University, College of Computer and Information Systems, Computer Science Department.

Hosam Alhakami received his B.Sc. degree in Computer Science from King Abdulaziz University, Saudi Arabia in 2004. From 2004 to 2007, he worked in software development industry, where he implemented several systems and solutions for a national academic institution. Following that, he started his postgraduate studies in UK, where he received his MSc degree in Internet Software Systems from Birmingham University, Birmingham, UK in 2009. Then he successfully acquired his PhD in Software Engineering from De Montfort University in 2015. His research interests include algorithms, semantic web and optimization techniques. He focuses on enhancing real-world matching systems using machine learning and data analytics in a context of supporting decision-making.

Tahani Alsubait is a faculty member of College of Computer and Information Systems. She earned her PhD in AI and instruction from the University of Manchester. She hold a Bachelor’s in Computer Science from King Saud University and a Master’s from King Abdulaziz University. Her research interests include knowledge representation and reasoning, data analytics and HCI.