A Trust-Based Security Model to Detect Misbehaving Nodes in Internet of Things (IoT) Environment using Logistic Regression

K Prathapchandran¹, T Janani²
¹,² Department of Computer Applications, Karpagam Academy of Higher Education (Deemed to be University), Coimbatore-021, Tamilnadu, India
E-mail: kprathapchandran@gmail.com, janani0609@gmail.com

Abstract. Ensuring authentication in the Internet of Things (IoT) environment is a crucial task because of its unique characteristics which include sensing, intelligence, large scale, self-configuring, connectivity, heterogeneity, open and dynamic environment. Besides, every object in the IoT environment should trust other devices with no recommendation or prior knowledge for any network operations. Hence, those characteristics and blindness in communication make security violations in the form of various attacks. Therefore, a trust-based solution is necessary for ensuring security in the IoT environment. Trust is considered as a computational measure represented through a relationship between trustor and trustee, explained in a particular context valued through trust metrics and evaluated by a trust mechanism. The proposed logistic regression-based trust model provides an efficient way to identify and isolate the misbehaving nodes in the RPL (Routing Protocol for Low Power Lossy Networks) based IoT network. It is one of the popularly used routing protocols in IoT, that builds a path especially for the constrained nodes in IoT environments. However, it is vulnerable to many attacks. The proposed model classifies and predicts the node's behavior (trusted or malicious). This model uses the logistic regression model to predict the node's behavior based on the integrated trust value which is computed from the direct trust, reputation score, and experience trust. It is primarily designed to address the black hole attack in the IoT environment. The mathematical analysis shows the possibility of the proposed work and the simulation results show the proposed model is better than the existing similar work.

Keywords: Internet of Things (IoT), Security, RPL, Authentication, Trust, and Logistic Regression.

1. Introduction
The Internet of Things (IoT) is the outcome of the recent trends that interconnects all kinds of devices and systems through the internet [1]. It is a novel framework where physical things are incorporated into networks and provide progressive and smart services for people. The associated “things” like sensors, actuators, and mobile devices will sense, observe, and gather various information about people's social life. Also, this information can be accumulated, intermixed, processed, tested, and mined to take out the fruitful message to facilitate very smart and universal services. The IoT is an emerging and appealing posterity network framework and service paradigm [2].

IoT has been broadly used in many applications, from large-scale smart energy grids to small-scale human wearable devices [1]. Although IoT applications and devices have been rapidly developing,
cyber-attacks also have been increasing, and they pretend to be a more crucial threat to the security and privacy of the IoT environment [3]. IoT faces many challenges because of its special characteristics such as heterogeneity, intelligence, connectivity, dynamic nature, scalability, and sensing. One distinguishing characteristic of IoT is that it has resource-restricted devices also known as "things" [1]. These things have minimal capacity like power, storage capacity, processing capability, etc. It restricts them to achieve a progressive security solution [6]. The restriction of the interconnected IoT devices brings security issues and privacy challenges [1]. For example, a remote opponent would gain the patient's details by inserting medical devices on the patient or get confidential information from the smart car. It may not only create financial loss to a person but also a risk in human life [3]. Most of the IoT applications use wireless communication which creates the system more susceptible to various attacks like identification threat, data packets eavesdropping, tampering the data packets, and other security-related problems [2]. The IoT can be considered as a union of non-homogeneous networks that fetch identical security issues available in the wireless sensor networks, mobile communications, and the computer network and also a few distinct and emphasize problems, such as privacy challenges, non-homogeneous device's authentication, access privilege challenges and protected routing among these non-homogeneous devices [4].

Although several essential challenges rest to complete the IoT vision, the most important among them is security [5], because it remains the primary barrier to the large-scale acceptance and implementation of IoT. The IoT devices are susceptible to various attacks for several reasons [1]. Therefore, without a powerful security establishment, threats, and malicious activities of the IoT will override all of its goodness [5]. Typical security concerns including authenticity, message confidentiality, and integrity are important for every element of the IoT systems. Among these requirements, authentication is most essential because it is an initial level of security. The elements in the IoT are objects, network connectivity, software, and application. The IoT application should ensure the system authentication because without proper authentication no other security requirements cannot be implemented properly, and it provides an initial level of security. Authentication is validating one's identity in communication and assuring the reliability of the source of communication. It is one of the fundamental goals of security and serves as a gateway in front of a security system to prevent malfunctions. Implementing conventional security techniques including lightweight cryptography and the secure protocol becomes a challenge for the IoT environment because of its restricted resources [7]. Trust management (TM) is an important concept in the IoT for trustworthy data integration and excavating, Quality of Services (QoS), context-aware smart services, improved privacy and security. It makes users overwhelmed by the perception of ambiguity, risk, and accepting usage of IoT smart services and application domains [8]. Therefore, the trust-based solution to ensure authentication will be more suitable for IoT devices because it uses limited resources. In an uncertain environment, trust facilitates interaction among IoT devices. So, implementing and maintaining trust throughout the interaction is necessary for IoT [7].

The contributions of the LogitRegTrust model are summarized as follows:

- Presented the basic introduction about logistic regression and how it classifies and predicts the behavior of the node in the network.
- Discusses the overview of the Routing Protocol for Low power and Lossy Network (RPL) and presents the construction process of Destination Oriented Directed Acyclic Graphs (DODAG) in RPL Network and Black hole attack in RPL.
- Mapping of the LogitRegTrust Model with the IoT based military environment is presented, and its mathematical analysis has also explained.
- The proposed model is involving trust computation among the IoT nodes using Direct Trust (DT), Reputation Score (RS), and Experience Trust (ET). Logistic regression is used to compute the Integrated Trust. There are two phases in Logistic Regression (LR), one is the training phase, and another one is the prediction phase. During the training phase, LR learns the node's behavior, based on the DT, RS, and ET, then it predicts the node's behavior (Trusted or
Malicious). The result will isolate the malicious nodes only trusted nodes will present in the network, hence authentication can be ensured and also can achieve security.

- The performance evaluation of the LogitRegTrust model is compared with the existing similar work to show the merits of the proposed model.

2. Background

2.1. RPL Overview

The proposed model is implemented with the RPL protocol. It is a proactive and distance-vector protocol for IPv6-based low Power and lossy networks. This routing protocol supports three fundamental traffic flows: Point-to-Point Traffic (P2P), Multicast to Point(M2P) traffic, and Point-to-Multicast(P2M) traffic [9]. The following section discusses the Destination Oriented Directed Acyclic Graph (DODAG) construction process in the RPL and the execution of the black hole attack in RPL.

2.1.1. RPL Protocol Description.

The RPL network is represented as a Directed Acyclic Graph (DAG), it may comprise many Destination Oriented DAGs (DODAGs). Each DODAG should have only one root, also known as Border Root (BR), which is connected with other roots through the backbone, and also connects with external networks. All nodes in the DODAG contain an identifier (IPv6 address), parent list, discovered neighbor list, rank, and other parameters. Trickle timer and ICMPv6 (Internet Control Message Protocol for IPv6) message dissemination is used to develop and manage the network topology. The DODAG Information Solicitation (DIS), DODAG Destination Advertisement Object (DAO), and DODAG Information Object (DIO) are ICMPv6 messages [10].

![Diagram of DODAG construction process in RPL.]

2.1.2. DODAG construction process in RPL.

Figure 1 illustrates the construction process of the RPL network. The DODAG construction method starts from the DODAG root. Initially, the DODAG root broadcasts the DIO message that contains the DODAG information. The nearby nodes that get the DIO control message of the border root will compute the distance cost and conclude whether or not to associate itself to DODAG. When the nearby nodes are associated with the DODAG, the border root becomes the parent of this node and the new node has a route to the border root. Afterward, the newly joined node in the DODAG estimates its rank value in the network and responds to the parent to notify its presence through the DAO message. The nodes that have not received any DIO messages in the network can inquire to their nearby nodes by sending DIS messages periodically [11].
The control messages that hold the routing details are broadcast periodically. The regularity is driven by the algorithm using the trickle timer that depends on how dynamically changes its network topology from a stable network. The primary task of a routing protocol is to construct a path for data packet transfer, assure routing topology information remains undamaged during transmission, data packets can be used by legitimate entities and that is accessible when required [12].

RPL is susceptible to various routing attacks against Wireless Sensor Networks (WSNs) as well as the IoT, so it is necessary to study routing attacks in RPL and implement security techniques in the RPL[13]. Therefore, the following section discusses the RPL adversary model.

2.1.3. RPL Adversary model – The Blackhole attack.

In this work, the adversary model assumes that an attacker is an internal node in the RPL network. The suspicious node introduces a black hole attack that drops the data packet that is transferred through this node.

Figure 3 illustrates the RPL network with no attacks. R is the root node or sink node, and all other nodes are internal and leaf nodes. In this example, the leaf node N11 transfers its data to the border root through the internal nodes N5 and N1. These nodes are trusted nodes; therefore, they forwarded the data packet to the sink node with no packet dropping.

Figure 3 depicts the example network scenario of the RPL network with a black hole attack. Node N5 is a suspicious node that launches the black hole attack. In this example, node N11 transfers its
data packet through suspicious node N5 that drops the data packets. Therefore, data will not be reached to the root node.

2.2. Logistic Regression Model

It is a statistical modeling method, where the dependent variable includes only two apparent values [14]. This technique is used to learn the connection between the dependent and independent variables. The independent variables can be in the form of discrete, continuous, or categorical. The proposed LogitRegTrust model uses logistic regression to classify and predict the node's behavior. It learns the node's behavior using DT, RS, and ET during its training phase, then it predicts the node's behavior accurately. In this model, the dependent variable is node behavior whether it may be trustworthy or untrustworthy, and the independent variables are DT, RS, and ET.

The following equation defined the logistic regression function

\[ p = \frac{1}{1 + \exp(-z)} \]  \hspace{1cm} (1)

\[ p = 1/(1 + \exp(- (b_0 + b_1 * x_1 + b_2 * x_2 + b_3 * x_3))) \]  \hspace{1cm} (2)

Where \( p \) is denoted as a possibility of the incidence of variable \( y \), and the range of \( p \) is (0 to 1).

The logit model gets real-valued inputs and generates a prediction. In this model, it takes DT, RS, and ET as input values and predicts the node's behavior (Trusted or Malicious) as the output. When the probability of output falls above the predetermined threshold value then the result of the prediction belongs to a trusted class otherwise malicious class.

For this LogitRegTrust model, the Logistic Regression takes four coefficients (\( b_0, b_1, b_2, b_3 \)) and three input values. For example,

\[ Z = b_0 + b_1 * x_1 + b_2 * x_2 + b_3 * x_3 \]  \hspace{1cm} (3)

Where \( b_0 \) denotes intercept or bias and assuming the input value of the coefficient \( b_0 \) is 1.0, \( x_1, x_2, x_3 \) are denoting input values for DT, RS, and ET.

The primary work of the learning algorithm is to find the correct estimation of the parameters (\( b_0, b_1, b_2, b_3 \)) from the sample data. To get the Logit parameters (coefficients), the stochastic gradient descent procedure is used to optimize the coefficients of the Logit model.

2.2.1. Logistic Regression by Stochastic Gradient Descent.

Stochastic Gradient Descent (SGD) algorithm accurately computes the maximum likelihood of logistic regression coefficients from input data. This is an advanced method to renew the regression parameters using a single training point (regression error) at a certain time. This algorithm is an online learning algorithm because classifiers update the training data incrementally at a particular time. Every sample data is given into the model at a certain time by this optimization algorithm. This model forecasts the sample data based on the current values of the parameters and computes updated parameters depending on the regression error in the forecasting and updates the model to decrease the error of the next forecasting. This task is iterated until the model is correct enough or for a finite number of repetitions [15].

The following equation is used to update the coefficient (\( b \)) at each iteration.

\[ b = b + \alpha * (y - p) * p * (1 - p) * x \]  \hspace{1cm} (4)

Where \( b \) denotes optimized coefficients or weight, \( x \) denotes input value, \( y \) denotes output and \( \alpha \) are learning rate. It determines how much the coefficients learn each time when it is updated. Best values might be in the range of 0.1 to 0.3. It should be initialized at the starting of the training sample.

3. Related Work

There is a broad range of literature that proposes the solution for security and trust in both general IoT and also in IoT based group communication like military scenarios. The existing work specifically in this area includes cryptography-based trust models and reputation-based trust models are discussed in this section.

Airehrour D et al. [16], proposed a SecTrust-RPL to assure security using trust in RPL protocol. This model resists against Sybil and rank attack. It estimates and determines the trustworthiness of the
node based on the successful data packets shared between two nodes with a particular time and positive acknowledgment from the connected IoT nodes. This model is implemented into the routing protocol called RPL in the IoT network. Alshehri, M. D et al [17], proposed a new centralized mechanism for managing trust in the IoT. This framework provides trustworthy communication between IoT nodes. It contains a dedicated centralized trust managed node also known as Super Node (SN). This system divides the IoT network environment into small regions called clusters for trustworthy interaction among the nodes. This cluster contains the Master Node (MN) which acts as a local trust manager. Each cluster comprises a lot of Cluster Nodes (CN) that communicate among themselves under the supervision of the master node. The super-node contains a central warehouse to keep the trust values of master and cluster nodes for the complete IoT network. The master node contains the local warehouse to keep the trust values of cluster nodes. The centralized trust approach is one of the mechanisms to provide solutions for security issues in IoT.

Djedjig N et al. [18], proposed a trust-oriented solution for RPL in the IoT environment to address both internal attack and external attack of RPL. This model ensures the trustworthiness between the IoT nodes. Each IoT node calculates the direct trust for its nearby nodes based on the direct observation, and opinion is collected from nearby nodes to compute the indirect trust. With these two parameters, the node's trust is computed, based on these trust value nodes choose its parent node. SeeberS et al. [19], proposed a trust calculating framework for cyber-physical systems that are embedded into the RPL. This system used a Trusted Platform Module (TPM) to establish trust and share a key among the legitimate nodes in the network. It provides secure communication among the legitimate nodes. In [20] authors proposed a trust-based parent choosing by building a secure path in the RPL network. This model avoids the suspicious node from the IoT network. It uses a threshold value to filter the suspicious node. The trust aware threshold method is used to estimate the node's trustworthiness. During the path-building process, every node in the network estimates its threshold value from the maximal and mean rank of its nearby nodes. When the rank of the node below the threshold value, then the node is suspicious which maliciously changes its rank to attract the nearby nodes. These kinds of nodes are avoided from selecting as a preferred parent node. In [21], the authors proposed the novel intrusion detection technique to find a black hole attack on RPL in the IoT network. The malicious node launches a black hole attack by broadcasting a relatively high routing metric to its nearby nodes, thus making most of the neighbor nodes select this node as a parent node. This model avoids the nodes that advertise relatively extreme routing properties and store these node identities in the blacklist. Verification is done only on the list that is available on the blacklist. Therefore, overall energy consumption is preserved by SIEWE within the network.

Conti M et al. [22], proposed a safe and extensible RPL routing protocol, primarily designed for IoT networks. This model used the lightweight remote attestation method to assure the software integrity of the IoT nodes in the network. Piggyback's attestation is worked on the control message of the RPL to avoid extra overhead results by adding attestation. SPLIT makes use of the extended property of the RPL and less energy usage, these two factors are important to IoT devices, because of its restricted resources. This model also recommended extendable, intellectual learning to detect the routing attacks like hello-flood, decreased rank attack, and version number change attacks in the IoT. In [23], the authors proposed the Self-Channel Observation Trust and the Reputation System (SCO is a secured routing in wireless ad hoc networks and CPS). After an interaction, this system classified a node behavior as trusted, authenticated, malicious, or selfish based on the performance evaluation. This evaluation outcome is the primary action of the IoT protocol, and it computes the trust using the SCOTRES's properties. Finally, it transfers the indirect recommendations to the trusted one-hop neighbor when it identifies the inspected node changes. It excludes the malicious node in selecting the route path.

Our work differs from the existing work mentioned above. In this model, trust computation is established by combining DT, RS, and ET. Logistic regression is used to compute the IT and predict the node's behavior. It identifies the nodes that launch a black hole attack. Unlike the other classification and prediction model, the Logistic Regression model provides high accuracy in
predicting malicious nodes and low false predictive rate. The deployment of this model in the RPL protocol is very simple. This model identifies and isolates the malicious node and selects the most trusted node for routing. By doing this, the system provides security and ensures authentication among the nodes. This model is more appropriate for resource-constrained devices because it uses simple mathematical computation.

4. Mapping of LogitRegTrust Model with IoT based military environment
Assuming the example network scenario is a highly disruptive military environment. Each soldier is attached to sensors or IoT objects to exchange information with one another on the battlefield. These nodes are used to transfer the soldier information such as health condition, current location, current situation, etc. Assuming one soldier in each team holds non-constrained devices that cannot be compromised by the attacker, and other nodes may be compromised by an attacker. A soldier can join or leave any time from one team to another without informing its neighbor soldiers, and also soldiers can communicate only with their teammates. A node attached to soldiers may change its behavior to perform its specific goals because it is seized by an attacker. In consequence, such nodes may slow down the network operations and expand infrastructure vulnerabilities. To identify these nodes, the LogitRegTrust model is used. This model considers DT, RS, and ET for trust evaluation because DT is not enough to identify the node's behavior. because malicious nodes may perform well with some nodes in an initial stage later, it launches the attacks. Therefore, the reputation score (Global Trust) is used which helps to predict the node's behavior accurately. Each soldier in the team calculates its neighbor soldier's DT through their interaction, then transfers their DT values as Recommendation Trust (RT) to the dedicated soldier who computes the RS for all nodes. Each soldier calculated IT using DT, ET, and RS. Logistic Regression predicts neighbor soldier behavior as trustworthy or untrustworthy. Based on the prediction, soldiers interact with the neighbor soldier, then update its trust value.

5. LogitRegTrust Model – The Proposed Model
There are two ways to provide authentication. One is the cryptography technique, and another one is trust-based authentication. Because of the resource-constrained, cryptography methods cannot be applied to IoT devices, it creates additional overhead and drains the node's energy. Therefore, trust-based security provides the best solution for ensuring nodes authentication in the system by analyzing the node's behavior. To analyze the node's behavior, Logistic Regression is used in the system, and it learns the node's behavior during its learning phase then it predicts the node's behavior, based on the prediction LogitRegTrust model selects the authenticated node for communication. This model is primarily designed to identify the blackhole attacks in the IoT environment. It identifies and isolates the malicious nodes that launch the black hole attack and select the trusted node for communication. In this way, it ensures the authentication in the system. The main reason for using logistic regression is to improve the prediction accuracy rate.

Typically, the black hole attack makes an impact on the following Quality of Service (QoS) metrics such as packet delivery ratio, end to end delay, and energy depletion. Therefore, in this model, we are considering the above metrics to estimate the trust values of IoT nodes/devices.

5.1. Network Model Assumptions
The LogitRegTrust model is developed with the following underlying assumptions.

- Each DODAG instance comprises at least one non-resource constrained devices, that node is called Dedicated Node (DN).
- The communication range of this DN will be obtainable by all IoT nodes in the RPL instances. This DN node is a predefined trusted node and it is not compromised by an attacker.
- The rest of the IoT nodes in the RPL network are resource-constrained mobile nodes. Since nodes can be easily compromised by an attacker.
• At any moment, malicious nodes may enter or leave the network.
• Attack identification and isolating only in an intra DODAG level (internal attacks).

5.2. Trust Computation.
In the LogitRegTrust model, Quality of Service (QoS) metrics are used to calculate the direct trust. The metrics are packet delivery ratio, end to end delay, and energy consumption. These trust metrics are described as follows.

5.2.1. Packet Delivery Ratio.
It is a proportion between the total amount of data packets forwarded by the source node and the total amount of data packets received by the destination node.
The following equation is used to compute the Packet Delivery Ratio (PDR) of the data packets over the time ‘t’.

\[
PDR(t) = \frac{TPR(t)}{TPF(t)}
\]  

Where TPR refers to the total amount of data packets received by the target node at a ‘t’ time and TPF is the total data packet forwarded by the origin node at a ‘t’ time.

5.2.2. Average Delay.
It includes every potential delay that occurred during path detection, propagation, re-transmission, and relay time. This metric plays a major role in the comprehension of the End to End Delay (EED) investigated by route discovery at a time ‘t’. It can be calculated as follows.

\[
EED(t) = \frac{(PRT - PTT)}{TNP}
\]  

Where PRT denotes the Packet Received Time and it is the time to reach the initial information of the packet to the destination node. PTT denotes the Packet Transmission Time which means the time to initial information of the packet is delivered from the origin node and TNP denotes the total amount of data packets transferred.

5.2.3. Energy Consumption.
The following equation provides the Energy Consumption (EC) of a particular node when packet transmitting, receiving, listening, and processing at a time ‘t’.

\[
EC(t) = (PR * RT) + (PS * RT) + (L * RT) + (P * RT)
\]  

Where PR denotes Packet Received, PS denotes Packet Sent, L denotes Listening of incoming packets, P denotes Processing the packets and RT denotes Required Time to do all the above operations.

5.2.4. Direct Trust Computation.
Direct Trust value is calculated from the equations (5), (6), and (7). The following Logistic Regression equation is used to compute the direct trust for each node.

\[
DT(t) = \frac{1}{(1 + \exp - (b_0 + b_1 PDR(t) + b_2 EED(t) + b_3 EC(t)))}
\]  

Where, DT(t) denotes dependent variable (Direct Trust)
PDR(t), EED(t), EC(t) are represented as Independent variables (QoS Metrics)
b’s are denoted as unknown regression coefficients. (Stochastic Gradient Descent use to learn coefficients)
b_0 denoted as constant

The range of the DT lies between 0 and 1. This equation predicts the node's behavior. If this value falls under the threshold, then the tested node is malicious and it might perform black hole attacks. Each node tests its neighbor node’s local trust based on its own experience.
5.2.5. Reputation Score computation.
Reputation is one of the important factors for trust calculation, which is a global trust. In this model, the Dedicated Node (DN) calculates the reputation score for all nodes. It periodically collects the Recommendation Trust (RT) from all nodes and updates the Reputation Score (RS). The following equation is used to compute the Reputation Score (RS).

\[ RS_D(t) = \frac{\sum_{k=1}^{m} RT_{NK,D}(t)}{m} \]  

(9)

Where \( m \) denotes the number of nodes that had direct experience with the Destination Node. 
\( D \) denotes the Destination Node. 
\( RT_{NK,D}(t) \) denotes the Recommendation Trust provided by the \( NK \)th node for the destination node. This node had direct experience with the Destination node. 
\( RS_D(t) \) denotes Reputation Score for Destination Node at ‘t’ time.

The following figure 4 illustrates the example scenario of the reputation score calculation.

![Figure 4. Reputation Score Calculation](image)

In figure 4 the nodes, N1, N2, N3 had direct experience with the destination node. DN receives RT from these nodes and calculates the RS for the destination node using equation (9), then it returns the RS to the source node. As discussed earlier the DN is one of the nodes in the network which has more power, memory, and processing speed than other nodes.

5.2.6. Experience Trust Computation.
Each node maintains the count of success and failure interaction with the neighbor nodes from its experience. Experience Trust (ET) value is calculated using the following equation.

\[ ET(t) = \frac{(SI(t) + 1)}{(SI(t) + FI(t) + 2)} \]  

(10)

Where SI denotes the successful interaction and FI denotes failed interaction. When the numerator has ‘+1’ and denominator has ‘+2’ which shows that at least two trials are observed out of which one is ‘successful’ and the other is ‘failure’ according to Laplace law.

5.3. Trust Aggregation
Trust aggregation is used to integrate all trust values which are calculated from Direct Trust (DT), Recommendation Score (RS), and Experience Trust (ET). The following equation is used to calculate the Integrated Trust (IT).

\[ IT(t) = \frac{1}{1 + \exp - (b_0 + b_1 * DT(t) + b_2 * RS(t) + b_3 * ET(t)))} \]  

(11)

Where, \( b_0, b_1, b_2, b_3 \) denotes Regression Coefficients. Now IT contains the output of the trained model, based on this value, predictions can be made.

5.4. Trust Update.
Trust is updated based on the change of satisfaction degree of the DN.
5.5. Identifying Malicious nodes.
Based on the IT values, nodes are classified into two classes, one is trusted nodes, and another one is the malicious node. Threshold values are used to separate the two classes. Depending on the importance of the application domain, the value of the threshold can be determined. If the Integrated Trust (IT) value is larger than or equal to the threshold, then the nodes belong to the trusted nodes, otherwise, the node belongs to malicious nodes.

| Level | Threshold | Description       |
|-------|-----------|-------------------|
| 1     | If IT >= Threshold | Trusted Node      |
| 2     | If IT < Threshold   | Malicious Node    |

Trusted nodes are selected for routing operation and malicious nodes are added to the blacklist and discard from the network. Dedicated Nodes broadcast the malicious node details to all other nodes in the network, thus making the other nodes also disconnect the link from the malicious nodes. By doing this, malicious nodes are removed from the network, and authentication is ensured.

Algorithm 1: LogitRegTrust Model

Step 1: Deploy N, IoT nodes in the RPL network (Both Trusted and Malicious Nodes)
Step 2: Over the period, all the nodes in the network computes its one-hop neighbor node’s Direct Trust (DT) and Experience Trust (ET) from equation (8) and (10) based on their own experience (Local Trust computation).
Step 3: All nodes store their one-hop neighbors’ DT value.
Step 4: Nodes send the Recommendation Trust (RT) to the Dedicated Node (DN) to calculate the Reputation Score (RS) for each node (Global Trust Computation). DN computes RS from equation (9).
Step 5: Each node computes the Integrated Trust (IT) value based on the DT, RS, and ET from equation (11). It is a prediction value. Logit used to predict the node’s behavior (Trusted/Malicious)
Step 6: if IT=>Threshold value then
    P=1 //predicted as a trusted value
    //Trusted and Authenticated Node, Select for the interaction
else
    P=0 //predicted as a Malicious node. Isolate this node and select another node to continue.
    End if
Step 7: Based on the prediction, nodes select their neighbors for the interaction.
Step 8: if interaction fails then
    Malicious node
    //Isolate this node. It might perform a black hole attack.
End if
End

The following figure 5 illustrates the overall structure of the LogitRegTrust model.
6. Mathematical Analysis of LogitRegTrust Model with IoT based military environment
The following example network scenario is taken for the mathematical analysis.

In the above figure, soldier S6 is a black soldier who performs a black hole attack. S2 is a dedicated soldier who computes the RS for all other soldiers in the team. Soldiers, S10, S11, S12 are the child nodes of the malicious soldier S5. These soldiers transfer their data packets through this malicious soldier to the sink. Instead of forwarding data, the malicious node drops the data packets. Each soldier calculates their neighbor soldier DT and ET value using the equation (8) and equation (10), then they receive RS from soldier S2. RS is calculated from equation (9). Finally, they compute the IT using equation (11).
6.1 Trusted Soldier IT calculation

For example, soldier S12 in figure 6 computes the IT for the parent soldiers S6 as follows. Assuming DT=0.8, RS=0.9 and ET=0.7

\[ \text{IT}_{s12, s6}(t) = \frac{1}{1 + \exp(-b_0 + b_1*DT + b_2*RS + b_3*ET)} \]

Assign \( p = \text{IT}_{s12, s6}(t) \)

**Training Phase:**
**First Training Instance**

In the first training instance, initialize 0.0 for \( b_0, b_1, b_2, b_3 \) in equation (11).

\[ p = \frac{1}{1 + \exp(0.0 + 0.0*0.8 + 0.0*0.9 + 0.0*0.7)} = 0.5 \]

**Second Training Instance**

New coefficient values computed using the prediction \( p(0.5) \) and the coefficient values (0.0) from the first training instance. Assuming the previous interaction between S12 and S6 was successful, so \( y = 1 \) otherwise \( y = 0 \). \( \alpha \) is a learning rate and that should be initialized before the training instance. The best range of these values is between 0.1 to 0.3. We take \( \alpha = 0.3 \). Coefficient is updated using the equation (4).

\[
\begin{align*}
  b_0 &= 0.0 + 0.3 \times (1 - 0.5) \times 0.5 \times (1 - 0.5) \times 1 = 0.0375 \\
  b_1 &= 0.0 + 0.3 \times (1 - 0.5) \times 0.5 \times (1 - 0.5) \times 0.8 = 0.03 \\
  b_2 &= 0.0 + 0.3 \times (1 - 0.5) \times 0.5 \times (1 - 0.5) \times 0.9 = 0.03375 \\
  b_3 &= 0.0 + 0.3 \times (1 - 0.5) \times 0.5 \times (1 - 0.5) \times 0.7 = 0.02625
\end{align*}
\]

Now updated coefficient \( b_0, b_1, b_2, b_3 \) are assigned in equation (11)

\[ p = \frac{1}{1 + \exp(0.0375 \times 1.0 + 0.03 \times 0.8 + 0.03375 \times 0.9 + 0.02625 \times 0.7)} = 0.5275 \]

| Training Instance | Coefficients | Prediction |
|--------------------|--------------|------------|
| TI-1               | 0.0375       | 0.0338     | 0.0263 | 0.53 |
| TI-3               | 0.0727       | 0.0583     | 0.0655 | 0.0509 | 0.55 |
| TI-4               | 0.1058       | 0.0847     | 0.0954 | 0.0741 | 0.58 |
| TI-5               | 0.1367       | 0.1094     | 0.1232 | 0.0957 | 0.60 |
| TI-6               | 0.1655       | 0.1325     | 0.1492 | 0.1159 | 0.62 |
| TI-7               | 0.1924       | 0.1540     | 0.1734 | 0.1347 | 0.64 |
| TI-8               | 0.2175       | 0.1741     | 0.1960 | 0.1523 | 0.66 |
| TI-9               | 0.2409       | 0.1928     | 0.2171 | 0.1687 | 0.67 |
| TI-10              | 0.2628       | 0.2103     | 0.2367 | 0.1840 | 0.68 |
| TI-11              | 0.2833       | 0.2267     | 0.2552 | 0.1983 | 0.70 |
| TI-12              | 0.3025       | 0.2421     | 0.2724 | 0.2117 | 0.71 |
| TI-13              | 0.3205       | 0.2565     | 0.2887 | 0.2444 | 0.72 |
| TI-14              | 0.3372       | 0.2698     | 0.3037 | 0.2560 | 0.73 |
| TI-15              | 0.3529       | 0.2824     | 0.3179 | 0.2671 | 0.74 |
| TI-16              | 0.3678       | 0.2944     | 0.3313 | 0.2775 | 0.75 |
| TI-17              | 0.3819       | 0.3056     | 0.3440 | 0.2874 | **0.76** |
| TI-18              | 0.3953       | 0.3164     | 0.3560 | 0.2968 | **0.76** |

Repeat the process until to get the accurate prediction value. Table 2 shows the coefficient and prediction for each training phase. In the Eighteenth training instance, accurate prediction is got for soldier S6. The prediction value is 0.76.

**Prediction Phase**

In the prediction phase, the Threshold value has to be fixed to classify and predict the node's behavior. In our military based application, the threshold value is fixed as **0.5**. Here \( p \) is the IT value. If IT is
greater than 0.5 then the node is trusted. In this example IT is 0.76, therefore the soldier S6 is an authenticated soldier, and this soldier is selected for communication.

6.2 Malicious Soldier IT calculation

Now, soldier S12 compute the IT for the parent soldier S5:

We assume DT=0.6, RS=0.4 and ET=0.5

First Training Instance

In the first Training instance, we initialize 0.0 for b0, b1, b2, b3 in equation (11).

\[ IT_{s12,s5}(t) = \frac{1}{1 + \exp(-b_0 + b_1 \times 0.6 + b_2 \times 0.4 + b_3 \times 0.5)} \]

We assign \( p = IT_{s12,s5}(t) \)

\[ p = \frac{1}{1 + \exp(-0.0 + 0.0 \times 0.6 + 0.0 \times 0.4 + 0.0 \times 0.5)} = 0.5 \]

Assuming previous interaction between S12 and S5 was unsuccessful, Therefore \( y = 0 \).

Second Training Instance

\( \alpha = 0.3, \ p = 0.5 \) and 0.0 for \( b_0, b_1, b_2, b_3 \).

Coefficient is updated using the equation (4)

\[ b_0 = 0.0 + 0.3 \times (0 - 0.5) \times 0.5 \times (1 - 0.5) \times 0.6 = -0.0375 \]
\[ b_1 = 0.0 + 0.3 \times (0 - 0.5) \times 0.5 \times (1 - 0.5) \times 0.4 = -0.0225 \]
\[ b_2 = 0.0 + 0.3 \times (0 - 0.5) \times 0.5 \times (1 - 0.5) \times 0.4 = -0.0150 \]
\[ b_3 = 0.0 + 0.3 \times (0 - 0.5) \times 0.5 \times (1 - 0.5) \times 0.5 = -0.01875 \]

Now updated coefficient \( b_0, b_1, b_2 \) and \( b_3 \) are assigned in equation (11)

\[ p = \frac{1}{1 + \exp(-0.0375 + 0.0225 \times 0.6 + 0.0150 \times 0.4 - 0.01875 \times 0.5)} = 0.48 \]

Repeat the process until to get the accurate prediction value. Table 3 shows the coefficient and prediction for each training phase.

| Training Instance | Coefficients | Prediction |
|-------------------|--------------|------------|
| TI-1              | 0            | 0          | 0          | 0          | 0.5        |
| TI-2              | -0.0375      | -0.0225    | -0.0150    | -0.0187    | 0.48       |
| TI-3              | -0.0737      | -0.0442    | -0.0294    | -0.0368    | 0.46       |
| TI-4              | -0.1086      | -0.0651    | -0.0434    | -0.0543    | 0.45       |
| TI-5              | -0.1422      | -0.0853    | -0.0568    | -0.0711    | 0.43       |
| TI-6              | -0.1745      | -0.1047    | -0.0698    | -0.0872    | 0.42       |
| TI-7              | -0.2055      | -0.1233    | -0.0822    | -0.1027    | 0.41       |
| TI-8              | -0.2352      | -0.1411    | -0.0941    | -0.1176    | 0.39       |
| TI-9              | -0.2638      | -0.1582    | -0.1055    | -0.1319    | 0.38       |
| TI-10             | -0.2912      | -0.1747    | -0.1164    | -0.1456    | 0.37       |
| TI-11             | -0.3174      | -0.1904    | -0.1269    | -0.1587    | 0.36       |
| TI-12             | -0.3426      | -0.2055    | -0.1370    | -0.1713    | 0.35       |
| TI-13             | -0.3668      | -0.2201    | -0.1467    | -0.1834    | 0.34       |
| TI-14             | -0.3900      | -0.2340    | -0.1560    | -0.1950    | 0.33       |
| TI-15             | -0.4123      | -0.2473    | -0.1649    | -0.2061    | 0.32       |
| TI-16             | -0.4337      | -0.2602    | -0.1734    | -0.2168    | 0.31       |
| TI-17             | -0.4543      | -0.2725    | -0.1817    | -0.2271    | 0.30       |
| TI-18             | -0.4741      | -0.2844    | -0.1896    | -0.2370    | 0.30       |

Predication Phase

In the Eighteenth training instance, accurate prediction is obtained. Soldier s5 IT value is 0.30 which is less than the threshold value 0.5. Therefore, the soldier s5 is a malicious soldier. Now soldier
s12 disconnects the link from the parent soldier s5 and selects s6 as a preferred soldier and transfers the data through s6 soldier.

7. Simulation Results and Discussion

7.1 Performance Evaluation Metrics

The LogitRegTrust model is evaluated in the Contiki 3.0 OS and the Cooja simulator. The LogitRegTrust model uses TMote Sky (Sensor nodes) as a mote type. The following table shows the simulation parameters of the proposed trust model.

| System Parameters                  | Values                          |
|------------------------------------|--------------------------------|
| Number of nodes                    | 50                             |
| Mote Type                          | TMote Sky                      |
| Simulation Time                    | 3600Sec                        |
| Network Coverage Area              | 300mx300m                      |
| Data Rate                          | 3072bps                        |
| Data Packet Size                   | 64 bytes                       |
| Traffic                            | UDP                            |
| Mac Layer                          | IEEE 802.15.4                  |
| Communication Range                | 50m                            |
| RPL Parameter                      | MinHopRankIncrease=256         |
| Routing Protocol                   | LogitRegTrust, Trust-based RPL |
| $\alpha$ - value                   | 0.3                            |
| coefficient values(b0, b1, b2 and b3) | 0.0                           |
| Threshold value                    | 0.5                            |

7.2 Simulation Results

The performance evaluation of the LogitRegTrust model is compared with the following cases.

1. The proposed model aims to identify the malicious nodes which perform a black hole attack. Therefore, it is necessary to know the impact of the malicious nodes. Here, increase the percentage of malicious nodes and measure the packet dropping ratio in RPL.
2. Increase the percentage of malicious nodes and compare the detection ratio of LogitRegTrust with Trust-based RPL.
3. The performance of the LogitRegTrust model is compared with Trust-based RPL [24] in terms of Delivery Ratio, Average Delay, and Throughput.

![Figure 7. Impact of Malicious nodes under normal RPL](image-url)
Scenario 1: The analysis is performed with the varying number of malicious nodes under normal RPL routing protocol. The observations in figure 7 show, when the number of malicious nodes increases, data dropping also increases.

![Detection Ratio vs Malicious Nodes](image)

Figure 8. Malicious nodes vs detection ratio

Scenario 2: Figure 8 depicts the detection ratio of the LogitRegTrust model and Trust-based RPL. This simulation aims to discover the suspicious nodes in the network. In this simulation, the number of malicious nodes increases from 0 to 80% to estimate the performance of the proposed LogitRegTrust model. Because of the global trust computation and implementation of the logistic regression in the proposed model, the detection ratio of the malicious node is greater than the Trust-based RPL. When malicious nodes increase, the detection ratio also increases in both models, however, the trust-based RPL can detect only 35% of malicious nodes, when the network comprises 80% of malicious nodes. But the proposed model can detect 72% of malicious nodes from the network.

Scenario 3: In this simulation, the performance metrics such as delivery ratio and the average delay and throughput of the LogitRegTrust model are compared with the Trust-based RPL.

Packet delivery ratio: It is a proportion between the total amount of data packets forwarded by the source node and the total amount of data packets received by the destination node. It is one of the significant metrics for evaluating the performance of the proposed model. This metric used to analyze the delivery ratio for the individual node and also for the whole network. Protocols are evaluated by varying percentages of the malicious nodes. These malicious nodes are increased from 0 to 80%. Figure 9 shows the delivery ratio of Trust-based RPL and LogitRegTrust. Results depict the proposed model's packet delivery ratio is higher than the Trust-based RPL. The reason is, the proposed model not only considers a single trust metric but also considers multiple trust metrics. The Trust-based RPL used forwarding ratio as a trust metric to detect behavior, but the proposed model uses forwarding ratio, average delay, and energy utilization to test the node's trustworthiness. Because of this multiple trust metric, the proposed model can easily detect the black hole attack and a centralized dedicated node removes these malicious nodes from the network. Therefore, the data packets are transmitted through only trusted nodes that increase the delivery ratio. Once malicious nodes are removed from the network, it will not be involved in any routing operation.
Average delay: It is measured as the average time needed to send a packet from source to destination. It is another important metric to measure the functionality of the proposed protocols. The presence of misbehaving nodes in the IoT network increases the delay. Figure 10 depicts the impact on the delay in the Trust-based RPL and LogitRegTrust model with the varying percentage of the malicious nodes. It shows that the average delay of the LogitRegTrust model is lesser than the Trust-based RPL. The existing Trust-based RPL used fuzzy logic to classify the node's behavior, but the proposed model uses the Logistic regression to predict the behavior of the node which accurately estimates the trust value and selects the most trusted nodes for routing, thus avoids malicious nodes from the network and decreases the average delay.

Average Throughput: The total amount of data packets transferred per unit time or an average number of successful information transferred per second over a communicating transmission channel is called throughput. It is represented in bits per second (bits/s or bps).
Throughput = \sum_{p=1}^{n} \frac{\text{packet transmitted}(p)}{\text{Total Time}}

![Figure 11. Malicious Nodes vs. Throughput.](image)

The average throughput of the proposed LogitRegTrust model is compared with the Trust-based RPL. As in figure 11, the average throughput of the proposed model is greater than the trust-based RPL. Because the proposed model uses global trust for indirect trust computation, which provides an accurate trust value than the local trust computation. Therefore, the proposed model can easily identify the node behavior at the initial stage of interaction. The misbehaving nodes are immediately removed from the network. When the number of malicious nodes decreases, the average throughput is increased in the proposed model. But, the existing Trust-based RPL uses the local trust for indirect trust computation, it can identify the malicious nodes after a certain number of interactions. The presence of the malicious nodes decreases the throughput of the Trust-based RPL.

8. Conclusion
The IoT applications are rapidly evolving over the period, at the same time security challenges are also increased. The success of the application highly depends upon its security. One of the primary security requirements for all applications is authentication, which assures the right identity of the objects. The proposed LogitRegTrust model ensures the authentication among the IoT nodes using the trust. This model uses the reputation score (global trust) for trust computation. The Trust-based RPL uses local trust for indirect trust value, but the proposed model uses the global trust to compute the indirect trust value of the node. Thus, the malicious nodes can be accurately identified by the LogitRegTrust model when compared to the Trust-based RPL. Besides that, the LogitRegTrust model uses multiple trust metrics and logistic regression machine learning techniques to predict the node's behavior based on direct experience and global trust. The existing Trust-based RPL used a single trust metric and fuzzy logic to identify the node's behavior. When compared to the fuzzy logic, the logistic regression predicts the node's behavior correctly. In this model, trusted nodes (authenticated nodes) are only involved in the routing operation, malicious nodes are removed from the network. The LogitRegTrust model can apply for all applications, but it will be more appropriate for group-based
military applications. In an open and remote battlefield environment, battlefield things require communication and cooperation to achieve a mission. Implementing the LogitRegTrust model in the battlefield network improves the mission's effectiveness by discarding the malicious battlefield thing from the network. The mathematical model has been proven the applicability of the proposed model. The proposed trust model has been embedded into RPL and the performance of the LogitRegTrust model is evaluated using a Cooja simulator. The performance evaluation shows the effectiveness of the LogitRegTrust with varying percentages of malicious nodes as compared to the Trust-based RPL in terms of forwarding ratio, average percentages, energy detection ratio, and throughput.

Acknowledgment
This research work is supported by Karpagam Academy of Higher Education (Deemed to be University), Coimbatore, Tamilnadu India through a Seed Project.

References
[1] Abera T, Asokan N, Davi L, Koushanfar F, Paverd A, Sadeghi A R, and Tsudik G 2016. Invited Things, trouble, trust. Proceedings of the 53rd Annual Design Automation Conference on - DAC ’16. DOI:10.1145/2897937.2905020.
[2] Agrawal S, Das ML. Internet of Things – a paradigm shift of future Internet applications. In: Proceedings of the Nirma University International Conference on Engineering (NUiCONE); 2011. p. 1–7.
[3] Zhou W, Jia, Y, Peng A, Zhang, Y, & Liu, P (2018). The Effect Security and Privacy: New Threats, Existing Solutions, and Challenges Yet to Be Solved. IEEE Internet of Things Journal, 1–1. DOI:10.1109/jiot.2018.2847733.
[4] Airhbour D, Gutierrez J, & Ray S K (2016). Secure routing for the internet of things: A survey. Journal of Network and Computer Applications 66, 198 213. DOI:10.1016/j.jnca.2016.03.006.
[5] Roman R, Najera P, & Lopez, J (2011). Securing the Internet of Things. Computer, 44(9), 51–58. DOI:10.1109/mc.2011.291.
[6] Hammi M T, Hammi B, Bellot P, and Serhrouchni A (2018). Bubbles of Trust: A decentralized blockchain-based authentication system for IoT. Computers & Security, 78, 126–142. doi:10.1016/j.cose.2018.06.004.
[7] Aljazzaf Z M, Perry M, and Capretz M A M (2010). Online Trust: Definition and Principles. 2010 Fifth International Multi-Conference on Computing in the Global Information Technology. doi:10.1109/iccci.2010.17.
[8] Yan Z, Zhang P, and Vasilakos A V (2014). A survey on trust management for Internet of Things. Journal of Network and Computer Applications, 42, 120–134. doi:10.1016/j.jnca.2014.01.014.
[9] Elyengui S, Bouhouchi R, & Ezzedine T (2015). A comparative performance study of the routing protocols RPL, LOADng and LOADng-CTP with bidirectional traffic for AMI scenario. 2015 International Conference on Smart Grid and Clean Energy Technologies (ICSGCE). doi:10.1109/icsgce.2015.7454267.
[10] Medjek F, Tandjaoui D, Abdmeziem M R, & Djedjig N (2015). Analytical evaluation of the impacts of Sybil attacks against RPL under mobility. 2015 12th International Symposium On Programming and Systems (ISPS). doi:10.1109/isps.2015.7244960.
[11] Tian H, Qian Z, Wang X, and Liang X (2017). QoI-Aware DODAG Construction in RPL-Based Event Detection Wireless Sensor Networks. Journal of Sensors, 2017, 1–9. doi:10.1155/2017/1603713.
[12] Glissa G, Rachedi, A, & Meddeh, A (2016). A Secure Routing Protocol Based on RPL for Internet of Things. 2016 IEEE Global Communications Conference (GLOBECOM). doi:10.1109/glocom.2016.7841543.
[13] Wallgren L, Raza S, & Voigt T (2013). Routing Attacks and Countermeasures in the RPL-Based Internet of Things. *International Journal of Distributed Sensor Networks*, 9(8), 794326. doi:10.1155/2013/794326.

[14] Hosmer D W, Jr. and S. Lemeshow, *Applied Logistic Regression* (John Wiley & Sons, New York, 1989).

[15] Al-Shihi G H, Bayat N, Fathi Ashtiani A, Tavallaii S A, Assari S, & Moharamzad Y (2008). Logistic regression models for predicting physical and mental health-related quality of life in rheumatoid arthritis patients. *Modern Rheumatology*, 18(6), 601–608. doi:10.3109/s10165-008-0092-6.

[16] Airehrou D, Gutierrez JA, Ray SK (2018) SecTrust-RPL: a secure trust-aware RPL Routing protocol for Internet of Things. *Future Gener Comput Syst*. https://doi.org/10.1016/j.future.2018.03.021.

[17] Alshehri M D, and Hussain F K (2017). A Centralized Trust Management Mechanism for the Internet of Things (CTM-IoT). *Lecture Notes on Data Engineering and Communications Technologies*, 533–543. doi:10.1007/978-3-319-69811-3_48.

[18] Djedjig N, Tandjaoui D, Medjek F (2015) Trust-based RPL for the Internet of Things. In: 2015IEEE Symposium on Computers and Communication (ISCC), pp 962–967 20.

[19] Sebe S, Sehgal A, Stelte B, Rodosek G D, Schnwlder J(2013) Towards a trust computing architecture for RPL in cyber-physical systems. In: *Proceedings of the 9th International Conference on Network and Service Management (CNSM 2013)*, pp 134–137.

[20] Iuchi K, Matsunaga T, Toyoda K, Sasase I (2015) Secure parent node selection scheme in route construction to exclude attacking nodes from RPL network. In: 2015 21st Asia-Pacific Conference on Communications (APCC), pp 299–303 22.

[21] Patel HB, Jinwala DC (2019) Black hole detection in 6LoWPAN based internet of things: an anomaly based approach. In: *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*. IEEE, pp 947–954 12.

[22] Conti M, Kaliyar P, Rabbani MM, Ranise S (2018) SPLIT: a secure and scalable RPL routing protocol for internet of things. In: *2018 14th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*. IEEE, pp 1–8.

[23] Hatzivasilis G, Papaeftathiou I, & Manifavas C (2017). *SCOTRES: Secure Routing for IoT and CPS*. IEEE *Internet of Things Journal*, 4(6), 2129–2141. doi:10.1109/jiot.2017.2752801.

[24] Airehrour, D., Gutierrez, J., & Ray, S. K. (2016). Securing RPL routing protocol from blackhole attacks using a trust-based mechanism. 2016. *26th International Telecommunication Networks and Applications Conference (ITNAC)*. doi:10.1109/atnac.2016.787879