Multimedia Tools and Applications (2024) 83:58915–58934
https://doi.org/10.1007/s11042-023-17894-2

Anomaly detection in WSN IoT (Internet of Things) environment through a consensus-based anomaly detection approach

Anitha C L[1] **· R. Sumathi**[1]

Received: 19 August 2022 / Revised: 2 November 2023 / Accepted: 13 December 2023 /
Published online: 28 December 2023

Abstract
The most essential part of any IoT (Internet of Things) model is the wireless network sensors (WSN). The application of these networks combined with the latest technologies relating to IoT provides fast, economical as well as flexible applications. Wireless sensor networks have various applications for IoT where devices combined with the sensors are used for data collection from various environments as well as monitoring of these environments. These networks are highly prone to attacks considering their characteristic nature, which includes self-organization, a topology that is dynamic, large-scale, and constrained on resources. Various models have been proposed for the detection of attacks in these Wireless sensor networks. Although, the recent survey studies on the attacks in this network aims at the methodologies for detecting only one to two kinds of attacks as well as have the absence of performance analysis in detail. This research work proposes a CSAD (consensus-based novel anomaly detection) approach in three steps; first step; each step includes a novel algorithm. A novel distributed algorithm is proposed to classify the anomaly and normal data packets. In the second step level based approach is used for decision implementation to identify the anomaly; also it is responsible for efficient packet transmission. The third step includes discarding the anomaly Moreover, the proposed model is evaluated by inducing the different malicious nodes, and an anomaly detected is observed. Further comparison with the existing model is carried out based on the classified and misclassified packet; through the comparative analysis, it is observed that the Consensus-AD (Anomaly Detection) approach simply outperforms the existing model. A comparative analysis is carried out considering the throughput for model efficiency. Moreover, comparative analysis shows that the proposed model outperforms the existing anomaly detection protocol. The existing model observes a throughput of 80.99% whereas the CSAD model observes a throughput of 81.81%.

Keywords Anomaly detection · WSN (Wireless Sensor Network) · IoT (Internet of things) · CSAD (consensus-based novel anomaly detection) · Data packets

Extended author information available on the last page of the article

1 Introduction

The Internet of Things is mainly utilized for the gathering of real-world information and transmitting it to the cloud. WSN (Wireless Sensor Network) is an essential part of IoT (Internet of things) wherein various applications of IoT are used for the monitoring of the surroundings as well as recording its constraints and conditions [1], these include smart city [2], smart home, healthcare intelligence and warning for disaster. Figure 1 presents the typical WSN architecture comprising BS (Base Station), Cluster Head, and Nodes; Data is sensed through the sensor nodes, sent to BS, and further accessed through the clouds.

Although, these WSNs are prone to various attacks considering their characteristic nature, which includes self-organization, a topology that is dynamic, large-scale, and constrained on resources. The attacks on the anomalies of the network result from the data that is gathered in the network. The information collected has a high value and is utilized for the detection of attacks on the network [3]. The data used for this purpose is defined as data related to security, termed security data as they are used in anomaly detection and to detect intrusions, threats, and attacks in the security. The generation of security data occurs in different types of applications relating to WSN that include healthcare intelligence, smart home, smart city, etc. Considering various methods of detection, some of the security data is normally sensory data that include indicator strength of a received signal, messages of acknowledgment, etc. whereas the other type of security data is special data that include fingerprints, which are required to be extracted particularly.

The detection of attacks supports the defense of security to resist intrusions as well as threats to the security of WSNs. Therefore, it has an essential role in the WSN being secure. The methods of attack detection that are existing and the data relating to security are essential for knowing the present state of the art for detecting the domain of attack as well as further research for the measurement of security in WSN. Although, there exist few studies in the literature review where the detection techniques of the attacks are mainstream and are relevant to the gathering of data security and analysis of WSN [4Ips]. The restrictions in the resource capability include less speed of computation, low capacity of memory, energy limitations, and communication bandwidth restrictions, which have reduced the

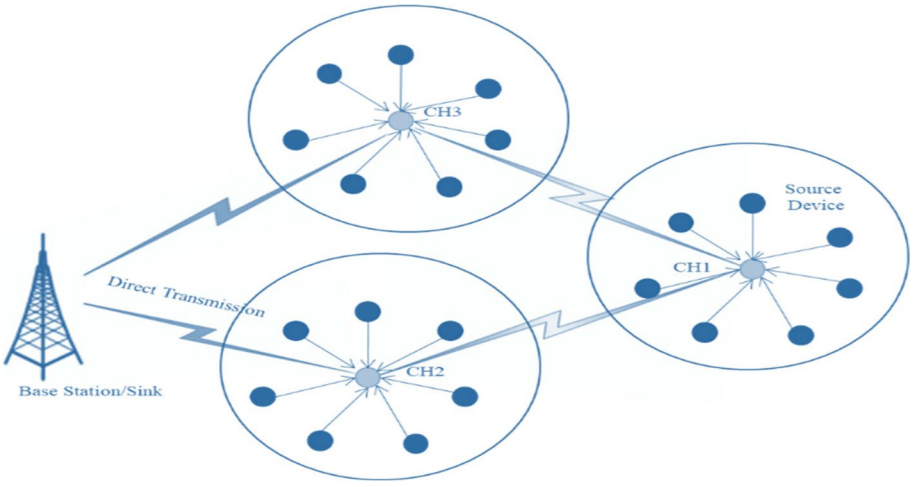

Fig. 1 WSN architecture

efficiency of these sensors being used. Further, the self-organization of the nodes in the WSN is highly vulnerable due to failure in communication because the transmission paradigm is unreliable.

Hence, these networks are mostly exposed either to random faults or to attacks relating to cyber that lead to sensory data anomalies. Practically, a higher number of anomalous events that tamper sensor data are of six types that involve degradation, spike, precision, drift, outlier, offset, and stuck-at [5]. These types are further grouped as anomalies of short-term including degradation, precision, spike, and outlier anomalies, and anomalies of long-term that include drift, stuck-at as well as offset anomalies [6].

- The detection of attacks as well as methods relating to the measure of security has been adapted, although the sink trust is missing.
- The literature is missing a method that is holistic for the detection of the attacks that are mainstream in WSN, which is extremely necessary for the measure of security.
- The parameters such as efficiency, accuracy as well as effectiveness are to be considered due to the node energy is limited and the resources in the WSN being constrained.
- Support for scalability is lacking in the methodologies that are existing. The attacks could occur on a huge scale in the network.
- Trust in the data and the quality of data are not considered proper for the methodologies that are existing. The process of collecting big data from various sources having varying quality for the detection of possible attacks on security measures that are trustworthy has been an issue.
- A dynamic setting threshold for the detection of the attack has to be done but it is not researched thoroughly so far.

The aggregation of data along with the protection of privacy as well as security is not researched perfectly and thoroughly yet. The aggregation of data is a method used to minimize the communication of the nodes and the energy of the node for the secure transmission of data. Although, the security methods have had negative impacts on the efficiency of energy for data aggregation techniques. Techniques need to be designed that satisfy the requirements of security wherein the overhead communication is decreased [7–9]. The various WSN layers may have a combination of more than a single attack that has been predefined that are launched in the stack of the protocol [10–13]. Some of these attacks are Denial service (DoS), jamming, and MITM (man-in-the-middle) attacks.

1. A jamming attack involves the physical layer having an initial jamming attack and the data link layer having an intelligent jamming attack.
2. The Denial-of-Service attack prevents the users that are legitimate from gaining access to the services and the information and consumes the resource of the nodes. The attack leads to congestion in the network while considering the physical layer. The basic attack of jamming is a type of DoS. The well-known protocols of communication are violated by the attacker considering the data link layer and the messages are transmitted continuously for the generation of collisions as well as lead to the retransmission of packets and loss of energy. Attacks such as collision, denial-of-sleeping, and intelligent jamming are categorized as attacks of DoS. Considering the network layer, the dishonest nodes do not route the required information or transmit the wrong information to the destination nodes. The attacks that belong to DoS include replay, Sybil, wormhole, grey hole, flooding, spoofing, and sinkhole attack. In the transport layer, the sensory nodes

are more prone to attacks of flooding as well as attack de-synchronization used for the generation of large counts of connections as well as requests of retransmissions and high consumption of energy.
3. The attack of man-in-the-middle is where an attacker eavesdrops secretly as well as changes the messages between the two nodes without any prior knowledge. The attack of eavesdropping for the physical layer as well as the attack of replay for the network layer come into the category of MITM attack.

Passive attacks cannot be easily sensed due to the missing emissions of radio that the adversaries produce. Privacy is breached when there are passive threats. The attackers are camouflaged during the passive attacks as well as disconnect the channel of communication for the gathering of data [14]. Destruction/tampering of nodes malfunction of nodes, eavesdropping, analysis of traffic, and disruption of nodes is classified as passive attacks.

The focus of the intrusion prevention system (IPS) is on monitoring and detecting network anomalies. The main distinction between IDS (Intrusion Detection System) and IPS is used in the prevention of attacks, whereas only the detection of the attacks is done by the IDS. IPS causes alarms to be raised during the detection of an anomaly, packets are dropped, connection resetting is performed or dishonest traffic is blocked using an IP address or URL (Uniform Resource Locator) that is malicious.

1.1 Motivation and contribution

Observations about sensory nodes show that they are normally related to the domains of time. The correlation is combined with the characteristics of WSNs because the further predictions that are observed are based on prior values. This largely motivates the algorithm proposed in this work, which is local detection. Further, considering WSNs, where normally deployed densely, each node is close to the other and the collaboration is facilitated for the nodes for a distinction among the events that are malicious as well as important emergency events. Hence, this research adopts the consensus approach for anomaly detection; further contribution of the research is given below points:

1. This research work aims to detect the anomaly that exists in the network through CAD (consensu anomaly detection); CSAD (consensus-based novel anomaly detection) is a lightweight security model with several phases. The initial phase includes designing an efficient energy utilization model along with setting the range of value.
2. The second phase includes the secure and optimal transmission of packets through a novel algorithm and the third phase includes the classification of data packets into normal and anomaly.
3. The last phase includes the identification of anomaly and discarding it; Evaluation is carried out considering the three distinctive malicious nodes i.e. 5, 10, and 15. The evaluation suggests major improvisation over the existing model.

This research is organized as follows: The first section starts with the background of WSN and its rising security concern; further, this section focuses on the different security aspects and the need for an anomaly detection approach from a security perspective. The second section highlights the different existing models for anomaly detection and their shortcomings. The third section presents the proposed methodology along with mathematical modeling and algorithm; performance evaluation is carried out in the fourth section.

2 Related work

IoT concerning WSN is prone to various attacks that lead to damage on a critical scale in the security of the network. The attacks that are related to the security of WSN are split into two classes: passive attacks and active attacks. These passive attacks are also classified as network interruptions, failure of the server, eavesdropping, analysis of traffic, and degradation of the network. In active attacks, the attacker gives up the activities and the roles of the network that is targeted. The main motive of any attacker is to create evident damage such that the security of the network cannot detect easily. Attacks such as flooding, DoS, black hole, jamming, Sybil, sinkhole, and wormhole are classified as active attacks.

The necessary and required surveys as well as the categorization of the issues about security are published in [7–9, 15]. In paper [16], an IDS (Intrusion Detection System) is proposed for various types of attacks that are malicious to be detected in an environment of IoT. The proposed method uses two methods for the reduction of dimensions as well as decreasing the count of the features that are to be utilized. The complexity of this is reduced by the use of a component principle and analysis for linear discrimination. Furthermore, two types of methodologies are used, namely, naive Bayes and KNN (k-nearest neighbors) for the activities that are malicious to be detected. In the paper [17], a detection system for intrusion based on blockchain-driven signature collaboration for IoT networks. In this proposed work, any intruder utilizes signatures or rules for the detection of malicious activities. Various other nodes share this data in the network for updating the database as well as improvement of the rate of detection. Although, there are possibilities of attacks within the network that increases due to the node inside being malicious or fake signatures that reduce the performance of the combined IDS. Hence, to solve this a methodology based on blockchain is proposed that is used commonly for databases that are distributed for intrusion detection.

In the paper [18], a detection system based on the lightweight host is proposed to be utilized in an IoT network HADES-IoT (host-based anomaly detection system for IoT devices). The method is based on devices, and proactive techniques are deployed in Linux. The distinct characteristic of this technique is that it is possible to load this in the kernel of the OS (Operating System). Hence, its usability increases in Linux for the installation of HADES-IoT. In paper [19], a client-based system is proposed for IDS where the detection of anomalies is performed which is termed as E-Spion. There are three various layers of security for which at every level the security increases. Although, there are some limitations to this method as the security level increases leading to an increase in overhead. Considering the first module, which there occurs a comparison of the system with the name of the processes that are running with their respective IDs, wherein it is done during the phase of learning for the malicious process to be separated. In the next module, the machine learning classifier is trained by the logs that are generated at the phase of learning which later monitored using parameters is.

The detection of anomalies and events in IoT devices is proposed using a multivariate long short-term memory (LSTM) autoencoder. In order to reduce power consumption and increase device longevity, the suggested approach also incorporates smart inference, a game-theoretical approach that dynamically alters the period of detection based on the stability of the data. Although this technology is pricey, it works well since some machine learning techniques are used at the node level. In the paper [20], an IDS is proposed in two steps for a secure system. In the initial step, a random model of the neural network is used for IDS based on the anomaly. In the next step, a new system for tag is proposed for the

design of the system, which is connected, to the memory of the model. The method of tag-check is utilized for the detection of anomalies in the network.

For the purpose of detecting anomalies, an integrated model of a convolutional neural network (CNN) and recurrent autoencoder is put forth. The classification performance cannot be improved by a simple CNN and autoencoder combination, especially for time series. In the paper [21], a technique for the identification of a new node is proposed and termed node identification over spoof attack (NISA). A synchronization framework is used considering reverse time, in which the clock skews of the sensory nodes are calculated at the WSN head. The radio information link that is correlated spatially is used for achieving the identification of nodes as well as the detection of attacks. Furthermore, a centralized approach is provided and NISA is distributed that covers two various scenarios of the multi and single hop, where there is used of multi-output and single-input CNN (Convolutional Neural Networks).

In the paper [22], a novel machine learning method is proposed that is supervised based on IDS that is used for the detection of many attacks. The proposed paper includes the analysis of all the design phases of IDS that start from the gathering of data until the analysis of feature engineering as well as trained models being built. The results that are produced from the experiments show the IDS being proposed is capable of detecting four various kinds of attacks seen by the models of machine learning at the phase of training. Table 1 show the comparison of literature survey.

3 Proposed methodology

Existing solutions for security issues discussed in the earlier section tend to detect any kind of malicious program; moreover, existing models are considered in two categories first is to protect and restrict the data through modifying the information flow another is designing access control. Moreover, this requires the modification of the application or the network. Hence, our research adopts a lightweight consensus protocol to monitor and detect the anomaly and further create a security model for discarding the malicious packets.

3.1 System model and preliminary analysis

CSAD model considers a network where sensor nodes are organized into clusters; clustering is carried out through the previously developed approach. Further, the proposed network has a large count of sensory nodes M that are placed in hospitals. For every node, there exists a range of transmission that is denoted as $Q_{maximum}$ and the implementation of the network is then given as a graph without direction as $F = (U, D)$ in which the group of nodes is given as $U = \{u_1, u_2, \ldots, u_M\}$ and D is defined as the distance between two nodes such that it is less than $Q_{maximum}$. We consider these two nodes to be u_a *and* u_b, where the link of data to be transmitted for any two different nodes has the edge of the node u_a, u_b that belong to D but do not have similar capacity. The capacity for it to transmit packets is denoted as B_{u_a,u_b} from the nodes $u_a tou_b$. Every node transmits the data packets using a communication that is multi-hop to the base station, the first phase of the model that is proposed is used for the consumption of energy. We assume that a data message that has δ bits is transmitted from one node to the other that is placed at a distance c from each other, therefore, the consumption of energy at this point is given by the equation below

Table 1 Literature survey comparison

Author	Methodology	Advantages	Research gap
[16]	Dimension reduction, Component analysis, Naive Bayes, KNN	Detection of IoT malicious activities, Complexity reduction	Lack of specific author, Details on performance, scalability, and comparison with existing methods required
[17]	Blockchain-driven signature collaboration	Collaboration for updating database, Improved detection rate	Evaluation of blockchain impact, Addressing malicious nodes and fake signatures, Performance evaluation
[18]	Host-based anomaly detection in IoT (HADES-IoT)	Lightweight host-based approach, Kernel-level integration in Linux	Performance evaluation, Scalability, Compatibility with non-Linux IoT devices
[19]	Client-based IDS, E-Spion	Multi-layer security, Machine learning-based anomaly detection	Overhead analysis, Comparison with existing IDS methods, Scalability, Real-world performance and implementation
[20]	Neural network-based IDS, Tag-check	Anomaly detection using a two-step approach, Tag-check method	Performance evaluation, Complexity and resource utilization, Real-world applicability
[21]	Node identification over spoof attack (NISA)	Synchronization framework, Multi-output/single-input CNN for node identification and attack detection	Evaluation of synchronization framework, Real-world testing, Scalability, Comparison with other node identification methods
[22]	Supervised IDS, Machine learning-based detection	Detection of multiple attack types, Comprehensive analysis of IDS design phases	Performance evaluation, Comparison with other IDS methods, Real-world applicability, Data gathering considerations

$$Energy_s = \begin{cases} \delta Energy_{ele} + \delta e_f c^2 \text{ when } c \text{ lesser than or equal to } c_0 \\ \delta Energy_{ele} + \delta e_{amp} c^2 \text{ when } c \text{ greater than } c_0 \end{cases} \quad (1)$$

In the above equation, the loss caused due to transmission is denoted as $Energy_{ele}$, the energy for power is given as e_f and the energy for amplification is denoted as e_{amp}, also the distance threshold is denoted as c_0. When a message of δ bits is transmitted there occurs consumption of energy, which is based on the dissipation in the network that occurs for this transmission which is given as,

$$Energy_{con} = \delta Energy_{ele} \quad (2)$$

Consider the rate at which the data packet is sent by w_a to a node u_a such that if the rate is found to be null then the node is not working. Hence, we consider a graph $F_w = (U, D, W)$ for a graph based on an IoT sensor network for which $W = (w_1, w_2, \ldots, w_m)^S$. where the rate at which the data is sent is denoted by W that varies concerning time. Assuming the traffic in the IoT sensor network is large enough to surpass the bandwidth of the network, there are packets of information that cluster up that lead to anomalies. When the packets of information are sent from one node to the other, u_a *to* u_b respectively, in which case the sending rate of the sending node u_a is greater than that of the receiving node u_b. This leads to the clustering and gathering up of the cache nodes in the network, which leads to anomalies.

The main motive of this proposed work is to effectively detect the anomalies in the IoT network. Other parameters have to be estimated for efficient detection and working of the network. The consumption of energy during this process, the packet transmission as well as the discarding of the anomalies are the focus of this paper base on an IoT network about healthcare systems. This increases the lifetime of the network and optimizes the network. The lifetime of the network is split into various stages such as $[R_0, R_1, R_2, \ldots, R_{q-1}, R_q]$, the working of this is such that we consider an initial node to work and end its working by the end of the stage R_0 and by the time the stage R_q is reached the node is completely out of process. The traffic in the network is also considered while framing the objective of the proposed work to detect anomalies. the parameters of traffic and consumption of energy are considered and evaluated at every stage of the network. Therefore, the anomalies can be detected even during high traffic conditions.

3.2 Proposed anomaly detection

The sole objective in this section of the proposed work is to detect the anomalies of the IoT network considering healthcare systems for WSNs that expand the lifetime of the network. To achieve this the methodology, consider a few aspects that include packet transmission, consumption of energy, traffic, and discarding of anomalies. The traffic constraint is of two kinds where the traffic is either non-sensitive or sensitive. There are three phases to this proposed anomaly detection methodology: the initial phase, packet transmission phase, and lastly, identification of anomaly and discarding phase. The IoT applications have various stages for which each stage has different importance based on the IoT devices used in the network; the proposed work senses data for various nodes based on priority. Before the transmission of data packets, every node that is involved in this process receives a priority status by considering the priority of the data that is being transmitted. The two main nodes that are involved in this process are the sender and receiver nodes, whereas the intermediate nodes in this process also have to consider the priority of the data being transmitted.

Figure 2 shows the proposed workflow, it comprises six blocks; the first block designs the system modeling, second block deals with the initialization phase where the anomaly detection is initialized. The third block includes the efficient data packet transmission and the fourth block includes the data packets classification as anomaly or normal data. Further, the anomaly is identified and it is discarded.

3.2.1 Initial stage

This phase runs only once in the network when the network process starts. A single hop process is detected by every node for its surrounding nodes after which the nodes that are being deployed are split into various stages. In the beginning, the stage value is given $S_{val} = 1$ and a packet request is transmitted to the nodes in the range $Q_{maximum.}$ The packets have their ID, S_{val} as well as the data regarding the location. When the node receives the packet with the stage value then it increases its value by one than the initial stage value, $(u_a) = S_{val} + 1$. All the nodes of the network within the value of twice the maximum range increase their initial stage value by one higher and give the value of their initial stage as the parent stage. This process is explained in detail by the algorithm given below in Table 2.

3.2.2 Packet transmission phase through a distributed model

In this stage, when the initial stage node gets a packet, the initial node distributes the requirements to the nodes that are deployed to match the required needs. In these types of IoT applications, the kind of information that is being used is highly important while the data is collected. Sometimes, a few parameters have information and data that is sensitive. Considering the packet transmission phase, the initial node identifies the stage as well as the location of the nodes from the data and evaluates the packet lifetime in the network. The life of the packet being transmitted can extend the transmission reliability as well as the overhead of the packet is decreased. This evaluation is performed as given below

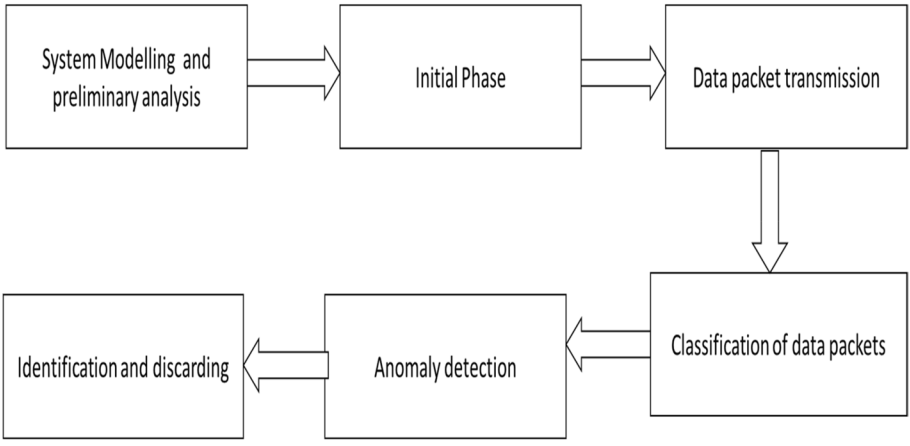

Fig. 2 Proposed workflow

Table 2 Anomaly detection

Steps	Initial Stage Detection Algorithm
Step 1	Input: sensory nodes are deployed Output: an initial stage value has been set to all the nodes
Step 2	$S_{val} = 1$//The initial stage value is set to 1
Step 3	A packet *STAGE* is sent within the $Q_{maximum.}$
Step 4	For each node u_M
Step 5	If the distance of the initial stage to the node u_m is $Q_{maximum.}$
Step 6	Then, S $=S_{val} + 1$
Step 7	And, the value of their initial stage as the parent stage $Parent = S_{val}$
Step 8	End if;
Step 9	End for;
Step 10	For the range 2 $Q_{maximum}$ the node u_M sends a packet *MODIFY_STAGE*
Step 11	If node u_a receives the packet
Step 12	And $S(u_a) > S(u_M)$ then
Step 13	$S(u_a) = S(u_M) + 1$
Step 14	$Parent(u_a) = u_m$ $Child(u_a) = u_m$
Step 15	Else discard packet
Step 16	End if;
Step 17	Every node u_m has and sends ID, S_{val} as well as the data regarding the location
Step 18	End for;

$$S_{Packet\_lifespan} = \sum_{1}^{dest(Stage)} (K_K + K_Q)O_k \tag{3}$$

In the above equation, the destination node value is given as (*Stage*), and the delay in the transmission is denoted as K_K, whereas the delay in receiving the packet is given as K_Q and the delay in processing is given as O_k. Table 3 displays the secure and efficient packet transmission.

In the IoT network, the initial node may receive multiple packets simultaneously for different of the same IoT devices. If this response to the packets has to be sent simultaneously then the packets face traffic and this causes anomalies in the network. Hence, the proposed work is performed based on the priority of the data. In addition, the request time for the response packets is considered while these data packets are being transmitted.

3.2.3 Identification of anomaly and discarding anomaly

After the phase of packet transmission, the final node sends a packet that updates the data to the initial node or the base station using an energy link that proves efficient in the network. When a high-priority data packet is sent by a node to the initial node, it is performed as per the specifications. Packets should consist of the necessary parametric values for the initial nodes to provide the required response packets. In this phase, the data about high-priority information is sent to the initial node because the initial node introduces the packet that consists of the data relating to the anomalies that occur which is sent to the *Parent*

Table 3 Secure and efficient packet transmission

Steps	Packet Transmission Algorithm
Step 1	Input: Packets from initial devices Output: Response to these Packets
Step 2	The initial node receives packets from the devices
Step 3	Initialize response to these packets as per priority
Step 4	If the packets required are unavailable; then
Step 5	The stage is identified by the initial node and the ID of the node u_a
Step 6	The packet lifespan is evaluated using Eq. (3)
Step 7	The life of the packet is set
Step 8	Send the *Response_Packet* with single-hop
Step 9	For every node u_M
Step 10	If the u_M node ID= u_a node ID
Step 11	The response packet is sent to the initial node
Step 12	End if;
Step 13	If $S(u_M) \neq S(u_a)$ then
Step 14	The response packets are sent to the *Child* nodes
Step 15	End if;
Step 16	If the lifespan of the packet $S_{Packet\_lifespan} = 0$ and $S(u_M) < S(u_a)$
Step 17	The packet is discarded
Step 18	End if;
Step 19	End for;

node at the nearest using the energy-based transmission link. A node may consist of multiple *Parents* as given in the initial stage, which results in many paths of transmission to the initial node.

We consider, a set of nodes $\{u_1, u_2, u_3, \ldots, u_o\}$ as the nodes that belong to the $Parent\{u_j\}$. While we assume that a high-priority packet has to be transmitted, the evaluation of the energy of *Parent* node is done by the node u_o by the equation given below

$$\mu(u_j) = \frac{\sum_{h=1}^{O} Energy_p(u_M)}{O} \tag{4}$$

Considering the above equation, the present energy condition is given as $Energy_p$, the number of nodes in $Parent(u_j)$ is given as O. The energy of the *Parent* is given as $R = \{r_1, r_2, r_3, \ldots, r_O\}$ which is more than or equal to the value of $\mu(u_j)$. The node (u_j) transmits the packet to the *Parent*, in this part of the process there are most chances for anomalies to occur. Therefore, an algorithm needs to be proposed where the anomalies are identified/detected as well as discarded. The details of the algorithm for this process have been described in detail below in Table 4.

3.3 Consensus leverage and data aggregation

In this part of the proposed work, we focus on the consensus detection of the anomalies and the management of the occurrence of these anomalies. The nodes that are deployed

Table 4 discarding the packet

Steps	Anomaly Identification and Discarding Anomaly
Step 1	Input: Anomaly identification by the nodes Output: Detected Anomaly sent to initial node or base station
Step 2	For every node u_O
Step 3	For every node $u_M \in Parent(u_O)$
Step 4	$variable = variable + Energy_p(u_M)$
Step 5	End for;
Step 6	$\mu(u_j) = \frac{variable}{Parent(u_O)}$
Step 7	For every node $u_M \in Parent(u_O)$
Step 8	If $Energy_p(u_M) = \mu(u_j)$
Step 9	$R = R \cup u_1$
Step 10	End if;
Step 11	End for;
Step 12	For every node $u_M \in R$
Step 13	If $maximumisgreaterthanEnergy_p(u_O)$ and $minimum(distance(u_k, u_O))$
Step 14	$maximum = Energy_p(u_O)$
Step 15	The node u_k is detected as an anomaly
Step 16	End if;
Step 17	End for;

in the IoT network, can avoid anomalies by selecting different transmission paths of the data packets, though these anomalies cannot be completely avoided, they can at least be decreased. The few unavoidable anomalies caused are detected based on consensus. The proposed technique has a classifier that includes various packets and channels of transmission. It is established in the above sections that the transmission of packets occurs based on priority. These classifications of priority are in three ways: Packets of high priority, packets of low priority, and control management packets. The type of the packet is mentioned for every packet in its header. It is the work of the classifier to identify and categorize the type of packet and place them in various queues depending on their priority classifications. A scheduling system based on a priority queue is proposed in this paper for various types of packets. If the priority queue has medium and low priority data in it, which is being processed and high priority data enters the queue. The transmission of the data packets in the queue stops and the transmission of the high-priority packets begins. Once all the packets of high priority are transmitted, the priority queue restarts its transmission. An anomaly is detected when the rate at which the packets are received exceeds the rate of transmission of the packets.

Here, packets are not transmitted to the *Child* nodes leading to an issue of missing packets or loss of packets. To resolve this, a different transmission route is selected by the node such that the anomalies reach a threshold value. At this time, the *Child* node sends the packet to the node which is responded to by an acknowledge packet. The change in the path is decided by the *Parent* node when the threshold value is reached when the queue is full by 95% of packets. Further, observed data is aggregated through the below equation

$$A_{\vartheta}^{o} = \sum_{l \text{ belongs to } X_o} A_l^o \tag{5}$$

Further, $A^o = [A_1^o, A_2^o, A_3^o, \ldots .., A_l^o]$ belongs to U^Q which denotes the data packet observed consensus is leveraged and aggregated data is given as

$$V_o(W_o) = Q\left(\sum_{l=1}^{Q} Q_l\right)^{-1} \tag{6}$$

4 Performance evaluation

The application of IoT has brought unprecedented convenience for human beings; however, attackers might use configuration vulnerabilities of the device to control the services, steal the data, hijack devices, and illegally operates the device. Moreover, these restrictions have not only led to major security risks but also possesses a major challenge to infrastructure service. Moreover, these issues can be solved by finding any abnormal behavior like anomaly detection that can detect any malicious activity with the data packets. Moreover, model simulation is carried out using the sensoria simulator.

System configuration includes the 2 TB of the hard disk along with 8 GB of RAM packed with 2 GB graphics; performance evaluation includes anomaly detection by inducing the different compromised nodes. Evaluation is carried out on anomaly detection; further evaluation is carried out by comparing several anomalies detected.

4.1 Energy utilization

Figure 3 shows the energy utilization of the proposed model by varying the number of different malicious Nodes concerning anomaly; in the case of 5, 10, and 15 malicious nodes, energy utilization by the CSAD model is 3.63, 3.58, and 3.53 millijoule (mJ) respectively.

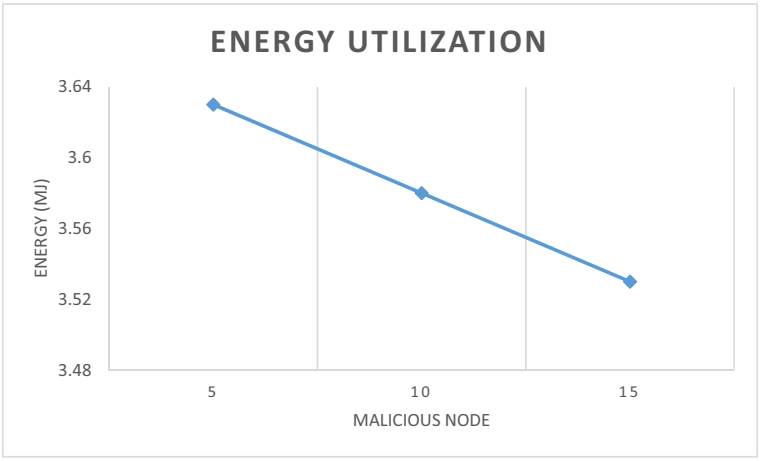

Fig. 3 Energy utilization

Through the below figure it is observed that with an increase in malicious nodes, energy utilization is optimal.

4.2 Anomaly detection

In general, anomaly shows the abnormal behavior of the sensor; evaluation is carried out considering the different malicious nodes as 5, 10, and 15 nodes. Figure 4 shows the anomaly packet detected for several malicious nodes induced is five; here the sensed range is 25 to 30, and anything above that is considered as the anomaly.

Similarly, Fig. 5 shows the anomaly detected for 10 malicious nodes; it is observed that with an increase in the number of malicious programs, there is an increase in the number of anomalies.

Figure 6 shows the anomaly detected for 100 nodes where 15 nodes are induced as malicious nodes.

4.3 Correct packet identification and misidentification

Anomaly detection is an essential process concerning security; however, due to various reasons mentioned earlier, there is the possibility of detection of the normal data as the anomaly, which is another major concern. In this section, a comparison of existing and proposed protocols is carried out to analyze the number of correct identification of anomaly and normal data.

Figure 7 shows the comparison of existing and CSAD model packet transmission over the simulation. In the below graph, 1 indicates a correctly classified packet and 0 indicates misclassified. It is observed that out of a total number of 100 packets, the proposed model can identify each packet correctly whereas the existing model fails to identify 9 packets.

Figure 8 presents the identification and misidentification comparison when 10 anomaly nodes are induced, Proposed Model classifies all the models correctly into normal data

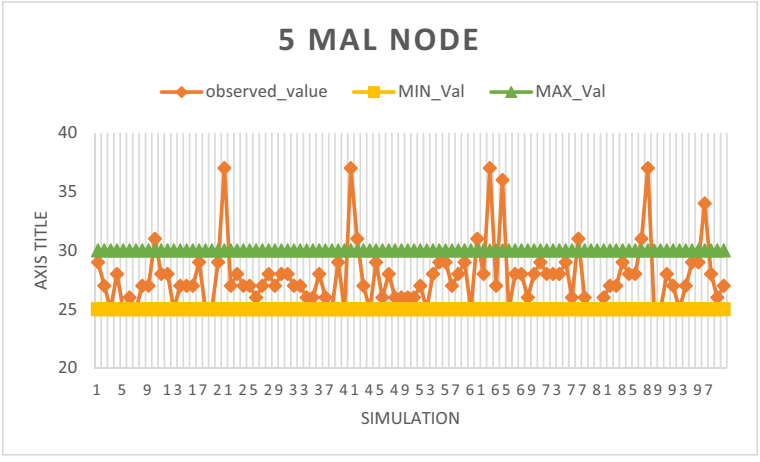

Fig. 4 5 Malicious nodes

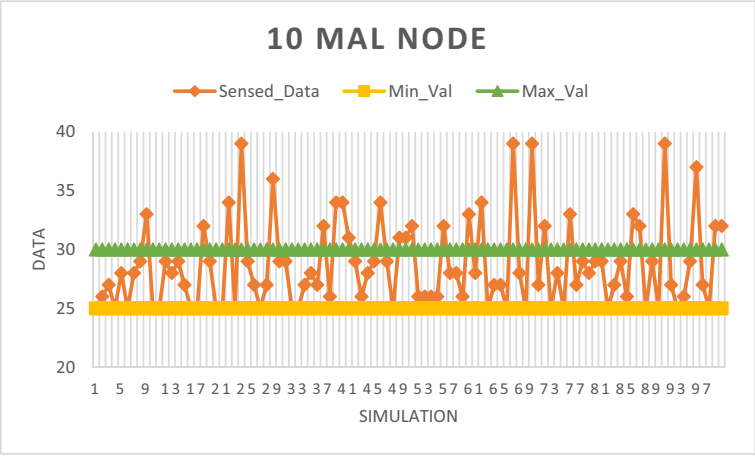

Fig. 5 10 Malicious nodes

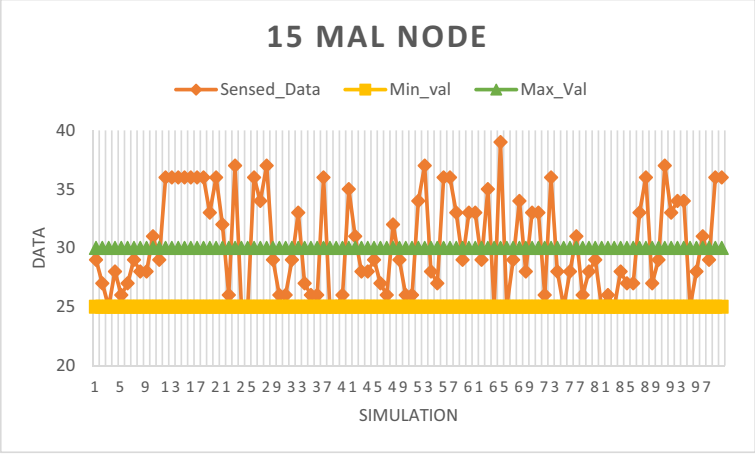

Fig. 6 15 Malicious nodes

and anomaly. Zero indicates misidentified packets whereas 1 indicates correctly identified packets.

Figure 9 shows the Packet identification and misidentification of packets when 15 nodes are induced, It is observed that the CSAD model classifies every packet and anomaly whereas the existing model misclassifies 35 packets.

4.4 Throughput and comparative analysis

This section of the research performs the comparative analysis based on calculated throughput as given in Fig. 10, which shows the improvisation of the proposed model over the existing one by varying the malicious nodes as 5, 10, and 15. Throughput is defined as

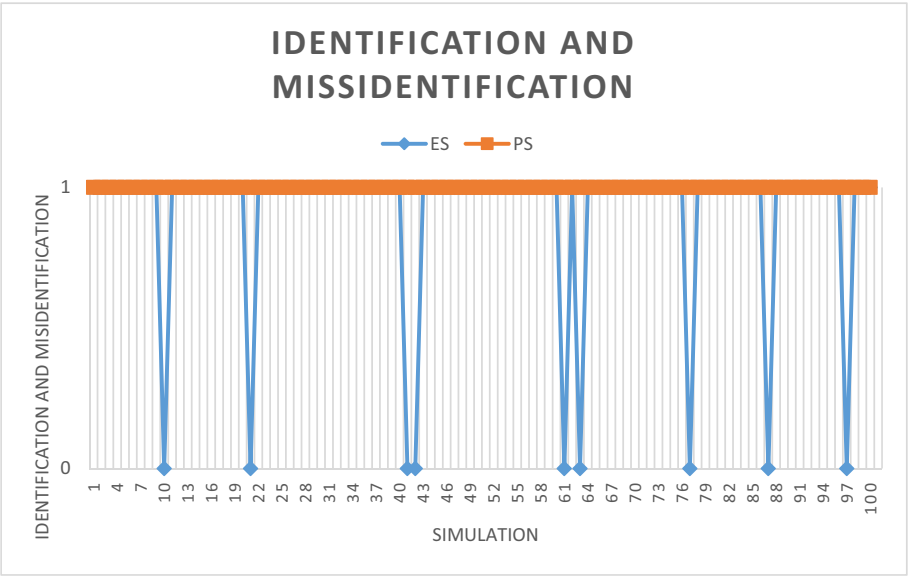

Fig. 7 Identification and misidentification

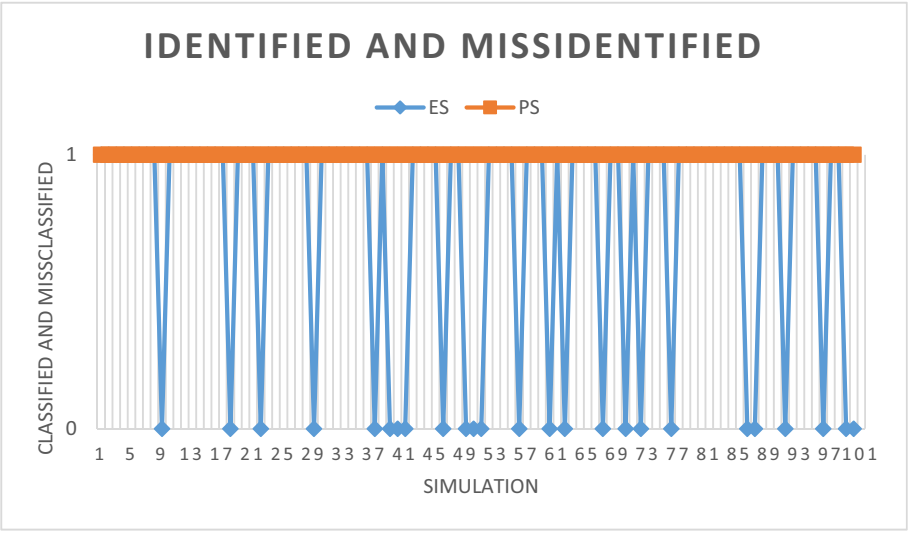

Fig. 8 Packet Identification and misidentification with 10 malicious nodes induced

the total amount of work done actually in a given time; it is considered one of the major metrics for the efficiency evaluation of model.

In the case of 5 malicious nodes, the existing model observes a throughput of 80.99% whereas the CSAD model observes a throughput of 81.81%. Similarly, for 10 malicious nodes, the existing model observes a throughput of 55.5% whereas the proposed model observes a throughput of 96.15%. At last, considering the 15 malicious nodes, the existing

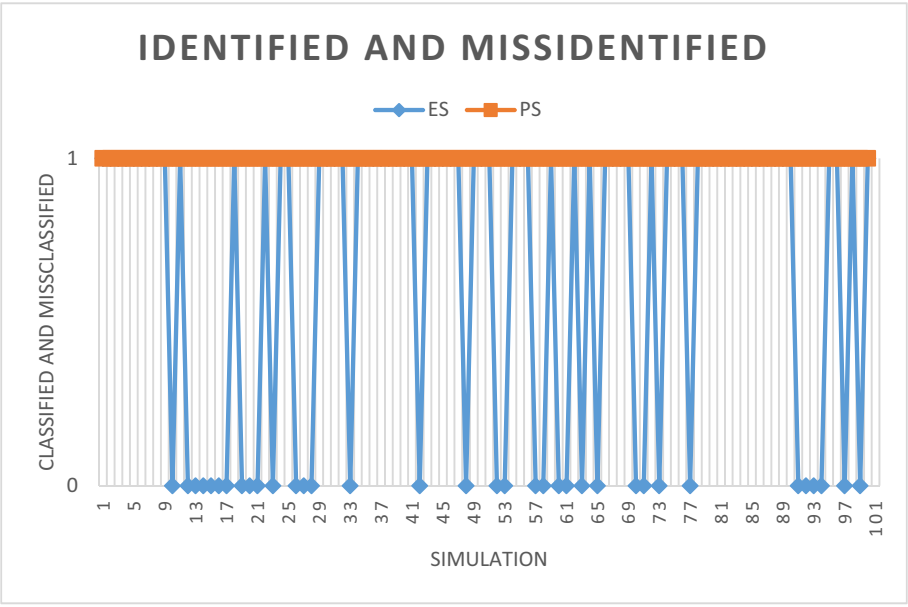

Fig. 9 Packet identification and misidentification

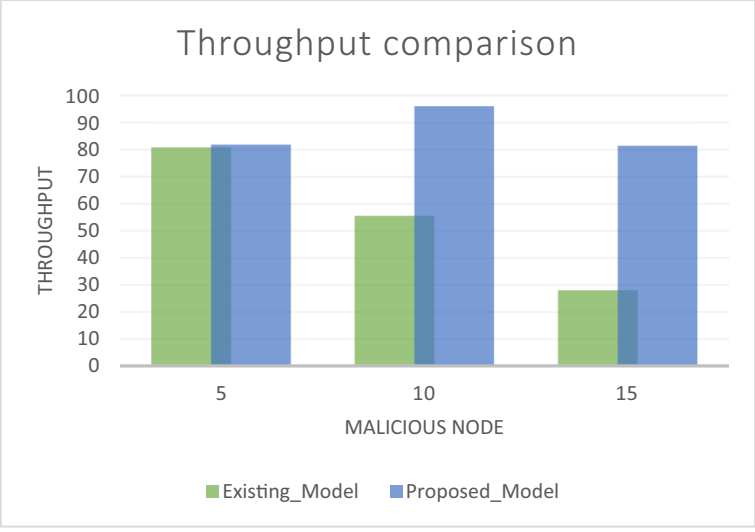

Fig. 10 Throughput comparison

model observes a throughput of 27.95% whereas the proposed model observes a throughput of 81.39%.

As given in above Fig. 10, for 5 malicious nodes, the proposed model improves its throughput by 0.82%; further proposed model observes 40.62% improvisation over the existing model. For 15 malicious nodes, the proposed model observes an improvisation of

53.44%. Through the analysis, it is observed that with an increase in the number of malicious nodes, the performance of the existing model degrades whereas the proposed model either increases its throughput or remains steady.

5 Conclusion

WSN-based IoT security has been a recently emerging topic in the industry as well as academia; However, WSN faces several security challenges whether physical failure or any kind of malicious attack which modifies the data packets. Hence, anomaly detection is an effective way to avoid the risk and secure the paper. This research work develops a CSAD (consensus-based anomaly detection); a three-step novel approach designed for anomaly detection and discarding. Moreover, this approach adopts the DADD (Data Aggregation and Data Dissemination) as developed in earlier work; further, this three-step approach includes data packet classification, efficient and secure packet transmission through a distributed approach, anomaly detection, and discarding the packets.

Model evaluation is carried out in several parts; the first detection of anomaly over inducing different malicious nodes is observed. Further evaluation is carried out by comparing the identification and misidentification of packets with the existing model. Energy utilization of the proposed model by varying the number of different malicious Nodes concerning anomaly; in the case of 5, 10, and 15 malicious nodes, energy utilization by the CSAD model is 3.63, 3.58, and 3.53 millijoule (mJ) respectively. In the case of 5 malicious nodes, the existing model observes a throughput of 80.99% whereas the CSAD model observes a throughput of 81.81%. At last, a comparative analysis is carried out considering the throughput for model efficiency. Moreover, comparative analysis shows that the proposed model outperforms the existing anomaly detection protocol. Although the CSAD model observes marginal improvisation, considering the WSN vulnerability other security aspects need to be considered such as the implementation of a blockchain-based approach.

Data availability No dataset is utilized in this research.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

1. Miao X, Liu Y, Zhao H, Li C (2019) Distributed online one-class support vector machine for anomaly detection over networks. IEEE Trans Cybern 49(4):1475–1488. https://doi.org/10.1109/TCYB.2018.2804940
2. Xie H, Yan Z, Yao Z, Atiquzzaman M (2019) Data collection for security measurement in wireless sensor networks: a survey. IEEE Internet Things J 6(2):2205–2224. https://doi.org/10.1109/JIOT.2018.2883403
3. Dang T-B, Le D-T, Nguyen T-D, Kim M, Choo H (2021) Monotone split and conquer for anomaly detection in IoT sensory data. IEEE Internet Things J 8(20):15468–15485. https://doi.org/10.1109/JIOT.2021.3073705
4. Jiang S, Zhao J, Xu X (2020) SLGBM: an intrusion detection mechanism for wireless sensor networks in smart environments. IEEE Access 8:169548–169558. https://doi.org/10.1109/ACCESS.2020.3024219

5. Islam K, Shen W, Wang X (2012) Wireless sensor network reliability and security in factory automation: a survey. IEEE Trans Syst Man Cybern Part C (Appl Rev) 42(6):1243-1256. https://doi.org/10.1109/TSMCC.2012.2205680
6. Abduvaliyev A, Pathan AK, Zhou J, Roman R, Wong W (2013) On the vital areas of intrusion detection systems in wireless sensor networks. IEEE Commun Surv Tutor 15(3):1223–1237. https://doi.org/10.1109/SURV.2012.121912.00006
7. Yao S, Li Z, Guan J, Liu Y (2020) Stochastic cost minimization mechanism based on identifier network for IoT security. IEEE Internet Things J 7(5):3923–3934. https://doi.org/10.1109/JIOT.2019.2961839
8. Desai SS, Nene MJ (2021) Multihop trust evaluation using memory integrity in wireless sensor networks. IEEE Trans Inf Forensics Secur 16:4092–4100. https://doi.org/10.1109/TIFS.2021.3101051
9. Laouira ML, Abdelli A, Othman JB, Kim H (2021) An efficient WSN based solution for border surveillance. IEEE Trans Sustain Comput 6(1):54–65. https://doi.org/10.1109/TSUSC.2019.2904855
10. Mohy-Eddine M, Guezzaz A, Benkirane S, Azrour M, Farhaoui Y (2023) An ensemble learning based intrusion detection model for industrial IoT security. Big Data Min Anal 6(3):273–287. https://doi.org/10.26599/BDMA.2022.9020032
11. Amaouche S, Guezzaz A, Benkirane S, Azrour M, Khattak SBA, Farman H, Nasralla MM (2023) FSCB-IDS: feature selection and minority class balancing for attacks detection in VANETs. Appl Sci 13:7488. https://doi.org/10.3390/app13137488
12. Mohy-eddine M, Guezzaz A, Benkirane S et al (2023) An efficient network intrusion detection model for IoT security using K-NN classifier and feature selection. Multimed Tools Appl 82:23615–23633. https://doi.org/10.1007/s11042-023-14795-2
13. Douiba M, Benkirane S, Guezzaz A et al (2023) An improved anomaly detection model for IoT security using decision tree and gradient boosting. J Supercomput 79:3392–3411. https://doi.org/10.1007/s11227-022-04783-y
14. Mohy-eddine M, Guezzaz A, Benkirane S et al (2023) An effective intrusion detection approach based on ensemble learning for IIoT edge computing. J Comput Virol Hack Tech 19:469–481. https://doi.org/10.1007/s11416-022-00456-9
15. Hazman C, Guezzaz A, Benkirane S, Azrour M (2022) lIDS-SIoEL: intrusion detection framework for IoT-based smart environments security using ensemble learning. Cluster Comput 26. https://doi.org/10.1007/s10586-022-03810-0
16. Pajouh HH, Javidan R, Khayami R, Dehghantanha A, Choo K-K-R (2019) A two–layer dimension reduction and two–tier classification model for anomaly–based intrusion detection in IoT backbone networks. IEEE Trans Emerg Topics Comput 7(2):314–323
17. Li W, Tug S, Meng W, Wang Y (2019) Designing collaborative blockchained signature-based intrusion detection in IoT environments. Future Gener Comput Syst 96:481–489
18. Breitenbacher D, Homoliak I, Aung YL, Tippenhauer NO, Elovici Y (2019) HADES-IoT: A practical host-based anomaly detection system for IoT devices. In: Proc. ACM Asia Conf. Comput. Commun. Secur., Auckland, New Zealand, pp 479–484
19. Mudgerikar A, Sharma P, Bertino E (2019) E-spion: A system-level intrusion detection system for iot devices. In: Proc. ACM Asia Conf. Comput. Commun. Secur., Auckland, New Zealand, pp 493–500
20. Abououf M, Mizouni R, Singh S, Otrok H, Damiani E (2022) Self-supervised online and lightweight anomaly and event detection for IoT devices. In: IEEE Internet Things J 9(24):25285–25299. https://doi.org/10.1109/JIOT.2022.3196049
21. Yin C, Zhang S, Wang J, Xiong NN (2022) Anomaly detection based on convolutional recurrent autoencoder for IoT time series. IEEE Trans Syst Man Cybern Syst 52(1):112–122. https://doi.org/10.1109/TSMC.2020.2968516
22. Sharma M, Elmiligi H, Gebali F (2021) A novel intrusion detection system for RPL-based cyber–physical systems. IEEE Can J Electr Comput Eng 44(2):246–252. https://doi.org/10.1109/ICJECE.2021.3053231

Authors and Affiliations

Anitha C L[1] · R. Sumathi[1]

✉ Anitha C L
clanitha@gmail.com

R. Sumathi
rsusit@gmail.com

[1] Department of Computer Science and Engineering, SIT, Tumkur, India