Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:55 Issue:04:2022
DOI 10.17605/OSF.IO/UQ3D6

DATA SECURITY AND PRIVACY IN IOT USING BUTTERFLY OPTIMIZATION ALGORITHM

SHIVAKUMARASWAMY G M
Assistant Professor, Department of Electrical and Electronics Engineering, Bapuji Institute of Engineering and Technology, Karnataka, India
Research Scholar, Department of Electronics and Communication Engineering, College of Engineering and Technology, Srinivas University, Karnataka, India.

Dr.RAJANNA GS
Research Professor, Department of Electronics and Communication Engineering , College of Engineering and Technology, Srinivas University, Karnataka, India.

Dr. ASHOKA K
Associate Professor, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Karnataka, India.

PRAVEEN K
Research Scholar, Department of Electronics and Communication Engineering, College of Engineering and Technology, Srinivas University, Karnataka, India.

ABSTRACT
Nowadays, the Internet of Things (IoT) application is most integrated with our daily lives and society, which are used for exposing the user to threat against their privacy. Moreover, privacy and security of data are some of the major issues in internet-based computing, which become manifolded in IoT for diversified technologies. Furthermore, IoT is adopted in many organizations and academics to protect their assets. So, the current research has proposed a novel machine learning optimized kernel framework for identifying the primary user and malware user using IoT devices. Also, access the primary user and deny malware users using the access history of the IoT device. Additionally, a Support Vector Machine (SVM) is imported to train the database and classify the malware. As well, optimize the kernel function with the help of the Butterfly Optimization Algorithm (BOA). After that, the developed framework analyses the IP address based on the threshold value of normal and malware users, then predicts the primary used based on the stored access history of NU and denies the malware user based on the access history of MU. Finally, IoT devices access the user and the proposed framework is implemented in the python tool. To check the reliability of the proposed framework launch spoofing attacks in the classification layer. Consequently, the performance metrics of the developed technique are compared with other prevailing techniques in terms of detection accuracy, False Prediction Rate (FPR), sensitivity, specificity, precision, and F-measure.

KEYWORDS: IoT application, data security, privacy, machine learning, normal user, malware, attacks, access history, database

1. INTRODUCTION

In the past decade, technologies in the digital epoch enable huge innovations and modernization of several computerized applications [1]. Among them, the IoT is a recent technology that implies a wide-reaching network including physical as well as virtual “things” interlinked with the internet [2, 3]. The interconnectivity among the devices and the internet makes it a global network of connected “things” [4]. In the future, there are possibilities that anything that will be connected can be connected. Some besides, every device has its own and unique ID to identify the device security issues of IoT are detailed in fig.1. Practically, most of the machines/devices are using

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:55 Issue:04:2022
DOI 10.17605/OSF.IO/UQ3D6

today will become smart appliances which are connected to the internet with each other [5, 6]. Moreover, many of the IoT devices like wearables, smartphones, lamps, headphones, to name a few are battery operated on concerning the minimum power utilities [7]. Furthermore, IoT is the concept of linking objects or devices over the internet and the popularity of the IoT increased quickly [8]. The technologies of IoT are helpful for several purposes that include education, communication, business development, and transportation [9]. Generally, IoT is introduced for communication through each other from remote areas or locations [10].

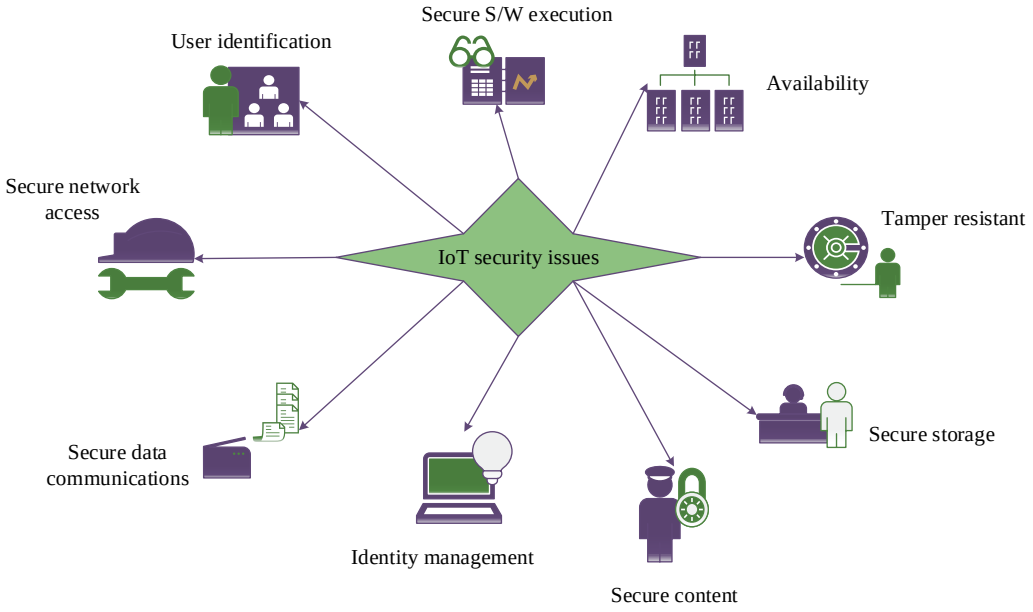

Fig.1 Security issues of IoT

As well, IoT-enabled devices are used for industrial applications and business purposes. Moreover, the app is used in business for attaining a competitive edge through competitors [11]. Also, it has the capability of connecting one device to another device and people by internet infrastructure [12]. Thus the devices are connected and communicated through several ways such as medical devices, video cameras, smartphones, consumers, and vehicle-to-vehicle communication [13]. But the main risk of IoT devices is an obvious chance of leaking sensitive information by unauthorized manipulation [14]. Also, the devices are transmitting the personal information of users like address, name, and credit card details [15]. If the IoT device contains some vulnerability means it can affect and damage the system also, the vulnerability of attacks causes data privacy [16].

Even though IoT enables interconnected smart appliances and smart autonomous devices, security and privacy are still tackling factor that restricts the performance and utilization of IoT in sensitive areas [17]. For instance, smartphones or laptops connected with IoT devices increase the risks of personal data leaking [18]. Generally, authenticating a user/device, identifying user/device and device heterogeneity is the most critical security and privacy issues in IoT [19, 20]. There are several techniques are developed to overcome security issues such as lightweight random recurrent and prediction model [21], complex event processing based machine learning method [22], and high interaction honeypot based machine learning technique [23] but still having the problem of low TPR, high FPR, high execution time and attack rate is high.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:55 Issue:04:2022
DOI 10.17605/OSF.IO/UQ3D6

The arrangement of this novel is structured as follows. The literature survey describing earlier related work based on data security and privacy is detailed in section 2 and the system model and the problems in existing system is described in section 3. Also, the process of the suggested technique is elaborated in section 4. The achieved outcomes are mentioned in section 5 and finally the conclusion about the suggested model is detailed in section 6.

2. RELATED WORKS

Some of the recent literature survey based on data security and privacy is detailed below,

Shahid et al [21] have proposed a lightweight random recurrent and prediction model for predicting the aforementioned attacks. Thus the performance of the developed technique is evaluated based on some parameters that also attain 99.20% in accuracy. Moreover, this technique enhances the attack detection rate also attains better results while comparing other existing techniques but it is more expensive and the noise rate is high.

By the increasing rate of IoT carried the most challenging task through detecting cyber-attacks and threats. Jose Roldan et al [22] have developed complex event processing-based machine learning methods for detecting various types of security attacks. Moreover, automatic code generation is designed to hide complexity from domain experts. Finally, check the capability of the developed technique to detect attacks through the malicious device but it takes more time to compute the process.

Jose Tomas et al [23] have developed high interaction honeypot-based machine learning technique for capturing the attacks and information created through attackers. Moreover, data generation is executed using a machine learning framework that detects the hidden pattern. Thus the developed framework attains a high level in prediction and less false negative rate but it has low sensitivity because of data complexity.

Data privacy and security are the most critical issues in internet-based computing such as mobile computing, IoT, and cloud computing. Ravi kumar et al [24] has studied the communication technologies, privacy, and security issues of IoT architecture. Also, discuss the various techniques which are applied for the privacy and security issues in IoT different layer architecture.

Kevin and Jorge [25] developed blockchain architecture of leverage properties for addressing the privacy and security of IoT applications. Moreover, the main aim of the developed technique is to improve the privacy of the user by implementing user control privacy. The user data are secured by anonymisation properties of blockchain and the technique attains better performance to enhance the security but the attack vulnerability rate is high while comparing other existing techniques.

The key steps of the present research work is summarized as follows,

- Initially, a group of normal and malicious users will be created.
- The normal users are identified as Nus and malicious users are identified as Mus.
- Here, the system will provide access to the users with ID NU and deny the users with ID MU.
- In order to identify Nus, a database is created and used to store the access history of the IoT devices by the Nus.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:55 Issue:04:2022
DOI 10.17605/OSF.IO/UQ3D6

- This database is further utilized to train the classifier to decide whether the request is from NU or MU.
- This decision-making will be performed by the efficiency of the ML algorithm SVM classifier which will be further improved by optimizing the kernel function through BOA.
- Moreover, the reliability of the system will be tested with several attacks such as spoofing and intrusion.
- Finally, the parameters are calculated and compared with other models with respect to accuracy, precision, sensitivity, specificity and F-measure.

3. SYSTEM MODEL AND PROBLEM DEFINITION

IoT is one of the interesting technology developments that may enhance connectivity around the world also increases the communication ability of information. Moreover, IoT captures information from the natural environment and human beings. The basic system architecture of IoT is detailed in fig.2. It contains a multi-layered network for sensing, data communication, and networking. Initially, the IoT device data are captured and send out through data communication which is received by the internet. Then the IoT security analyses the data flow depending upon the data usage of the user. The IoT data are transferred from the IoT device through the cloud to the internet or vice versa.

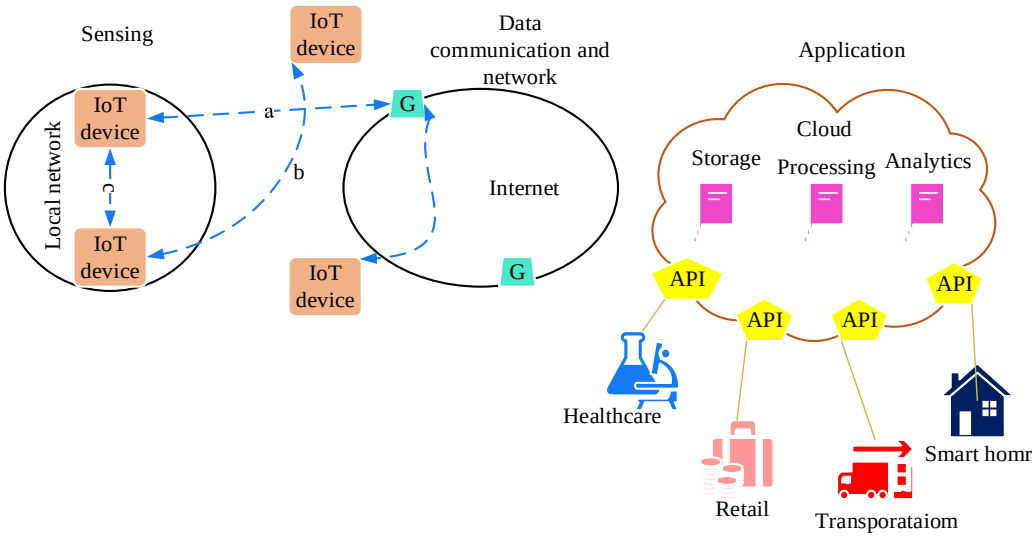

Fig.2 Basic system model

But it obtains some problems for data communication such as malicious attacks and malware. Thus the attacks may damage the system and hack the information of a user. Also, the IoT devices connected with hardware or software have the chance to leak sensitive data through unauthorized access. These factors increase the security and privacy risks in IoT so current research work is motivated to overcome the attacks and enhance data security and privacy using IoT. Here, the model is developed as a multi-objective problem.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:55 Issue:04:2022
DOI 10.17605/OSF.IO/UQ3D6

4. PROPOSED METHODOLOGY

To establish an enhanced security and privacy protocol for IoT devices, this proposal implements a Machine Learning (ML) algorithm named SVM to predict the primary user and malicious user. Furthermore, the detection performance will be improvised by optimizing the kernel function using a modified BOA. In data security and privacy of IoT, authenticating the normal user provides privacy as far as preventing attacks enables security. This proposal aims to assure optimal, secure as well as privacy-enables IoT technology. The design of the proposed methodology is illustrated in fig.3.

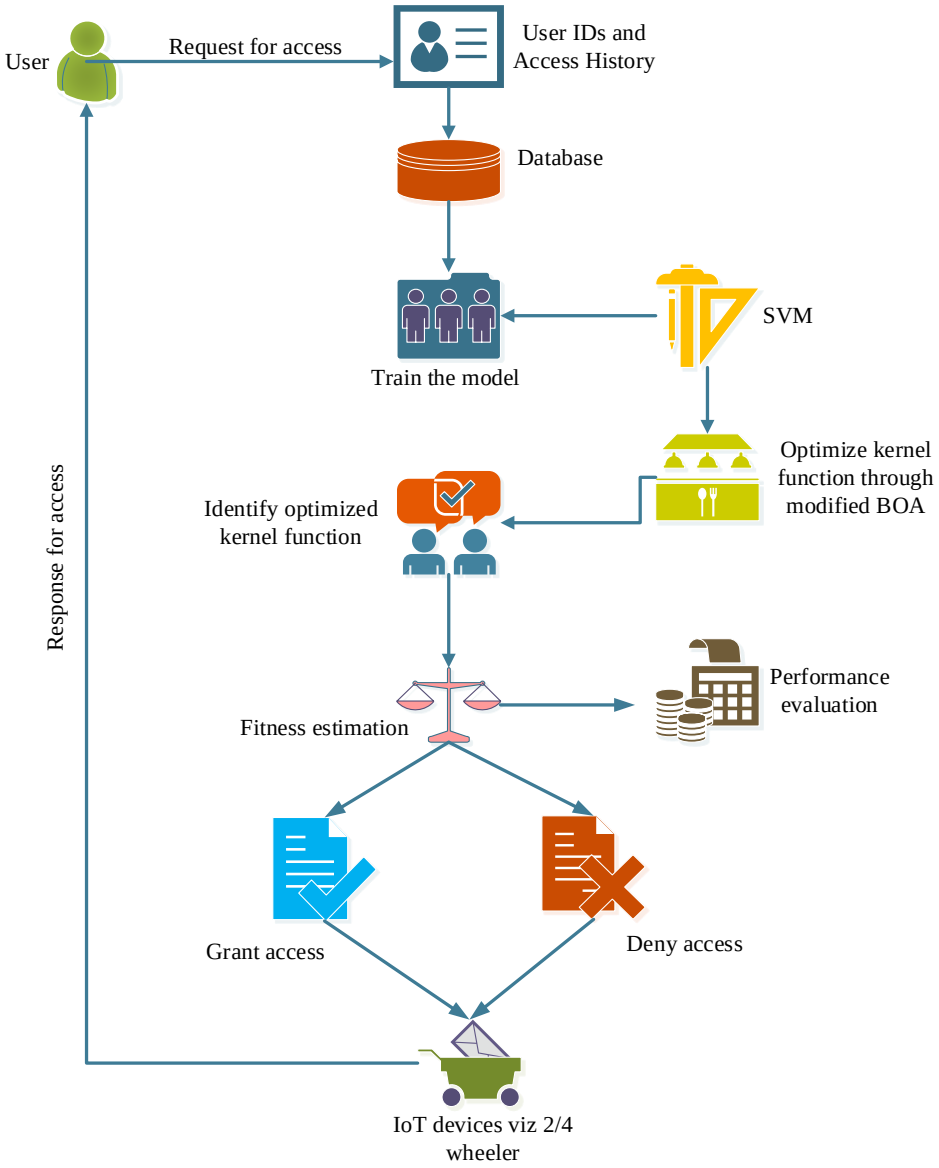

Fig.3 Proposed methodology

Initially, the dataset of normal users and malware users is collected and trained in the system. Then the normal user dataset is stored as a name of Nus and the malware user dataset is stored as the name of Mus. Furthermore, normal user datasets are stored in the access history of IoT devices. It can utilize whether the request is from

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:55 Issue:04:2022
DOI 10.17605/OSF.IO/UQ3D6

Mus or Nus. In the classification identify the malware using optimized kernel function through BOA. It includes three processes such as preprocessing, feature extraction, and classification.

4.1 Dataset collection

The normal user and malware user datasets are collected from HitGub and the dataset contains the user activity, access history of each user. Total 337941 datasets are collected from the HitGub and imported to the python environment, which contains 24% of malware and 76% of normal user datasets. The training process used 70 % of the dataset and the testing process used 30% of the dataset. Also, the dataset contains some features for identifying whether the user is normal or malware.

4.2 Preprocessing

The collected dataset contains missing values, errors, irrelevant data, noise, and training flaws which are removed using preprocessing layer. Thus the preprocessing improve the quality of the dataset also remove unwanted noise or error in the dataset. Then the preprocessing is obtained using eqn. (1).

$$P(i)=\sum_{j=1}^{M} k^{+}(i,j)-k^{-}(i,j) \tag{1}$$

Where, M is the constant, $k^{+}(i,j)$ is denoted as relevant data and $k^{-}(i,j)$ is considered as errors and noise present in the dataset.

4.3 Feature Extraction

Furthermore, the process of extracting features is helpful to identify the malware and normal user depending upon some features of each user. In this stage, extract only relevant features from the dataset so it can enhance the detection performance of the developed framework. Then the features are extracted based on source IP, destination IP, count, address, type, location, operation, timestamp, values, and so on. Additionally, the features extraction is obtained by eqn. (2).

$$\sigma(m_{ab})=\frac{d(a)p(a)+p(i)+\mu(a)}{d(d)p(b)+p(i)+\mu(b)}\left(\sigma^{*}(m_{a})\right) \tag{2}$$

Let, $\sigma(m_{ab})$ is denoted as feature extraction of user Id and access history. Moreover, $d(a)$, $p(a)$and $\mu(a)$ is considered as source IP, source address and source type. Also, $d(b)$, $p(b)$and $\mu(b)$ is considered as destination IP, destination address and, destination type.

4.4 Process of SVM with an optimized kernel function

Frequently, extracted features are stored in the database which is designed for storing the access history of the IoT device. Thus the database is further utilized in the classification layer to identify whether the request is coming from Mus or Nus. Furthermore, train the database using SVM which is used to analyze the sparse data and high dimensional data with the recognized pattern. As well, prediction of primary

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:55 Issue:04:2022
DOI 10.17605/OSF.IO/UQ3D6

user and malware user involves two steps such as training and prediction. Moreover, a database is trained using eqn. (3).

$$T_r = \left\{\left(a_i, b_i\right), a_i \in M^k, b_i \in \left\{-1,+1\right\}\right\}_{i=1}^{M} \tag{3}$$

Where, a_i is denoted as the training samples and M^k is considered as fixed length vector. Moreover, b_i is denoted as the associated binary label of M vector. Then the trained SVM searches the input dataset which separates the positive and negative samples such as ID MU and ID NU. Then the developed framework is determined a binary linear classifier which is mainly focused on the original sample of SVM and the hyper plane is parameterized using eqn. (4).

$$M^k + o(k) + a = 0 \tag{4}$$

Let, $o(k)$ is denoted as the vector orthogonal of hyper plane which separates the training data for classification using eqn. (5).

$$b(a) = sign\left(M^k + o(k) + a\right) \tag{5}$$

Where, $b(a)$ is denoted as accurate classification of training data. Frequently, the optimized kernel function is updated to the SVM for predicting the primary and malware users also enhance the prediction accuracy of malware. In this stage, the fitness of butterfly optimization is updated to the kernel function for correct predicting of malware users. Generally, the basic purpose of BOA is to sense the fragrance to the other butterfly for searching food also move randomly towards the local search. The fitness of BOA is used to identify the request is from Nus or Mus. because it can search and monitor each IP address of all users. Moreover, classification is occurred based on the comparison of the testing vector which is obtained by eqn. (6)

$$b(a) = \sum_{i=1}^{M} \left(a_i + b_i\right) M^k \alpha^k \tag{6}$$

Let, α^k is denoted as fitness function of BOA that is used for measuring the comparison of two vectors in kernel function. Thus the optimized kernel function identifies the similarity of normal and malware users by feature extraction of the trained database. Whether the primary and normal user is identified based on the sequence of calls because it contains access history of ID NU and ID MU. Thus the size of the subsequence are identified using eqn. (7)

$$D(s,t) = \sum_{M \in n} \sum_{a=s} \sum_{a=t(i)} \phi^{a(s)a(t)} \tag{7}$$

Let, n is denoted as subsequence size and ϕ is considered as decay factor based on the contribution of matching IP address. Moreover, n is considered as the length of the subsequence.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:55 Issue:04:2022
DOI 10.17605/OSF.IO/UQ3D6

Algorithm: Optimized kernel function of SVM to predict malware

Start
{
Initialize dataset
//Normal and malware user dataset
Store dataset
{
$Normal\_user \rightarrow NUs$
$malware\_user \rightarrow MUs$
}
Preprocessing
// remove noise and errors present in the dataset
For all k=1,2,3,4…..i
{
Remove errors
}
End for
Feature Extraction
// Identify the malware and normal user based features
For all $a \rightarrow Nus$
{
Normal user feature extracted
//source, destination, address, IP, timestamp, etc.
}
For all $b \rightarrow Mus$
{
Malware user feature extracted
}
End for
Update to developed SVM
{
Train the database
$T_r \rightarrow a_i, b_i, M^k$
// a_i - training samples
// M^k - fixed length vector.
// b_i - associated binary label of M vector
Binary linear classifier
For all $b(a)$=classified trained data
{
Store access history
// stored using IoT device
}
Launch attack
Update BOA in kernel function
Identify size of sequence user

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:55 Issue:04:2022
DOI 10.17605/OSF.IO/UQ3D6

{
$D(s,t) \rightarrow size, decay\_factor, length$
}
Classification
if $(D(s,t) \leq 0.1)$
{
Normal user
//ID NU
}
Grant access
else if $(D(s,t) \geq 0.1)$
{
Malware user
//ID MU
}
Deny access
End if
Neglect attack
ACCESS to normal user
Output
}
End

After that, developed SVM with optimized kernel function classify the primary user and malware user present in the dataset using IoT devices. The developed framework continuously watches the stored access history of the normal and malware user using IoT device. Then classify the primary user and malware user by defining the threshold value of sequence size of each user. Furthermore, prediction of normal or malware is obtained using eqn. (8).

$$Y_s(t) = \begin{Bmatrix} D(s,t) \leq 0.1 & \rightarrow Normal\_user \\ D(s,t) \geq 0.1 & \rightarrow malware\_user \end{Bmatrix} \quad (8)$$

While the sequence size threshold value is less than 0.1 means that the user is considered a normal user and the system can access the user. But the sequence size threshold value is greater than 0.1 means that the user is considered a malware user and the system can deny the user. At last, the developed framework response to the normal user to access the data and provide data privacy and security using smart IoT technology.

5. RESULTS AND DISCUSSION

The designed model is elaborated in Python and the successive score was validated by comparing the performance of the designed model with other approaches in terms of accuracy, specificity, sensitivity, F-measure, precision. Initially, a group of normal and malicious user datasets is collected from the net source and trained to the system. Moreover, a database is generated for storing the access history of the IoT device. That database is employed for training the classifier and the classifier is designed with

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:55 Issue:04:2022
DOI 10.17605/OSF.IO/UQ3D6

an optimized kernel via BOA. Finally, the developed technique accesses the normal user and denies the malware user. At last, check the reliability of the developed technique by testing spoofing attacks.

5.1 Case study

Generally, the application of IoT in the industrial sector enhances the efficiency, security, and production of industrial operations. Also, the internet is one of the important backbones in modern life such as e-commerce, e-learning, and e-mail.

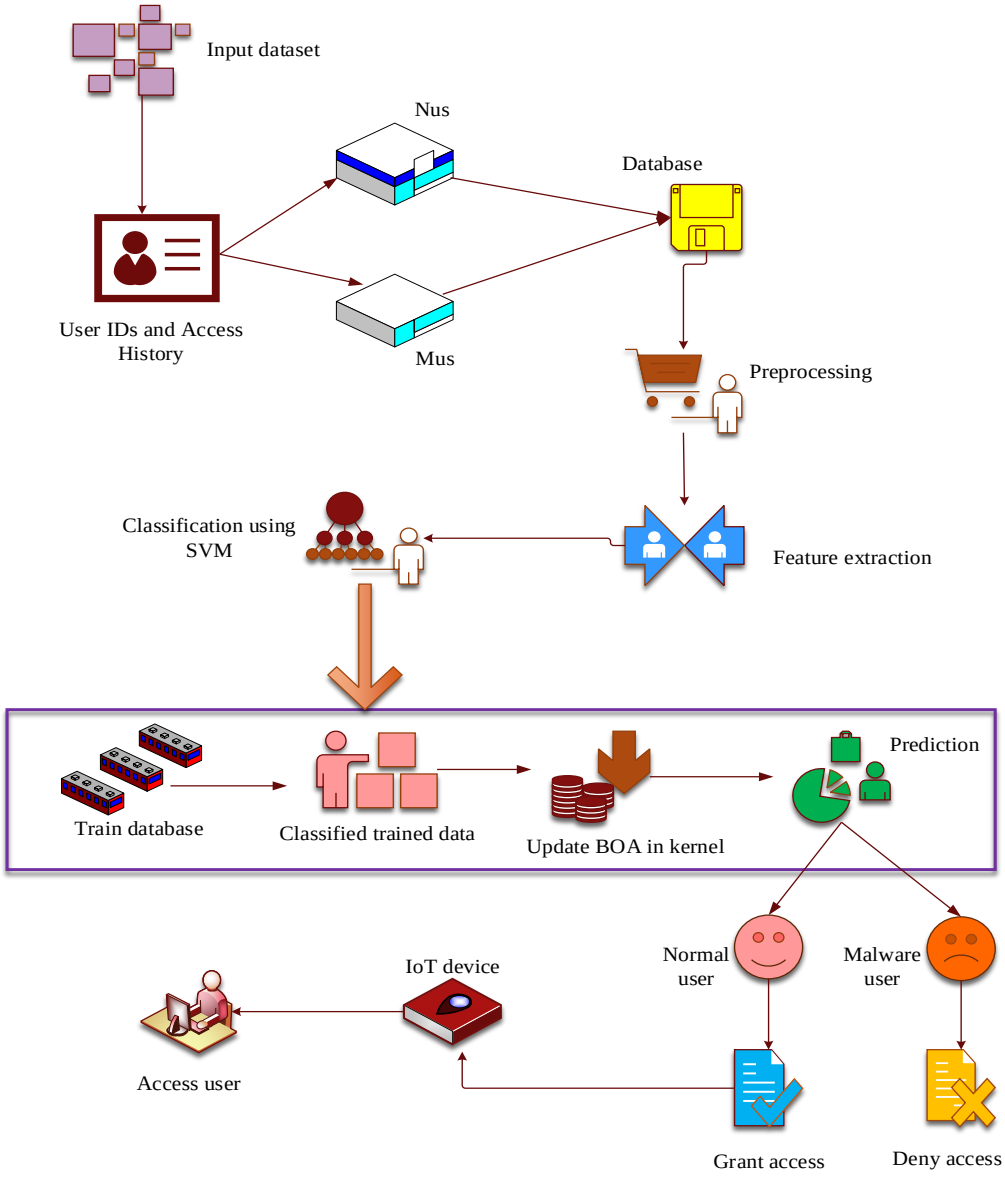

Fig.4 Process of developed architecture

Moreover, IoT is one of the applications of the internet it acts, connects, communicates, and reacts with each other without human interaction. The developed framework is imported into a python environment to secure data privacy and enhance detection performance. Initially, user datasets with various IP addresses are collected from the net source and trained to the system. Then the collected dataset is stored in

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:55 Issue:04:2022
DOI 10.17605/OSF.IO/UQ3D6

the IoT device based on the Nus and Mus. Moreover, unwanted noise, errors, and training flaws are removed in preprocessing layer at the next feature extraction are processed to extract the relevant information from the dataset which is helpful to predict the primary or malware user. Additionally, the classification layer used SVM for predicting the normal and malware user. In this stage, the optimized kernel function is updated to the SVM through BOA. It can analyze the IP address whether the request is from Nus or Mus. Then launch an attack in the classification layer to check the fidelity of the developed framework. The main purpose of spoofing attacks is for a user to steal data, spread malware; the attacker impersonates an authorized device and bypasses the access control system. Thus the attacks have affected the system with a different IP address which is identified and neglected with the help of optimized kernel function. Finally, the developed framework overcomes the issues of the multi-objective problem better performance in detection accuracy also enhance the performance of data privacy and security using IoT device.

5.2 Performance metrics

The planned model is implemented in the python tool and the success rate of the designed scheme was analysed with comparison assessment in terms of, accuracy, sensitivity, specificity, F-measure, and precision. Thus the achieved performance is compared with other existing techniques such as Lightweight Random Neural Network (LRaNN) [21], Detect Botnet Infection (DBI) [23], Multistage and Elastic Spam Detection (MESD) [26], Android Malware Detection System (AMDS) [27], Detect Android Malicious using Ensemble Learning (DAM-EL) [28], and Intrusion Detection in IoT (ID) [29].

5.2.1 Accuracy

The accuracy of the developed framework is identifying the malware user based on the access history of IoT devices. Moreover, the system will provide access to the user with the ID NU and deny the user with ID MU. Furthermore, the accuracy of identifying malware users can be expressed using eqn. (9)

$$Accuracy = \frac{IP + IN}{IP + AP + AN + IN} \tag{9}$$

Where, IP is denoted as a true identification of malware user, IN is represented as a true negative identification of malware user. Moreover, AP is expressed as a false positive identification of malware user and AP is called a false negative identification of malware user.

Table.1 Validation of accuracy

No. of. epoch	Accuracy (%)						
	LRaNN	DBI	MESD	AMDS	DAM-EL	ID	Proposed
1	99.20	98.1	91.3	98.8	98.39	97.36	99.54
5	98.1	97.2	89.6	97.6	98	96.01	99.34
9	97.34	95.03	87.06	95.34	97.3	94.89	99.21
13	97	94	85.98	94	95.54	92.4	99
21	95.43	93.54	84	92.04	94	91.13	98.67

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:55 Issue:04:2022
DOI 10.17605/OSF.IO/UQ3D6

The achieved accuracy rate of the proposed technique is compared with other existing replicas such as LRaNN, DBI, AMDS, MESD, ID and DAM-EL. Thus the LRaNN and DBI replicas attained 99.20% and 98.1% in accuracy, also the MESD method gained 91.3% accuracy. Moreover, the DAM-EL and AMDS techniques achieved 98.39% and 98.8%. Additionally, the developed technique achieves 99.54% accuracy. The comparison of accuracy with the exciting technique is detailed in fig.5 and table.1.

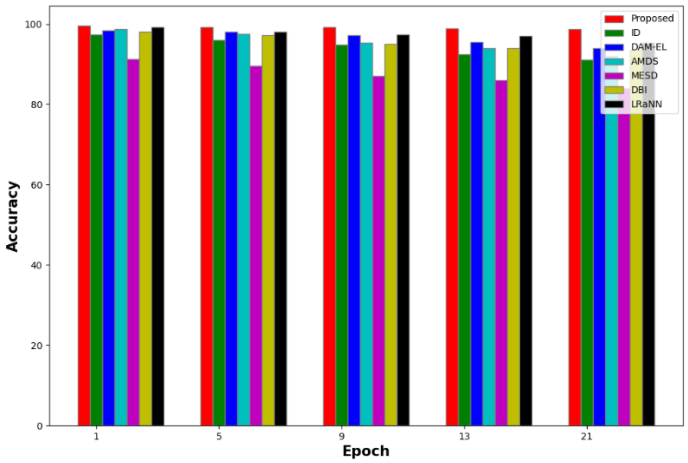

Fig.5 Accuracy comparison

5.2.2 Precision

The computation of precision (P) is operated for recognizing the success of the proposed technique while accessing the permission to the normal user and denying malware user. In addition, the measurement of precision rate is obtained using eqn. (10) and comparison of precision has been detained in table.2.

$$P = \frac{IP}{IP + AP} \tag{10}$$

Table.2 Validation of Precision

No. of. epoch	Precision (%)						
	LRaNN	DBI	MESD	AMDS	DAM-EL	ID	Proposed
1	99.11	95.7	90	95.89	95.03	97.72	99.88
5	98.72	94.32	89.12	94.2	94.78	97	99.67
9	98	93.01	88	93.02	92.34	96.54	99.22
13	97.54	92.4	86.97	92.34	92	95.8	99
21	97	91	85.34	91.3	90.01	94	98.72

Generally, DBI and LRaNN replicas attained 95.7% and 99.11% in precision; also MESD method gained 90% precision. Moreover, the AMDS and DAM-EL methods attained almost 95% in precision; also ID technique achieved 97.72%. Additionally, the developed technique achieves 99.88% in precision. Thus the comparison of precision is detailed in fig.6.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:55 Issue:04:2022
DOI 10.17605/OSF.IO/UQ3D6

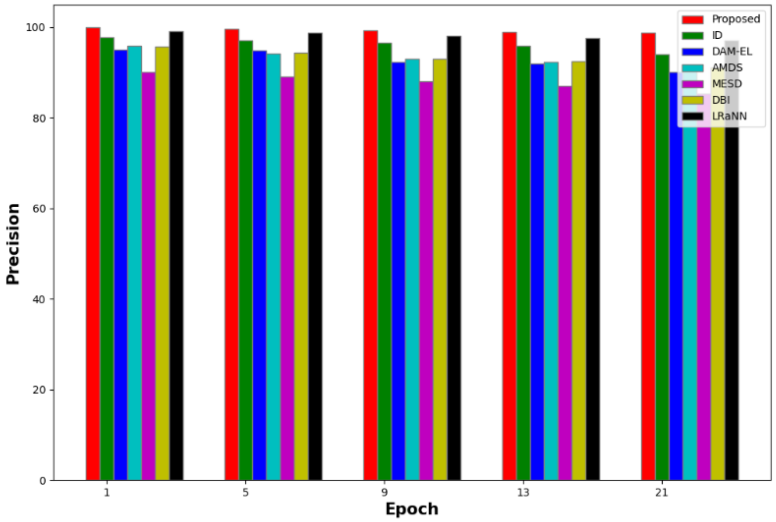

Fig.6 Precision comparison

Here, the existing approaches have achieved a lower precision value while comparing developed techniques has obtained 99.88% high precision value than other methods.

5.2.3 Sensitivity

Measurement of sensitivity is developed to access the normal user and deny the malware user which is mainly predicting the malware user based on access history. Additionally, sensitivity is the term of true positive value to the addition of false-negative and true positive value. Moreover, the sensitivity calculation of the developed method was obtained using eqn. (11),

$$Sensitivity = \frac{IP}{IP + AN} \tag{11}$$

Table.3 Validation of Sensitivity

No. of. epoch	Sensitivity (%)					
	LRaNN	DBI	MESD	AMDS	ID	Proposed
1	99.13	93.9	89.5	98.20	96.92	99.56
5	98.67	92.89	88.12	97.3	95.1	99.23
9	97.6	91.5	86.7	96	93.45	99
13	97	90.02	84.3	95.23	92.01	98.62
21	96.6	89	82	94.12	91.1	98.12

The achieved sensitivity rate of the proposed technique is compared with other existing replicas such as LRaNN, DBI, AMDS, MESD, and ID. Thus the LRaNN and DBI replicas attained 99.13% and 93.9% in sensitivity, also the MESD method gained 89.5% sensitivity. Moreover, the AMDS and ID techniques achieved 98.20% and 96.92%. Additionally, the developed technique achieves 99.56% sensitivity. The comparison of sensitivity with the exciting technique is detailed in fig.7 and table.3.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:55 Issue:04:2022
DOI 10.17605/OSF.IO/UQ3D6

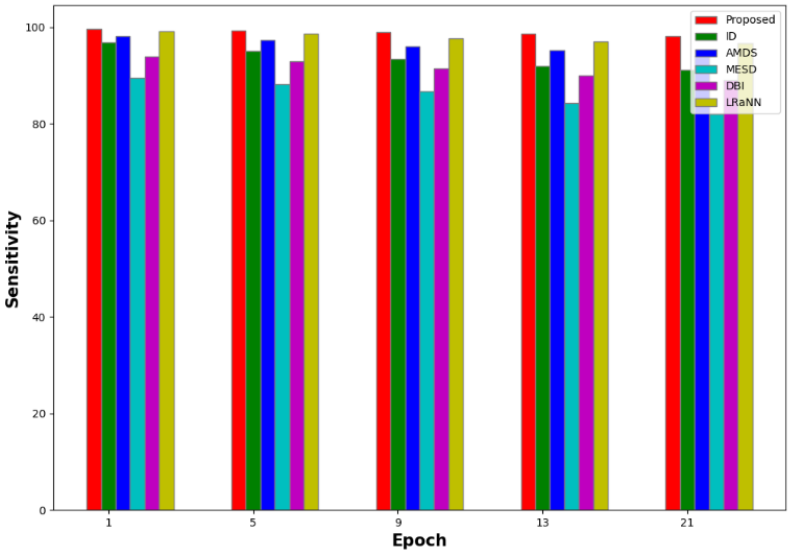

Fig.7 Sensitivity comparison

5.2.4 Specificity

Specificity is defined as the degree that is utilized for identifying the amount of true negatives that are recognized accurately. Also, specificity is employed for calculating the efficiency of detecting malware using the access history of IoT devices. The mathematical expression to calculate specificity is detailed in eqn. (12), and the comparison of specificity is shown in table.4.

$$Specificity = \frac{IN}{IN + AP} \tag{12}$$

Table.4 Specificity validation

No. of. epoch	Specificity (%)					
	LRaNN	DBI	MESD	AMDS	ID	Proposed
1	98.76	92	90.5	98	95.12	99
5	98	91.03	89.67	97.12	94.88	98.67
9	97.12	90.45	88	96.39	93.13	98.12
13	96.67	88.3	87.60	95.12	92	97.78
21	96.01	87	86.12	94	91.67	97.24

The achieved specificity rate of the developed technique is compared with other existing techniques such as DBI, MESD, LRaNN, ID and AMDS. Moreover, the ID replica attained 95.12 for one epoch. Also, LRaNN and DBI techniques gained specificity rates are 98.76% and 92%. The developed MESD and AMDS methods attained specificity rates as 90.5% and 98%. The developed framework attains a high specificity rate while comparing other techniques to predict malware as 99%. The comparison of specificity is illustrated in fig.8.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:55 Issue:04:2022
DOI 10.17605/OSF.IO/UQ3D6

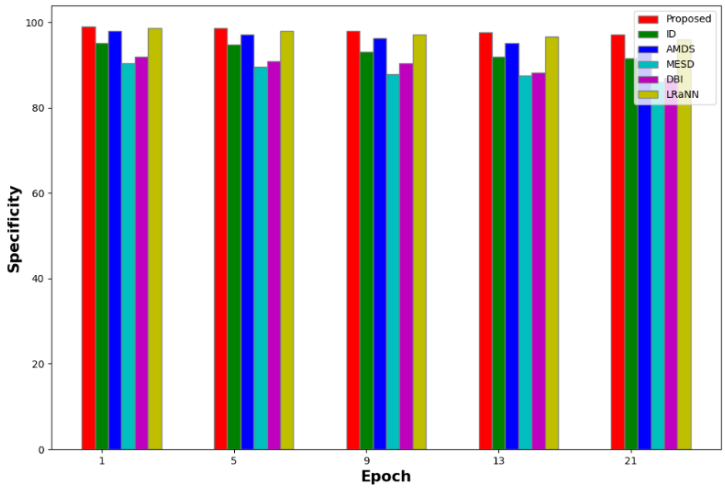

Fig.8 Comparison of Specificity

5.2.5. F-measure

It is the measure of model accuracy of the dataset used for evaluating binary classification systems. Also, it is calculated by combining both precision value and recall value, which is calculated using eqn. (13),

$$Fmeasure = \left(2\frac{P * R}{P + R} \right) \tag{13}$$

Table.5 Validation of F-measure

No. of. epoch	F-measure (%)				
	LRaNN	DBI	MESD	AMDS	Proposed
1	99.20	94.8	91.5	97.02	99.57
5	98.56	93.56	90	96.34	99.12
9	98	93.02	88.12	95.02	98.78
13	96.28	92.62	87.87	94.12	98.28
21	95.98	92	86	93	97

The DBI replica attained 94.8% in F-measure, and the LRaNN method gained 99.20% F-measure. Moreover, the MESD and AMDS techniques gained 91.5% and 97.02%. Additionally, the developed technique achieves 99.57% in F-measure and the comparison of F-measure with the exciting technique is detailed in table.3 and fig.9.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:55 Issue:04:2022
DOI 10.17605/OSF.IO/UQ3D6

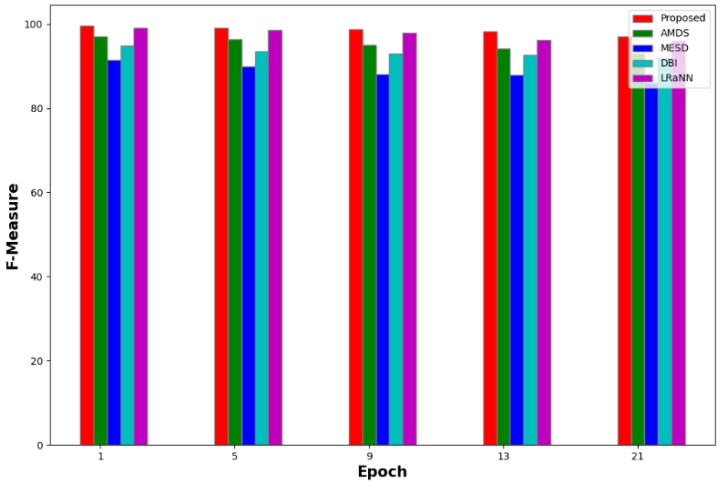

Fig.9 F-measure comparison

5.2.6 False Positive Rate (FPR)

Generally, FPR is the ratio among quantity of negative identification of malware users which are wrongly categorized as the false positive and total quantity of actual negative events. Thus the FPR is measured using eqn. (14).

$$FPR = \frac{AP}{AP + IN} \tag{14}$$

The gained FPR rate of the developed technique is compared with other existing techniques such as AMDS, ID, and DAM-EL. Furthermore, DAM-EL technique gained 0.016% in FPR and AMDS technique achieved 0.014% in FPR. Additionally, ID method attained FPR rate as 0.37% but the developed replica attain low FPR as 0.001%. The comparison of the FPR is detailed in fig.10.

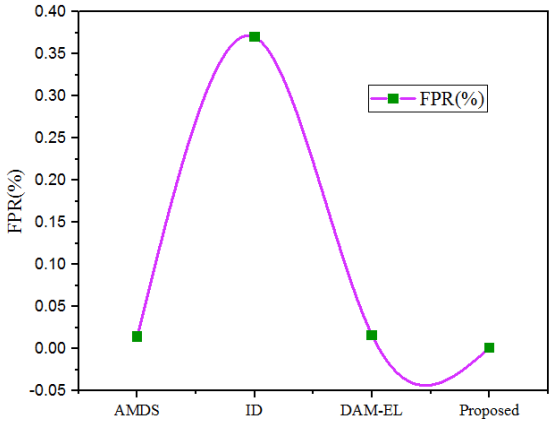

Fig.10 FPR comparison

5.3 Discussion

The proposed model has shown good performance by attaining the best results in accuracy, recall, precision, specificity, FPR, and f-measure. Thus, the developed scheme identifies the normal user and malware user from Nus and Mus. The normal

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:55 Issue:04:2022
DOI 10.17605/OSF.IO/UQ3D6

user database is stored in the access history of IoT, which can grant access to a normal user and deny malware users. To check the reliability of the developed technique, launch spoofing attacks. Furthermore, the proposed technique secures data privacy using IoT devices from malware users. Thus the developed technique attains high accuracy for detecting malware users.

Table.5 Overall performance metrics

	Performance assessment with key metrics				
Methods	**Accuracy**	**precision**	**Sensitivity**	**Specificity**	**F-measure**
LRaNN	99.20	99.11	99.13	98.76	99.20
DBI	98.1	95.7	93.9	92	94.8
MESD	91.3	90	89.5	90.5	91.5
AMDS	98.8	95.89	98.20	98	97.02
ID	97.36	97.72	96.92	95.12	-
Proposed	99.54	99.88	99.56	99	99.57

The outstanding metrics comparisons are tabulated in table.5, in all parameter validation, the proposed technique has gained the finest results. Moreover, the developed framework gained high detection accuracy as 99.20%, low FPR as 0.001%, and high sensitivity as 99.13%. Hence, the robustness of the proposed technique is verified and it can identify the normal user and malware users using IoT devices.

6. CONCLUSIONS

In this paper, a novel optimized SVM framework has been developed to detect malware and attacks using IoT systems also enhance data privacy and security. Thus the developed framework accesses the normal user and denies the malware user based on the stored access history of the IoT device. More than 330000 datasets are imported to the system which is trained and tested using the python tool. At first, training errors and noise are removed using preprocessing, and then significant features are extracted in the process of feature extraction. In the classification layer, train the database using SVM which is modified by kernel function with optimized BOA. The developed technique predicts the normal and malware user based on the threshold level of each user IP. Finally, access the normal user with an IoT application or IoT device. Furthermore, to check the designed model's reliability, harmful attacks are launched in the IoT environment. Hence, the proposed architecture achieved less FPR and high detection accuracy. It can improve data privacy and security using IoT devices for predicting malware and attacks.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:55 Issue:04:2022
DOI 10.17605/OSF.IO/UQ3D6

REFERENCE

1. Kirubasri, G., et al. "A Recent Survey on 6G Vehicular Technology, Applications and Challenges." 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO). IEEE, 2021.
2. Akhtar, Ahmad Naeem, et al. "Spectrum decision framework to support cognitive radio based IoT in 5G." Cognitive Radio in 4G/5G Wireless Communication Systems (2018): 73.
3. Devi, M., et al. "Data Science for Internet of Things (IoT)." International Conference on Computer Networks and Inventive Communication Technologies. Springer, Cham, 2019.
4. García, Cristian González, Liping Zhao, and Vicente García-Díaz. "A User-Oriented Language for Specifying Interconnections Between Heterogeneous Objects in the Internet of Things." IEEE Internet of Things Journal 6.2 (2019): 3806-3819.
5. Aazam, Mohammad, Sherali Zeadally, and Khaled A. Harras. "Deploying fog computing in industrial internet of things and industry 4.0." IEEE Transactions on Industrial Informatics 14.10 (2018): 4674-4682.
6. Avula, Vasavi, et al. "The Internet Of Everything: A Survey." 2021 13th International Conference on Computational Intelligence and Communication Networks (CICN). IEEE, 2021.
7. Mamdiwar, Shwetank Dattatraya, et al. "Recent advances on IoT-assisted wearable sensor systems for healthcare monitoring." Biosensors 11.10 (2021): 372.
8. Balaji, S., Karan Nathani, and R. Santhakumar. "IoT technology, applications and challenges: a contemporary survey." Wireless personal communications 108.1 (2019): 363-388.
9. Alam, Tanweer. "Blockchain cities: the futuristic cities driven by Blockchain, big data and internet of things." GeoJournal (2021): 1-30.
10. Diène, Bassirou, et al. "Data management techniques for Internet of Things." Mechanical Systems and Signal Processing 138 (2020): 106564.
11. Huseien, Ghasan Fahim, and Kwok Wei Shah. "A Review on 5G Technology for Smart Energy Management and Smart Buildings in Singapore." Energy and AI (2021): 100116.
12. Ghosh, Ashish, Debasrita Chakraborty, and Anwesha Law. "Artificial intelligence in Internet of things." CAAI Transactions on Intelligence Technology 3.4 (2018): 208-218.
13. Kumar, Priyan Malarvizhi, et al. "Ant colony optimization algorithm with internet of vehicles for intelligent traffic control system." Computer Networks 144 (2018): 154-162.
14. Meneghello, Francesca, et al. "IoT: Internet of threats? A survey of practical security vulnerabilities in real IoT devices." IEEE Internet of Things Journal 6.5 (2019): 8182-8201.
15. Rizvi, Syed, et al. "Identifying the attack surface for IoT network." Internet of Things 9 (2020): 100162.
16. Rizvi, Syed, et al. "Threat model for securing internet of things (IoT) network at device-level." Internet of Things 11 (2020): 100240.
17. Abiodun, Oludare Isaac, et al. "A Review on the Security of the Internet of Things: Challenges and Solutions." Wireless Personal Communications (2021): 1-35.
18. Siboni, Shachar, Asaf Shabtai, and Yuval Elovici. "Leaking data from enterprise networks using a compromised smartwatch device." Proceedings of the 33rd Annual ACM Symposium on Applied Computing. 2018.
19. Nour, Boubakr, et al. "Security and privacy challenges in information-centric wireless internet of things networks." IEEE Security & Privacy 18.2 (2019): 35-45.
20. Rahman, Mahbubur, and Hamid Jahankhani. "Security vulnerabilities in existing security mechanisms for IoMT and potential solutions for mitigating cyber-attacks." Information Security Technologies for Controlling Pandemics. Springer, Cham, 2021. 307-334.
21. Latif, Shahid, et al. "A novel attack detection scheme for the industrial internet of things using a lightweight random neural network." IEEE Access 8 (2020): 89337-89350.
22. Roldán, José, et al. "Integrating complex event processing and machine learning: An intelligent architecture for detecting IoT security attacks." Expert Systems with Applications 149 (2020): 113251.
23. Garre, José Tomás Martínez, Manuel Gil Pérez, and Antonio Ruiz-Martínez. "A novel Machine Learning-based approach for the detection of SSH botnet infection." Future Generation Computer Systems 115 (2021): 387-396.
24. Kumar, P. Ravi, Au Thien Wan, and Wida Susanty Haji Suhaili. "Exploring data security and privacy issues in Internet of Things based on five-layer architecture." International journal of communication networks and information security 12.1 (2020): 108-121.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:55 Issue:04:2022
DOI 10.17605/OSF.IO/UQ3D6

25. Carvalho, Kevin, and Jorge Granjal. "Security and Privacy for Mobile IoT Applications Using Blockchain." Sensors 21.17 (2021): 5931.
26. Feng, Bo, et al. "Multistage and elastic spam detection in mobile social networks through deep learning." IEEE Network 32.4 (2018): 15-21.
27. Feng, Pengbin, et al. "A novel dynamic Android malware detection system with ensemble learning." IEEE Access 6 (2018): 30996-31011.
28. Wang, Wei, et al. "DroidEnsemble: Detecting Android malicious applications with ensemble of string and structural static features." IEEE Access 6 (2018): 31798-31807.
29. Prabavathy, S., K. Sundarakantham, and S. Mercy Shalinie. "Design of cognitive fog computing for intrusion detection in Internet of Things." Journal of Communications and Networks 20.3 (2018): 291-298.