An Intrusion Detection System using GA and SVM classifier for IoTs

Nilesh Kunhare
Atal Bihari Vajpayee Indian Institute of Information Technology and Management
Gwalior, India
nilesh954@gmail.com
Ritu Tiwari
Indian Institute of Information Technology
Pune, India
ritutiwari@iiitm.ac.in
Joydip Dhar
Atal Bihari Vajpayee Indian Institute of Information Technology and Management
Gwalior, India
jdhar@iiitm.ac.in

***Abstract*—With the emergence of technologies and the increasing number of users of the Internet, such as the Internet of Things (IoT) paradigm, there are new and modern efforts to invade networks and computer systems. Many researchers have developed various artificial intelligence-based algorithms to detect these attacks. For network security they focus on machine learning models that are used in IoT and intrusion detection. In this study, we have proposed a machine learning based intrusion detection system which is a combination of support vector machine (SVM) and Genetic Algorithm (GA). GA is basically used for feature selection and parameter optimization of SVM models. The performance of a GA-based SVM model is evaluated on an KDD Cup 99 intrusion database. First, GA optimizes the KDD Cup 99 dataset and then optimizes the weight and parameters of the SVM model. The experimental result presents that the SVM model gives prominence in terms of detection rate, accuracy, false positive rate and false negative rate and also compared to other literature works.**

I. Introduction

The expansion of computer networks leads to the rapid increase of sharing information through the internet. The information shares using the internet to various organizations, educational institutes, banking sectors, medicines, transportation, marketing, and other essential services. The exchange of resources and information through several devices results in numerous security issues over the last few decades. Any vulnerable activity, intrusions, or malicious activities on the network give rise to severe disasters that violate security policies, i.e., confidentiality, integrity, and information availability. The preventive measures are not sufficient, like network firewalls, password authentication, and virtual private networks. An IDS is a tool that monitors the network packet passes across the network and generates alerts for any suspicious activities to the administrator for appropriate actions. There are two IDS variations: misuse-based IDS detects malicious activities by comparing them with recognized activities. Commercial systems mainly adopt these systems to identify predefined patterns and have a low false alarm rate. However, anomaly IDS observes network traffic; if any deviation is found from the normal, it triggers an alert. These systems can diagnose unknown attacks.

The hacker's vulnerabilities are exploited by using a well-known information-gathering technique, port-scanning. This technique is defined as the list of open ports in the network ready to respond when a connection attempts with them. The network administrator uses this technique for troubleshooting and monitoring purposes. However, the attacker uses port scanning to gain information about the target system to access the network and resources. There are various types of port scanning techniques which are illustrated in the Figure 1. It does not harm or damage the system directly; it helps the attacker find the list of open ports by sending a message to every port. The response received from the listening port indicates the weakness of that host. The information gathered includes the host's operating system and other relevant information required for launching any attack. With computer

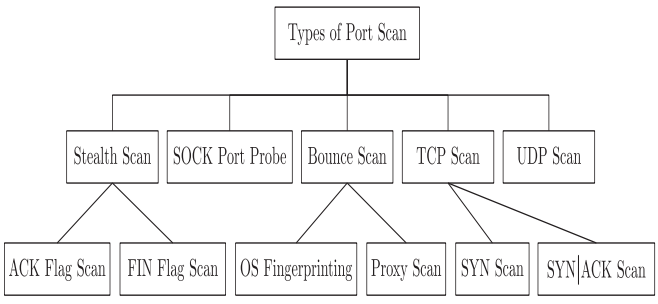

Fig. 1. Types of port scanning

networks and increasing Internet access day by day, network security is a fundamental necessity. Developing and popularising information and network infrastructure is increasingly critical for network information security [1, 2] Compared to standard network security technologies (e.g., firewalls), the practical importance of a smart human-centered Intrusion Detection System (IDS), which takes the initiative to warn and

intercept network intrusions. Network security is based on the issue of how to maximize the performance of smart network IDS [3]. Currently, the use of intelligent IDS is recognized as important network security and external attack defense approach. However, in new attacks, the original IDS also has a poor detection rate and a high overhead when using audit data, and hence machine learning methods are commonly used for the detection of intrusions. SVM, a popular supervised learning algorithm based on statistical learning theory, gives a higher performance than conventional methods of learning to solve the problem based on classification of speech recognition and pattern recognition [4].

The problem of small samples, nonlinearity, and high dimensionality can be resolved using SVM compared with other classification algorithms. However, in the new Big Data age, SVM faces a challenge with the long duration of training and testing, low true positive rates, and high error rates that hinder the use of SVM in network intrusion detection. For improved detection efficiency, the selection of the SVM function, weighting, and setting of SVM parameters play a critical role. GA demonstrates an excellent ability of global optimization through population search and share information among people. GA can conveniently avoid local optima, unlike the standard multi-point search algorithm [5].

II. Internet of the Things

"The Internet of Things (IoT) is a system of interrelated computing devices, mechanical and digital machines, objects, animals, or people that are provided with unique identifiers and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction." These physical devices have sensors that provide information at a regular interval for smart intelligence and decision-making. It is estimated that one trillion internet protocol (IP) addresses will be connected via IoT to the internet by 2022. The IoT's main objective is to make human life smart and comfortable by using the objects or sensors that connect them directly. IoT applications include smart cities, weather forecasting, smart transportation system, health care, agriculture, the education sector, and so on [6, 7, 8]. IoT devices are vulnerable to threats. The exchange of an enormous amount of data and connectivity with several devices led to various cyber-attacks. The devices used for communication in IoT are low power consumption and lightweight. The conventional security standard is not sufficient for IoT because the security protocols are centralized, but IoT supports distributed and scalable services. Thus, there is a requirement of the intrusion detection system (IDS) that solves and mitigates IoT environments' security problems. The IDS must handle various technology domains of IoT, perform analysis of the network packets, and generate the response. The IDS must be deployed that deals with a massive volume of data, fast response, and the ability to adapt to low processing conditions.

A. Characteristics of IoT

The characteristics of IoT are enlisted below:

- Interconnectivity: Anything connected for communication and information sharing globally.
- Heterogeneity: The communication takes place between different networks and platforms.
- Dynamic: The mode of devices dynamically changes like sleep, wake up, connecting/ disconnecting. It also includes the location and speed of the devices.
- Scaling of devices: It is related with management of the devices required during the communication. The efficient handling of the devices connected or not connected.
- Safety: The devices communicating with each other through the internet, thus safety of these devices becomes vital.
- Connectivity: A good connectivity result in the accessibility of the required information and compatibility provides the production and consumption of data.

III. Related Work

Malik et al.[9] suggested blockchain technology with a hybrid algorithm, including the combination of Sea Lion Optimization and Whale Optimization algorithm in vehicular ad-hoc networks. Lee et al.[10] proposed the authentication of IoT-based systems and data protection using blockchain techniques. Li et al. [11] suggested a multi-classification algorithm using double SVM for the detection of intrusions in the system. Kumar et al. [12] suggested the detection of distributed DoS (DDoS) attacks using fog computing which is a distributed framework. The author(s) implemented artificial techniques, including RF and XGBoost algorithm, for decision-making. The rapidly increasing technology in the recent decade is IoT in various applications, including wireless and wired networks. The author(s) [13] performs a detailed survey in the technologies, including machine learning, blockchain, and artificial intelligence, for the representation of various security and challenges in these domains. Calik et al. [14] studied various security and privacy issues raised in IoT-enabled domains, including monitoring devices, smartphones. The key challenges also discussed to handle the security challenges in this domain. Gao et al. [15] discussed the security and integrity validation techniques of blockchain methods in IoT-based systems. Kunhare et al. [16] suggested PSO with a random forest algorithm for feature selection in IDS to detect intrusions in the system effectively.

Intrusion detection has become a critical problem in security infrastructure in this era of big data. Several machine learning methods in IDSs, including fuzzy logic [17], K nearest neighbor KNN [18], the artificial neural network (ANN) [19], support vector machine (SVM) [20], and the artificial immune system (AIM) approaches [21], are applied to differentiate between attack and usual network access. SVM worked better than conventional classification technology [22], and many researchers have suggested SVM-based IDS [23]. Though the SVM model can increase IDS performance through detection rate and minimize error over conventional algorithms (e.g., neural networks), there is still space for improvement. With the number of internet users' data rising, IDS' performance degrades Tin terms of training time and classification accuracy. We use the GA technologies to solve these problems and provide quick and precise optimization to allow IDS to find an optimized SVM-based detection model.

In [24], it was proposed that the GA improves the intrusion IDS based on SVM and selects the optimal subset of features for SVM. However, the SVM error rate was not taken into account. In order to increase the detection capacity for SVM in complex nonlinear systems, the IDS approach based on the least square wavelet kernel was developed [25]. The algorithm is thus relatively long to train and test. In [20], the SVM Kernel parameters have been optimized by the heuristic genetic algorithm. The genetic operator is modified dynamically by a heuristic approach, and the model's performance is taken to optimize the SVM classification model based on the Gaussian kernel. However, the effect of feature weighting on SVM detection precision was not considered. In [26], the feature sub-sets and parameters of the SVM were optimized through the coarse-grained parallel GA. It has suggested a new fitness function including classification precision, number of features, number of supporting vectors, but it took a long time to train the SVM. In [27], GA was chosen as one of the most effective methods in the vast space to look for the best solution in the search space. However, with later population evolution, the lack of good genes and the algorithm's delayed convergence could lead to increased crossover and mutation likelihood.

To sum up, while several SVM-based approaches have been developed in past few decades to detect network interference, there are still some weaknesses in the above algorithms: The raw dataset confuses the classifier because of redundant features and leads to an erroneous detection. A variety of sensitive features are neglected by conventional feature filtering (such as PCA), so that a classifier does not have optimal sensitivity. In GA's case, the training period is longer, and the error rate is higher when choosing the best function subset to optimize an SVM IDS. The value of features is not sorted until choosing the best sub-set feature. We are thus proposing the combination of the GA with the SVM.

IV. Material and Methods

The following figure 1 represents the proposed method of intrusion detection using GA with an SVM classifier. GA is used in this study to select features and optimize the parameters of the SVM model. The proposed method consists of four main parts of the IDS, as shown in Figure 2 . These are feature selection, feature weight and optimization of parameters, SVM training, and finally, classification of data.

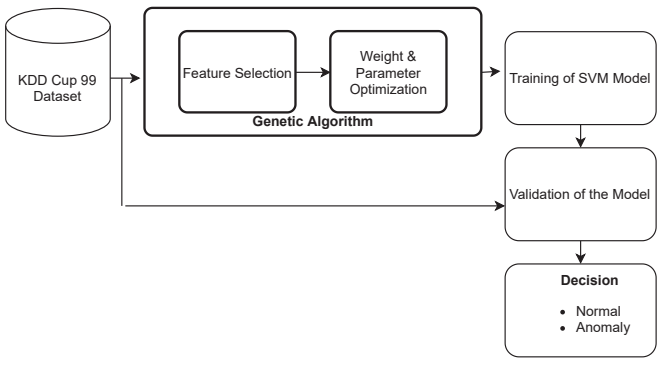

Fig. 2. Proposed model for IDS using GA-SVM approach

A. Network Dataset

KDD CUP 1999, benchmark and widely used intrusion databases for evaluating different entry-level strategies. It is based on the DARPA 98 database (DARPA 98 IDS testing program, run by Lincoln Labs), which contains 5 million simulation records mimicked in the U.S. Airforce Military. However, Tavallaee et al. proposed the NSL-KDD database [28] due to various limitations in the 1999 KDD commission, such as unequal distribution of samples, duplication, and repetition of records, etc. Each sample in the NSL-KDD database is defined by 41 conditional features followed by a class label. Any network behavior that deviates from "Normal" is considered an attack (class label).

B. Genetic Algorithm

NP-hard problems face a complicated issue of Feature selection problem. It is possible to state as follows: m instructive feature settings are chosen from the n number of elements or features, where the value of n is greater than that of m. Instructive features are anticipated to need minimal calculation requirements and demonstrate greater precision. Under this topic, the highly efficient selection of features using the genetic algorithm is illustrated in-depth. A genetic algorithm [29] is a global search algorithm for adaptive heuristics that are promoted from the process involved in the evolution of living beings. Build a solution by looking for a solution space on a global scale. It uses the functions of selecting, reassembling, and mutation operations to find solutions to the problem. In the problem of selecting a factor-based in GA, the first random number of chromosomes are produced according to the characteristics defined as input. The elements are encoded as genes on the chromosome, where ?1? represents the ability of the element and ?0? represents the absence of the element. Every chromosome in humans demonstrates distinct characteristics and is known as a distinct personality.

When chromosomes have been produced, fitness associated with every chromosome is evaluated. The degree of severity represents the degree of chromosome suitability for problem-solving. Chromosomes with more significant amounts of immunity are chosen to be parents of the next generation. The top two human chromosomes are selected in elitism; afterward, different selection strategies are incorporated to opt for the remaining parent chromosomes. Roulette wheel and feature selection strategies based on the level of a few widely used GA selection strategies [30]. In this exercise, a contest-based selection process is included. Once the parent chromosomes of the coming generation opted, they performed crossover and mutation operations. In crossover operations, A few chromosome blocks are exchanged at random., and in conversion operations, certain fragments of chromosomes are moved from 1 to 0 or 0 to 1 depending on crossover values and mutations, respectively. Crossover operators help generate new interest for each generation, while conversion operators help prevent the problem of clutter in local areas. This process is repeated until the correct solution is found. Features of the right solution have been chosen to be the most experienced features.

C. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a binary controlled and supervised learning algorithm that searches for a suitable hyperplane to draw a line between the two categories. This line breaker performs a risk analysis of the structure of mathematical learning theory by selecting specific parameters according to the limits separating the data points [31]. The hyperplane is specified as $f(x)$ = w x + b, in which w is the weight vector and b is the bias. Our separator defines the hyperplane and systematically classifies these two categories: traffic congestion and normal traffic. The points nearest to the hyperplane are being called support vectors, and the distance between them is known as the margin. On the other hand, the database is not always categorized or separable. This problem is solved by using slack variables and defining a less reliable margin. Additionally, using kernel functions, we can convert non-linear SVM to a linear task by inserting a database map into the top feature space.

D. Working of GA-SVM model

Network congestion or traffic data are entered, characteristic chromosomes are formed, chromosomes according to the fitness method suggested in this report are evaluated, chromosomes have a high degree of fitness method value as the appropriate chromosome is selected, and the optimum attribute subset is decoded. Feature weights and SVM criteria chromosomes are formed as per the subset of the optimum attribute. By evaluating the chromosome with the highest accuracy of the classification and choosing it as the optimum chromosome, the optimum SVM parameters/criteria and feature weights are decoded. The initial data is divided at random into small parts of equal size (say k) to keep the first, second, ... k, and remaining small parts of $k-1$ sub-portions used as training data to train the SVM. The subdivisions of the first, second, ... k are listed as test data and cumulative predictive k effects. The advantage of this approach is that all test sets are not dependent on each other and will improve the findings' correctness. In this experiment, we selected $k = 10$ and combined the k results to measure the SVM classifier's efficiency.

E. Steps for feature selection using GA

Step 1: Initial population and original data are generated.

Step 2: To design the SVM training model and apply the training dataset with decoding parameters, then apply the test dataset for validation of the model, and use the fitness function on the predicted results obtained to calculate each chromosome.

Step 3: To find the termination condition of GA: If the maximum number of iterations (100) or the currently obtained maximum fitness value is subtracted from the previous step, the maximum fitness value is equal to or less than 0.001. Go to Step 5. else go to Step 4.

Step 4: To perform a selection, crossover, and mutation operation on the parent population of GA, afterward a population of offspring is created, then go to Step 2.

Step 5: Decode the chromosome that has the maximum fitness value and obtains the optimal feature subset.

The chromosome feature here is a binary coded value like 1110000111000?..10011, with the value 0 representing that the feature index is not chosen while 1 indicating that the feature index is chosen. Fitness function is a key element in GA that can test a person?s readiness to survive. In this report, a new GA fitness function is suggested in order to reduce the rate of error (error rate) and improve the positive true rate while selecting optimum set of features. Each set of features is examined by the new fitness function using 3 different parameters, namely, the true positive rate (TPR), the error rate (Error) and the number of selected features (SF). The equation used is as follows:

$$Fitness(F) = W_1 * TPR + W_2 * error + W_3 * SF \quad (1)$$

The proportion of samples appropriately predicted by the classifier in all samples having actual positive category is known as true positive rate(TPR), and the equation used for calculation is as follows:

$$TPR = TP/((TP + FN)) \quad (2)$$

The proportion of the samples predicted incorrectly by the classifier in all samples, is known as error rate(ER) and the formula used for calculation is as follows:

$$error = ((FP + FN))/(\sum N) \quad (3)$$

Where W_1 is the TPR weight value, W_2 is the Error weight value, and W_3 is the selected feature number weight value. Generally, W_1 and W_2 can be set from 75% to 100% depending on user requirements. In this paper, W_1 is set to 40%, W_2 to 50%, and W_3 to 10%, to find the highest TPR, lowest Error, and smallest subset of the features. The value of TP, FP, FN, and TN are obtained from the confusion matrix.

F. Steps for feature weight and parameter optimization of SVM model using GA

SVM faces two problems after selecting the optimum subset of features: how to choose the optimum parameters for SVM and how to sort the importance of the feature. These two problems should be solved simultaneously because the kernel parameter influences the weighting feature and vice versa.

Chromosome Design for SVM model: The chromosome design has a real numeric code. The RBF kernel function of SVM is used to convert a completely indivisible problem into a fragmented or non-fragile state. The RBF kernel parameter γ represents the data distribution to the space containing new features. The severity of incorrect classification in the linearly inseparable cases is represented by the parameter X. Because the SVM parameters and the weighting feature parameters interact with each other, both feature weights and parameters must be included by the chromosome.

The values of all genes present in the chromosome lie between 0 and 1, where 0 and 1 are inclusive (i.e. 0 <= value of gene <= 1). The gene values of X and γ are represented by the two genes x and y respectively, and the gene values of the

feature weights (the weight of the unselected feature $W = 0$) are represented by W_1 through W_n. Thus, x and y are mapped to $[X1, X2]$ and $[\gamma 1, \gamma 2]$ using the SVM parameters X and γ to obtain the following formulae:

$$X = X_1 + x(X_2 - X_1) \tag{4}$$

$$\gamma = \gamma_1 + y(\gamma_2 - \gamma_1) \tag{5}$$

Step 1: Chromosome formation and initial data.

Step 2: To convert chromosomes into parameters of SVM X, γ and feature weights. The train and test datasets multiply the feature value of an instance according to the given equation:

$$A_{ij} = K_{ij} * W_j \tag{6}$$

Where K_{ij} and A_{ij} are the before and after the transformation value of j^{th} field of i^{th} instance and W_j represent the weight.

Step 3: Parameters X, γ, and training datasets are applied to train SVM classifiers. The transformed test data set is then applied to validate the performance of the SVM model. Then fitness function is applied to evaluate each chromosome on the validation result.

Step 4: Check the termination condition of GA: If Yes go to Step 6 else go to Step 5.

Step 5: To perform crossover, selection and mutation operation on the parent population of GA, and then a population of offspring is created, then go to Step 2.

Step 6: The chromosome value is decoded and get optimal parameter of SVM with feature weights.

G. Fitness Function

If we talk about decoding process, the ith attribute training and test datasets is multiplied by the correlated with weights W_i $(i = 1, .., N)$, and the SVM with RBF kernel function or method is developed based on X, as well as transformed/modified training data sets. The classification correctness of the test data is used to evaluate chromosome quality. The fitness method is indicated accurately, and the computation formula is provided as follows:

$$Fitness = ((TP + TN))/(\sum N) \tag{7}$$

H. Simulation Results

In this section, the publicly available dataset KDD Cup 99 is applied to evaluate the proposed GA-based SVM model results and then compare the results with other existing works. The observation result presents the effectiveness of the GA-SVM model with analyzing classification time. The proposed GA-SVM model is implemented on Python using Scikit library. The performance of the model is evaluated in terms of false-positive rate (FPR), false-negative rate (FNR), precision or detection rate (P), accuracy, and sensitivity (TPR). These parameters are calculated as follows:

$$FPR = FP/(FP + TN) \tag{8}$$

$$FNR = FN/(TP + FN) \tag{9}$$

$$Precision(P) = TP/(TP + FP) \tag{10}$$

The basic parameters of GA-SVM model setting are defined as: number of generation (100), population size (500), maximum and minimum crossover probability (0.9 & 0.5), maximum and minimum mutation probability (0.1 & 0.0001), SVM parameter X range (0.001 ? 1000) and range of γ (0.0001 - 64) respectively.
In this study, two experiments are performed using GA. In the first experiment, GA optimized the feature subset of the dataset. Out of 41 features, GA has selected 19 features according to the procedure discussed in Section 2. After that SVM classifier is applied to classify the dataset. The second experiment GA is also optimized for the SVM model parameters with selected weights, according to the steps explained in section 2.4.2. The obtained results from both experiments are shown in the following Table I and Figure 3.

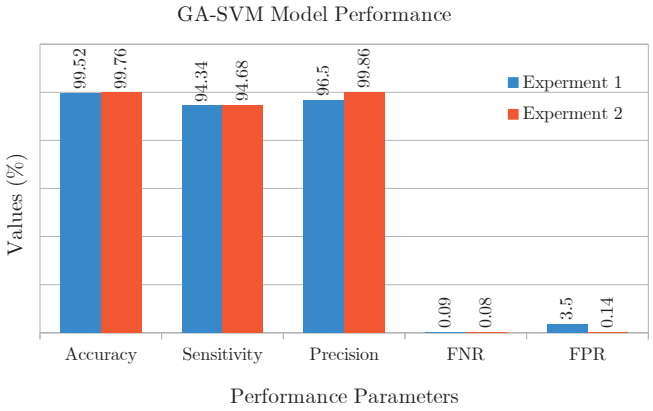

Fig. 3. Comparison performance of the GA-SVM model for both experiments

V. Conclusion

This article presents a GA-SVM model (Alarm Intrusion Detection Algorithm) elicited from a genetic algorithm (GA) and the SVM (support vector machine) algorithm used in an intelligent human-centered IDS. Firstly, this study takes advantage of the GA crowd search technique and the ability to share knowledge between people by maximizing crossover and GA mutation probability. It accelerates the convergence of algorithms and reduces the SVM training time. To reduce the SVM error rate and maximize the true positive rate, a new fitness method is proposed. Finally, simultaneously increase the parameter of the kernel, the penalty parameter X, and the function weights and improve the correctness of SVM. Simulation and experimental findings demonstrate that the advanced GA-based better intrusion detection technology

TABLE I. PERFFROMANCE COMPARISON WITH EXISTING WORKS

S.No.	Authors	Proposed Model	FPR	FNR	Detection rate (P)
1	Gharaee, H et al. [32]	GF-SVM	0.31	2.5	–
2	Feng, W et al. [33]	CSVAC	6.01	1.0	94.86
3	Yerong, T et al. [34]	HGA-SVM	–	–	91.38
4	**Proposed approach**	**GA-SVM**	**0.14**	**0.08**	**99.86**

introduced in this report enhances the accuracy of intrusion detection, precision rates, and a true positive rate, reduces SVM training time, and lowers the false-positive rate.

REFERENCES

[1] C.-F. Tsai, Y.-F. Hsu, C.-Y. Lin, and W.-Y. Lin, "Intrusion detection by machine learning: A review," *expert systems with applications*, vol. 36, no. 10, pp. 11 994–12 000, 2009.

[2] S. M. H. Bamakan, H. Wang, T. Yingjie, and Y. Shi, "An effective intrusion detection framework based on mclp/svm optimized by time-varying chaos particle swarm optimization," *Neurocomputing*, vol. 199, pp. 90–102, 2016.

[3] A. Sultana and M. Jabbar, "Intelligent network intrusion detection system using data mining techniques," in *2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*. IEEE, 2016, pp. 329–333.

[4] L. Oneto, F. Bisio, E. Cambria, and D. Anguita, "Statistical learning theory and elm for big social data analysis," *ieee CompUTATionAl inTelliGenCe mAGAzine*, vol. 11, no. 3, pp. 45–55, 2016.

[5] Y. Zhang, L. Wang, W. Sun, R. C. Green II, and M. Alam, "Distributed intrusion detection system in a multi-layer network architecture of smart grids," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 796–808, 2011.

[6] H. Arasteh, V. Hosseinnezhad, V. Loia, A. Tommasetti, O. Troisi, M. Shafie-khah, and P. Siano, "Iot-based smart cities: A survey," in *2016 IEEE 16th International Conference on Environment and Electrical Engineering (EEEIC)*. IEEE, 2016, pp. 1–6.

[7] I. Khan, F. Belqasmi, R. Glitho, N. Crespi, M. Morrow, and P. Polakos, "Wireless sensor network virtualization: A survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 553–576, 2015.

[8] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," *Computer networks*, vol. 54, no. 15, pp. 2787–2805, 2010.

[9] N. Malik, P. Nanda, X. He, and R. P. Liu, "Vehicular networks with security and trust management solutions: proposed secured message exchange via blockchain technology," *Wireless Networks*, vol. 26, no. 6, pp. 4207–4226, 2020.

[10] C. H. Lee and K.-H. Kim, "Implementation of iot system using block chain with authentication and data protection," in *2018 International Conference on Information Networking (ICOIN)*. IEEE, 2018, pp. 936–940.

[11] J. Li, "Iot security analysis of bdt-svm multi-classification algorithm," *International Journal of Computers and Applications*, pp. 1–10, 2020.

[12] P. Kumar, R. Kumar, G. P. Gupta, and R. Tripathi, "A distributed framework for detecting ddos attacks in smart contract-based blockchain-iot systems by leveraging fog computing," *Transactions on Emerging Telecommunications Technologies*, p. e4112, 2020.

[13] B. K. Mohanta, D. Jena, U. Satapathy, and S. Patnaik, "Survey on iot security: challenges and solution using machine learning, artificial intelligence and blockchain technology," *Internet of Things*, p. 100227, 2020.

[14] Z. B. Celik, E. Fernandes, E. Pauley, G. Tan, and P. McDaniel, "Program analysis of commodity iot applications for security and privacy: Challenges and opportunities," *ACM Computing Surveys (CSUR)*, vol. 52, no. 4, pp. 1–30, 2019.

[15] W. Gao, W. G. Hatcher, and W. Yu, "A survey of blockchain: Techniques, applications, and challenges," in *2018 27th international conference on computer communication and networks (ICCCN)*. IEEE, 2018, pp. 1–11.

[16] N. Kunhare, R. Tiwari, and J. Dhar, "Particle swarm optimization and feature selection for intrusion detection system," *Sādhanā*, vol. 45, pp. 1–14, 2020.

[17] L. A. Zadeh, "Fuzzy logic," *Computer*, vol. 21, no. 4, pp. 83–93, 1988.

[18] S. Malhotra, V. Bali, and K. Paliwal, "Genetic programming and k-nearest neighbour classifier based intrusion detection model," in *2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence*. IEEE, 2017, pp. 42–46.

[19] R. Sen, M. Chattopadhyay, and N. Sen, "An efficient approach to develop an intrusion detection system based on multi layer backpropagation neural network algorithm: Ids using bpnn algorithm," in *Proceedings of the 2015 ACM SIGMIS Conference on Computers and People Research*, 2015, pp. 105–108.

[20] T. Mehmood and H. B. M. Rais, "Svm for network anomaly detection using aco feature subset," in *2015 International symposium on mathematical sciences and computing research (iSMSC)*. IEEE, 2015, pp. 121–126.

[21] M. Tabatabaefar, M. Miriestahbanati, and J.-C. Grégoire, "Network intrusion detection through artificial immune system," in *2017 Annual IEEE International Systems Conference (SysCon)*. IEEE, 2017, pp. 1–6.

[22] R. Vijayanand, D. Devaraj, and B. Kannapiran, "Support vector machine based intrusion detection system with reduced input features for advanced metering infrastructure of smart grid," in *2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS)*. IEEE, 2017, pp. 1–7.

[23] D. S. Kim, H.-N. Nguyen, and J. S. Park, "Genetic algorithm to improve svm based network intrusion detection system," in *19th International Conference on Advanced Information Networking and Applications (AINA'05) Volume 1 (AINA papers)*, vol. 2. IEEE, 2005, pp. 155–158.

[24] M. S. Pervez and D. M. Farid, "Feature selection and intrusion classification in nsl-kdd cup 99 dataset employing svms," in *The 8th International Conference on Software, Knowledge, Information Management and Applications (SKIMA 2014)*. IEEE, 2014, pp. 1–6.

[25] Y. Guang and N. Min, "Anomaly intrusion detection based on wavelet kernel ls-svm," in *Proceedings of 2013 3rd International Conference on Computer Science and Network Technology*. IEEE, 2013, pp. 434–437.

[26] Z. Chen, T. Lin, N. Tang, and X. Xia, "A parallel genetic algorithm based feature selection and parameter optimization for support vector machine," *Scientific Programming*, vol. 2016, 2016.

[27] K. S. Desale and R. Ade, "Genetic algorithm based feature selection approach for effective intrusion detection system," in *2015 International Conference on Computer Communication and Informatics (ICCCI)*. IEEE, 2015, pp. 1–6.

[28] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the kdd cup 99 data set," in *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*. IEEE, 2009, pp. 1–6.

[29] J. H. Holland *et al.*, *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press, 1992.

[30] A. Mishra and A. Shukla, "A new insight into the schema survival after crossover and mutation for genetic algorithms having distributed population set," *International Journal of Information Technology*, vol. 10, no. 2, pp. 165–168, 2018.

[31] A.-C. Enache and V. Sgarciu, "Enhanced intrusion detection system based on bat algorithm-support vector machine," in *2014 11th International conference on security and cryptography (SECRYPT)*. IEEE, 2014, pp. 1–6.

[32] H. Gharaee and H. Hosseinvand, "A new feature selection ids based on genetic algorithm and svm," in *2016 8th International Symposium on Telecommunications (IST)*. IEEE, 2016, pp. 139–144.

[33] W. Feng, Q. Zhang, G. Hu, and J. X. Huang, "Mining network data for intrusion detection through combining svms with ant colony networks," *Future Generation Computer Systems*, vol. 37, pp. 127–140, 2014.

[34] T. Yerong, S. Sai, X. Ke, and L. Zhe, "Intrusion detection based on support vector machine using heuristic genetic algorithm," in *2014 Fourth International Conference on Communication Systems and Network Technologies*. IEEE, 2014, pp. 681–684.