Journal of Positive Psychology & Wellbeing
2022, Vol. 6, No. 1, 2707 – 2725
http://journalppw.com
ISSN 2587-0130

A TRUST-BASED ROUTING IN IoT ENVIRONMENT USING REINFORCEMENT LEARNING: application for the well-being of people

Carlos Eduardo Andrade Cuadrado[1], Edgar Francisco Llanga Vargas[2], Mercy Esthela Guacho Tixi[3], Carlos Volter Buenaño Pesántez[4], Nilton Chucos Baquerizo[5], Andrea Damaris Hernández Allauca[6]

[1]*Escuela Superior Politécnica de Chimborazo*
https://orcid.org/0000-0002-2769-7202
c_andrade@espoch.edu.ec
[2]*Escuela Superior Politécnica de Chimborazo.*
https://orcid.org/0000-0002-8577-2864
edgar.llanga@espoch.edu.ec
[3]*Escuela Superior Politécnica de Chimborazo*
https://orcid.org/0000-0001-9821-7256
me_guacho@espoch.edu.ec
[4]*Escuela Superior Politécnica de Chimborazo*
https://orcid.org/0000-0002-4170-2290
cbuenano@espoch.edu.ec
[5]*Universidad Nacional Agraria de la Selva*
https://orcid.org/0000-0001-5596-4454
nilton.chucos@unas.edu.pe
[6]*Escuela Superior Politécnica de Chimborazo- Grupo de Investigación y Transferencia de Tecnologías en Recursos Hídricos (GITRH) y Grupo de Investigación en Turismo (GITUR).*
https://orcid.org/0000-0001-6413-5607
andrea.hernandez@espoch.edu.ec

ABSTRACT—

Many application domains gain considerable advantages with the Internet of Things(IoT) network. It improves our lifestyle towards smartness like smart cities, smart health, smart home, smart vehicle, smart grid, etc. The ubiquity of IoT permits numerous heterogeneous smart devices interconnected through the internet to provide smart services. IoT devices are mostly resource-constrained regarding memory, processing capacity, battery, etc. So, it is highly susceptible to security attacks. Traditional security mechanisms can not apply to these devices due to their restricted resources. A trust-based security mechanism plays an important role to ensure security in the IoT environment because it consumes only fewer resources. Thus it is most essential to evaluate the trustworthiness among IoT devices. The proposed model improves trusted routing in the IoT environment by detecting and isolating malicious nodes. This model uses Reinforcement Learning (RL) where the agent learns the behavior of the node and isolates the malicious nodes to improve the network performance. The proposed work focuses on IoT with the Routing Protocol for Low power and Lossy network(RPL) and counters the black hole attack. The simulation results show that the proposed RLTrust model provides better performance than the existing one.

KEYWORDS -IoT, Reinforcement Learning, Trust, security, RPL

I. INTRODUCTION

In recent years, Many academics and public research institutions are focusing on the Internet of Things(IoT). The central concept of IoT is to interconnect loosely defined smart things and make them communicate with other things, the environment, and computing devices[1].

The Internet of Things is the latest trend, and it incorporates many technologies in it. We are gradually entering into the IoT era in which communication takes place between humans and things, and between things itself. In information and communication technology(ICT), IoT brings a new dimension that connects anyone from anyplace at any location. It combines the physical world with the information world. One of the important components of the IoT is the sensor that gathers data from the environment and controls the environment if it requires any changes[2]. Human lifestyle and business gain significant benefits due to the development of IoT. Most of the typical application uses the concept of IoT including greenhouse monitoring, telemedicine monitoring, smart electric meter reading, and intelligent transportation[3]. With the advancement in technology, IoT has become popular and also developed rapidly. These technologies are remote connectivity through fault-tolerant networks, embedded systems, wireless communication, and microelectronics-mechanical systems[4].

Although IoT provides influential advantages, it also faces critical challenges. Some notable challenges are presented here. Devices in the IoT environment are usually open to the public, and it uses wireless communication that creates susceptible to system security. IoT interconnects numerous heterogeneous embedded mobile devices and applications that make difficulties in scalability, dynamic adaptability, and compatibility[4]. The important component of the IoT is the internet in which most of the attacks have occurred. IoT devices are resource-constrained including limited processing capacity, low memory, and energy. Also, a new set of problems will occur because of the high mobility of smart objects and services[5].

However, with the rapid development and broad acceptance of IoT, it is possible to have many kinds of attacks and it violates the security of IoT devices. Because of its limited resources. Traditional security mechanisms cannot be implemented because it consumes more resources of IoT devices. So, IoT systems require a lightweight security mechanism which should handle maximum security attacks. IoT applications consist of a set of things in the network where IoT devices search other devices for the service request. Before accessing the services from the service provider, the trustor node should ensure the trustworthiness of the trustee node[5].

Typically, trust and reputation management are interchangeable. However, both are distinct at a certain point. Trust is dynamic but the reputation is static. Trust is a faith of the node belief based on the qualities of the neighbor nodes while reputation is the opinion about the neighbor node[6].

Trust Management can provide security and it is an essential one in the IoT. Trust-based security is the category of soft security in which the object's behaviors are measured over time to identify the misbehaving nodes. In a changeable IoT environment, objects can misbehave at any time and disrupt the services and performance of the system. Hence, maintaining trust between the objects is a significant aspect[7].

The main goal of this paper is to provide security in a distributed and open IoT environment by selecting trusted IoT devices without the central trust manager. It can be achieved by selecting the devices that have high trust value.

Contribution

The primary contributions of the trust model are listed as follows:

- Presented the fundamental introduction for Markov Decision Processes and Reinforcement Learning.
- Discuss the overview of the RPL and Blackhole attack in the RPL.
- A proposed model involving trust computation among the nodes using both Direct Trust(DT) and Recommendation Trust(RT). In DT calculations trust metrics are aggregated using a weighted linear equation. Recommendation received from common friends and aggregated using the arithmetic mean. Then Composite Trust(CT) is calculated using DT and RT and it gives a Reward in the Q-learning algorithm which generates Q-value.

Node has maximal Q-value and greater than the threshold value is a trusted node that will be involved in routing operation. Nodes are malicious when their Q-value falls below the predefined threshold value, The malicious nodes are avoided from the network operation. The main focus of this paper is to find the data drop attacks and isolate the misbehaving node in the network. By the way, security can be achieved in the IoT network.

- The performance of the trust model is compared with the existing similar work under blackhole attacks to show the merits of the proposed model.

2. BACKGROUND

This section explains the RPL routing protocol and presents a brief description of the Markov Decision Processes and Reinforcement Learning.

2.1 RPL Protocol Description

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. It builds a Destination Oriented Directed Acyclic Graph (DODAG) using the IoT nodes. A DODAG contains the nodes including router, host, gateway, etc. These are all arranged themselves into a specific form of topological structure to perform routing in Low Power and Lossy Networks (LLNs). An individual RPL network consists of several concurrent RPL Instances executed in the same period, it can be recognized by the RPL Instance ID. A single IoT network also contains several DODAGs and it is identified by the DODAG ID(unique IPv6 address). The fundamental aspect of this routing protocol is self-organization, auto-repairing. loop prevention and identification, clarity, and provide several border roots or sink. To construct and manage DODAG, RPL uses various types of messages including DODAG Information Solicitation(DIS), DODAG Information Object(DIO), DODAG Advertisement Object(DAO), DODAG Advertisement Object Acknowledgement(DAO-ACK). First, the DODAG construction process is accomplished in two different ways. 1) Nodes joined in the DODAG network broadcast the DIO messages to its nearby nodes. 2)Node does not receive any DIO message and may request DIS messages to DODAG. The DODAG allows the trickle timer, the member node of the DODAG has to transfer the DAO messages to DODAG at a particular time interval. Then, the DODAG transfers the DAO-ACK messages to all other nodes in the network.

Objective Function (OF) is used to choose the best route between DODAG nodes. It adopts various advantages and restrictions to choose the best path and choose the preferred parent among the various preferred choices. Every node in the RPL has a unique rank value with the 16 bit which shows the distance between the node's current place and border root. This rank value is used to manage the connection between parent and child nodes and prevent loops in the network[8].

2.2*Markov Decision Processes(MDP)*

It is a model based on the consecutive decision by the agent, fixed in an environment. State and the agent in an environment select a particular action at each discrete time. Based on the action and state an agent can obtain the reward and also an environment alter its current position to a new position stochastic-ally[9].

It is explained by a five-tuple S, A, T, R, γ, where S is a definite group of states. A is defined as a group of actions. T is a transition function T(s,a,s'). R is a reward function. R(s, a) indicates the reward acquired by an agent when an agent takes an action 'a' in state 's' and $\gamma \in [0, 1)$ is the discount factor for future reward. π represents the policy, where state $s \in S$ and action $a \in A$, for each state policy π select the action. For a given policy π the value of the Q-function (Q^{π}: $S \times A \rightarrow R$) is described as the expected discount total of all rewards that can be obtained by an agent over an unlimited state transition path beginning from state s taking action

$$\pi(s): Q^{\pi}(s, a) = E[\sum_{k=0}^{\infty} \gamma^{k} r_k | s_k = s, a_k = a, k = 0, 1, 2...] \text{ ----------(1)}$$

where r_k is the reward obtained from the action a_k taken at state s_k, where k is the series index for states and actions. Maximizes the range of every state-action pair is optimal policy:

$$\pi^*(s) = \arg\max_{a \in A} Q^*(s, a) \quad \text{----------(2)}$$

where

$$Q^*(s, a) = E_{s'|s,a}[R(s, a, s') + \gamma \max_{b \in A} Q^*(s', b)] \quad \text{----------(3)}$$

Where R(s, a, s') is the reward for selecting an action 'a' at state s and transfer to state s'. equation (3) is known as Bellman optimal equation. So, identifying an optimal policy is the same as discovering the optimal Q function, that can be solved repeatedly called Value Iteration[10].

2.3 *Reinforcement Learning*

It is a machine learning technique where an agent directly communicates with the surroundings and finds out the control policies depend on their experiences and rewards. Mostly, it is modeled as a Markov Decision Process. It is attached to the environment through action and perception. For each step of communication, an agent gets an input 'i' and a few implications of the present state's' of the environment. Based on this information the agent selects the action 'a' to produce an output. The action modifies the state of the surroundings and the state transmission value that is interacted with the agent via a scalar reinforcement learning signal, r[11]. The goal of the RL is to acquire the highest long-term reward for an MDP environment, Although the model of the environment is difficult to learn or unknown. The method that chooses the maximum long-term reward defines the agent's behavior at a specific time is called a policy[12].

2.4 Q-Learning

Q-Learning is a Reinforcement Learning algorithm that does not depend on a state transition method to work[12]. It is a model-free approach, the goal of this method is to find the optimal decision policy by studying the value of a function Q(s, a).whenever an action 'a' is performed then the agent gets an immediate reward 'r' from the environment. To make future decisions, the Q-learning algorithm updates its Q-value by using reward and expected long-term rewards. The following equation defined one step Q-learning rule

$$Q'(s, a) = (1 - \alpha)Q(s, a) + \alpha[r + \gamma \max_{a \in A} Q(s', a')] \quad \text{----------(4)}$$

where Q(s, a) indicates the quality of action a at state s,

α denotes the learning rate that models the value of updating Q-values.

γ denotes the weight of future rewards.

Q'(s, a) denotes the expected future reward at state s by taking action a[11].

3. RELATED WORK

In recent days, a lot of research work has been done in the IoT to ensure security through trust management. It improves the security in IoT where each node evaluates its neighbor's node's trust value by direct experience or recommendation from other nodes in the network. This section presents a summary of an existing research work on trust evaluation schemes on IoT that includes cryptography-based trust models and reputation-based trust models.

Lahbib A er al[13]., proposed trust model using link and node trust which is implemented into the RPL. This model uses the QoS metric for the trust computation of neighbor nodes. It is also considered the reputation of trust and energy consumption. When the IoT network builds, it

ensures the trust among nodes. This trust also maintains during maintenance. It concentrates on both node trust and link trust to ensure security. Ben Saied et al [14]., Proposed a context-aware and multi-service method to ensure trust in IoT. It assigns a changeable trust score for collaborating nodes based on several contexts and functions. This model collects ratings from other nodes and assigns a recommendation score for each node. This recommendation score is adjusted after each interaction in the learning phase.

Hellaoui et al[15], proposed a dynamic security model for IoT that provides the trust-based solution. This paper primarily deals with selfish behaviors and internal attacks. This model computes the trust level that used to identify the security threats among nodes. Glowacka, et al[16], Presents a trust-based situation awareness technique where the nodes react for the threat depending on the situation awareness knowledge. This system considers both direct and recommended trust. Each node monitors the interaction among its surroundings and receives a recommendation from the trusted nodes. It used to classify the surrounding nodes and identify the intrusion and take action for detecting threats.

In [17], the authors proposed the new trust in RPL using the trust metric to improve RPL security. This system used selfishness, energy, and honesty trust metrics to enhance security. In [18], the authors present a trust and reputation model for the large number of sensor networks in IoT/CPS. Trust established among the nodes with collaboration. This trust mechanism detects the untrustworthy nodes in the network. They used fuzzy theory based trust and reputation mechanisms for IoT/CPS environments which analyze the global trust relationship and local trust relationship. This secure routing approach prevents several attacks via the dynamic replaying of routing information.

In [19] authors proposed the solution to mitigate black hole attack and selective forwarding attack in MWSN(Medical Wireless Sensor Networks in the IoT. They provide the solution with the cryptography hashes and also use threshold-based analysis and neighborhood watch to identify and rectify the selective forwarding and black hole attacks. In [20] authors developed a cryptography solution for version number and rank attack. This system avoids Version Number attack and falsifies the Rank by the malicious nodes. Version number attack makes a load on energy and it consumes more energy, to provide a solution for this attack they created a hash chain, and also the member of this chain also creates the rank chain. In[21], the authors proposed cryptography methods to protect against internal attacks like rank Spoofing and rank replay. It is based on topological authentication. They round trip messages to validate the upward path. The child node sends an authentication message after it receives a message from its parent node. Each parent node checks the child node rank from the testing message. Upward node checks whether the child rank is greater than its rank and also checks the difference of the rank.

The proposed work differs from the existing research work mentioned above. This model used a Q-learning algorithm to choose the trusted IoT nodes for communication. This model ignores the malicious nodes which perform a black hole attack. Once, malicious node is identified, then it will not be involved in any routing operation. Only trusted nodes involved in the routing operation, in this way security, is ensured in the IoT environment.

4. Reinforcement Learning based Trust Model(RLTrust Model)

The proposed model used Q-learning to construct the trust model for the IoT environment. In a highly distributed IoT environment, the devices that successfully joined the network may change its behavior at any time to perform its specific goal, because this node is seized by an attacker. Selecting the trust metric is important in a behavior-based trust model. Because it depicts the past behavior of the node's performance(including communication, routing, data processing, etc) which determines the node's future behavior. Each node in the network uses a trust metric to classify and isolate the malicious nodes. Once malicious nodes are isolated, they are not used for communication anymore. This model is mainly designed to mitigate the black hole attack.

4.1 Network Model Assumptions

The trust model developed with the following underlying assumptions.

1. The Network model is based on the pure distributed Internet of Thing(IoT) environment. There is no centralized trusted node in the network, therefore each node should maintain its neighbor node's trust value for communication.

2. The IoT nodes are heterogeneous which have different capabilities in terms of energy, processing, memory, etc.

3. Restricted Resources: Battlefield Things are small in size and their memory capacity, energy also limited. It may get drained due to sensing, monitoring, updating, and processing capacity. These things are compromised by an adversary.

4. Dynamic Topology: Battlefield Things may leave or join any network at any time.

5. There are two types of nodes in the example network scenario: Trusted and Malicious. The trusted node performs well in terms of forwarding data packets, cooperativeness, etc. A malicious node performs malicious activities to disrupt the main network topology. It drops the data packets to degrade the network performance.

4.2 Adversary Model

In this paper, the behavior of the IoT nodes is considered as malicious, if the node performs data packet drop attack. In the IoT environment, these attacks cause severe problems.

Black hole Attack In this kind of attack, the misbehaving nodes drop all the data packets that are supposed to forward to their neighbor nodes[22].

Figure 1. Sample RPL Network with No Attacks

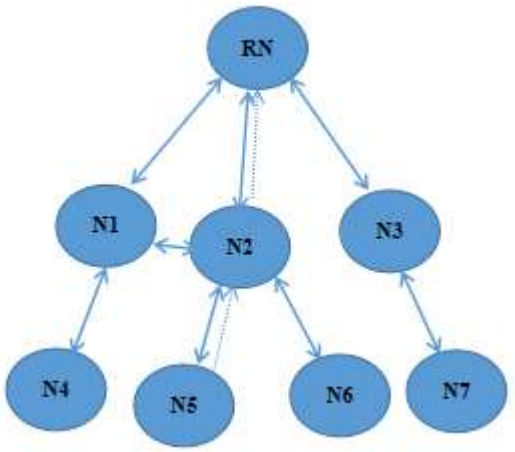

Figure 2. Sample RPL Network with No Blackhole Attacks

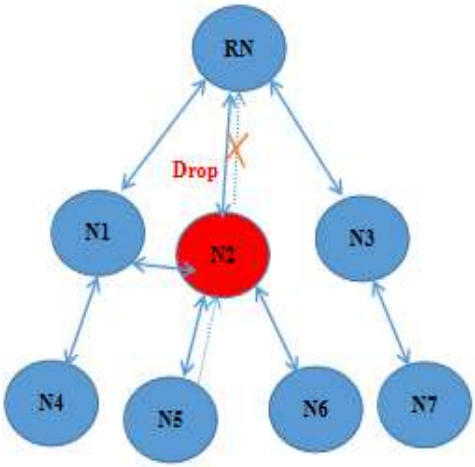

Figure. 1 illustrates the example network scenario in the IoT environment without any attacks. All IoT nodes involved in the network are trusted and authenticated nodes.

Figure.2 shows the example battlefield environment with a black hole attack. The node N2 launches this attack which drops the data packets that are forwarded through this node.

4.3 Design of RLTrust

The design of RLTrust contributes two important components: one is trust computation and another one is identifying node' s behavior.

4.3.1 Trust Computation

The proposed model uses both direct and recommendation trust for trust computation. Direct trust is derived from the neighbor nodes, it is the first-hand information and it can be obtained easily. Indirect trust is second-hand information that is derived from other trustworthy third party nodes. Recommendation

trust is an essential feature in any trust computation system[23].

To obtain accurate trust for each node, the proposed model uses both social and Quality of Service trusts. QoS trust means the faith of the node that it can transfer the data packets to the destination nodes. These trust metrics can be received from the communication and information networks[24] which is measured in terms of delivery ratio, power consumption, bandwidth, average delay, etc. The proposed model considers the packet delivery ratio to compute the QoS of a node. Social trust refers to the social connection between the IoT device owners, it is measured in terms of intimacy, honesty, centrality, etc[25] and it is derived from social networks. This model considers honesty and recommendation to measure the social trust of the node. Honesty is one of the social metrics that is computed based on the previous interaction. The recommendation metric measures the number of correct recommendations given by the particular node.

In addition to that, each node in the IoT network also maintains the friendship list which is based on the successful interaction. For example, node i interact with node j, if the interaction is successful, then node j is added on the friend list of node i.

To estimate the Direct Trust(DT), the proposed model aggregates packet delivery ratio, honesty, and opinion trust properties. Indirect Trust(IT) is received from the common friends of two nodes. Node is selected for routing operation based on the Composite Trust(CT) which combines both direct and indirect trust.

In the IoT environment, selecting the route path is critical to transfer the packet from source to destination. To optimize the routing path routing trust is calculated. If the routing trust below the determined value then the source node selects another trust route. The Reinforcement Algorithm is used in this model which is a powerful alternative to handle network conditions as they present in the real world. It provides trust-based optimize routing in the IoT environment. In the initial stage RL agent(node) randomly selects the neighbor node for interaction, after an interaction it gives reward for the node. In this model delayed reward is given based on the direct and recommendation trust.

4.3.1.1*Direct trust (DT)*

For any trust-based model, it is essential to collect the data from neighbor's nodes for trust computation. Direct Trust computed through direct interaction. Various types of metrics are needed to calculate the direct trust of neighbor nodes. In the proposed model, packet delivery ratio, honesty, and recommendation are used to evaluate the direct trust value. To aggregate, all these metrics weighted linear equations are used in this model. Node A evaluates the direct trust value of node B as follows.

$$DT_{A,B}(t)=w1.PDR(t)+w2.H(t)+w3.O(t) \quad \text{----------(8)}$$

$w1+w2+w3=1$

PDR(t)- Packet Delivery Ratio at ‘t’ time.

H(t)- Honesty(based on number of successful interaction)

O(t)-Opinion (based on number of correct recommendation provided by the node B)

Packet Delivery Ratio(PDR)

The Packet Delivery Ratio (PDR) is computed as the ratio between the total number of transmitted packets and the total number of the received packets. The following mathematical notation used to compute the PDR.

$$PDR_{A,B}(t) = \Sigma TPT(t) / \Sigma TPR(t) \quad \text{----------(5)}$$

TPT- Total amount of Packets Transmitted by node B at ‘t’ time.

TPR- Total amount of Packets Received by node B at ‘t’ time.

Honesty(H)

All nodes in the network calculate its neighbor node’s honesty value based on the ratio of successful and failure interaction between them. For example, node i establish the node j honesty value by direct observation. This value is calculated using the beta function as follows.

$$H_{A,B}(t)= SI(t)+1 / SI(t)+FI(t)+2 \quad \text{----------(6)}$$

SI- Total Number of Successful Interaction at ‘t’ time.

FI- Total Number of Failure Interaction at ‘t’ time.

Where the numerator has '+1' and the denominator has '+2' which indicates that at least two trials were observed out of which one was 'successful' and other was 'failure' according to Laplace law.

Opinion(O)

It is a social trust metric that measures the number of correct opinions provided by the neighbor nodes. For example, node i receives an opinion about node k by node j. Assume node j provides high trust value for node k. Based on this opinion, node i interacts with node k. If the interaction is successful, then node j has given correct opinion otherwise false opinion. If the node gives the correct opinion then the correct opinion count will be increased. This metric is measured as the proportion between the total number of correct opinions and total opinions.

$O_{A,B}(t)= \Sigma CO(t)/\Sigma TO(t)$ ----------(7)

CO- Total number of correct opinions given by B to A at 't' time.

TO- Total number of opinions given by B to A at 't' time.

4.3.1.2 Indirect Trust(IT)

It is computed from the recommendation trust provides by the neighbor nodes. However, compromised nodes may launch recommendation attacks like self-promoting, ballot stuffing, bad-mouthing, etc. To deal with these attacks, the recommendation is received only from the trustworthy nodes in uni-cast mode[26].

Assuming the common friends are trustworthy nodes in the proposed model. Nodes request a recommendation to common friends in uni-cast mode instead of broadcasting the request to all nodes because it consumes less energy. and also it reduces the computation process by avoiding recommendations from unknown or malicious nodes. So, this system does not require any filtering technique to select the recommendation trust.

Common Friend(CF)

Each node maintains the friend list and it dynamically updates based on its interaction. In this model, nodes receive recommendations only from the common friends between two nodes. For example node A wishes to interact with the B node than node A requests and receives recommendations only from common friends between A and B. The following formula used to identify the common friends between nodes A and B.

$CF=A\cap B/A\cup B$ ----------(8)

Where A represents the friends of node A and B represents the friends of node B. $A\cap B$ depict the number of common friends, $A\cup B$ denotes the total number of friends of node A and B. CF reflects the ratio between common friends of A and B and Total friends of A and B.

Indirect Trust Computation

The arithmetic mean is used to calculate indirect trust values. For example, common friends of node A and B are m1, m2 then the node A calculates the indirect trust value for B using the following equation.

$$IT_{A,B}(t)= \sum_{i=1}^{n} DT_{mi}(t) /n \qquad \text{----------(9)}$$

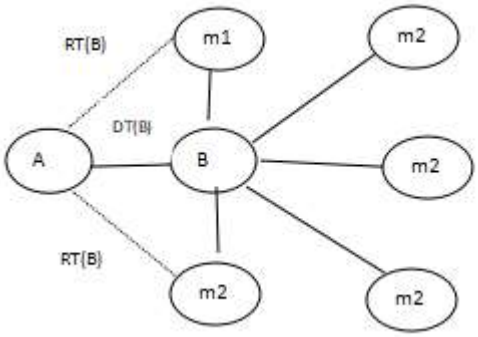

Figure 3. Indirect Trust Computation

Figure 3. shows that nodes A and B have two common friends node m1 and m2. node A request and receive recommendation only from node m1 and m2 which transfer its direct trust as recommendation trust to node A.

4.3.1.3 Composite Trust(CT)

Composite Trust is estimated using the DT and IT. The following equation used to compute the CT.

$CT_{A,B}(t)=w1DT_{A,B}(t)+w2IT_{A,B}(t)$----------(10)

Where w1+w2=1

CT is given as a reward in equation(12).

4.3.2 Q-Learning Based Trust Routing In IoT Environment

Q-learning is the model-free technique of Reinforcement learning(RL), it receives rewards from the environment. To solve the RL problem in a practical way, Markov Decision Process(MDP) is needed. In this model, state-space S refers to the set of neighbor nodes(state), action A refers to selecting the next forwarder(node) for the routing(RL Action) by an agent. The node that currently holds the packet is considered as an agent. Each node in the network is considered as an independent learning agent. T is the transition function that relates to each state, actions, and events which occur when a data packet is transferred from one node to another. R is the reward function obtained from the environment as feedback based on its performance that is needed to update the Q-value. Over time the agent learns from the original action and selects the most trusted node.

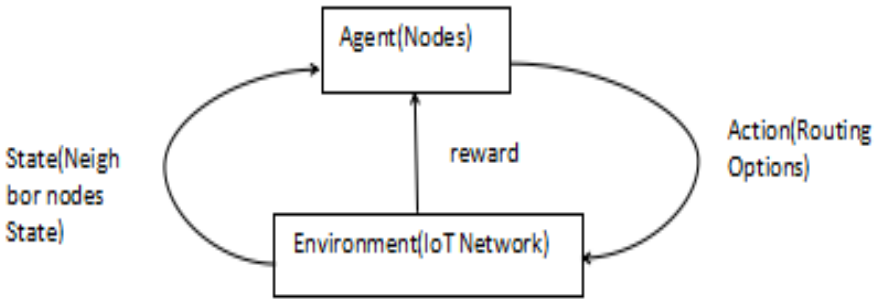

Figure 4. The Q-Learning task in Trust-based routing in the IoT environment

Q-learning based routing algorithm is best suitable for a distributed environment because an initial level deployment of all nodes in the IoT environment does not get enough information to evaluate the neighbor nodes. So, trial and error mechanism is required to learn the behavior of the neighbor nodes. This is possible in Q-Learning.

In the proposed model each node in the IoT environment maintains the Q-table which contains the information about its one-hop

neighbor. This model primarily takes three metrics to calculate direct trust that is PDR, honesty, and recommendation. Assuming, if the node gives correct recommendation and performs an expected task that means successful interaction(forwarding data packets) then the node is trusted. Indirect trust values received form common friends it avoids untrustworthy recommendation. Direct Trust(DT) and Indirect Trust(IT) are aggregated to calculate Composite Trust(CT).

Q- the learning rate is defined as follows

$$Q[s, a] = Q[s, a] + \alpha*(R + \gamma*Max[Q(s', A)] - Q[s, a]) \quad \text{----------(11)}$$

Where Reward R is given based on the node's Composite Trust.

Max[Q(s', A)] - selecting the node with maximal Q value among the set of possible nodes.

4.3.3 Identifying Malicious Nodes

The RL learns the neighbor node's state and determines whether the node behavior is malicious or not. Each agent in the network can take two actions, the node can select the trusted node or malicious node.

Level	Threshold	Meaning
1	If Q[s, a]> TH	Trusted Node
2	If Q[s, a]<= TH	Malicious Node

Table 1. Threshold Table

If Q values below the threshold value, then the chosen node is a suspicious otherwise trusted node. when all neighbor nodes are trusted, then the node with the highest Q-value will be selected for routing operation. In a route selection process, the primary attention is to select the most trusted node, by the way, security is ensured. It can be achieved through the reward function. The reward is given to the agent(node) based on its performance. The components in the reward function based on the different sets of trust metrics. Each agent estimates the trustworthiness of the node based on its previous experience and recommendation from its common neighbors.

4.3.4 *Routing Trust (RT)*

It is used to select the trusted route path for forwarding the data packets. For example, the node wishes to transfer the packet to the root node via the intermediate nodes, the node evaluates the routing trust. This trust value is obtained from the intermediate nodes who indirectly connect the source and destination. Total routing trust is computed using Geometric Mean(GM). For example node A and node B are source and destination and m1,m2 are intermediate nodes who indirectly connect the nodes A and B then the following equation is used to compute the Routing Trust.

$$RT_{A,B}(t) = \left[\prod_{i=1}^{n}(DT_{mi}(t))\right]^{1/n} \quad \text{----------(12)}$$

Source node evaluates the Routing Trust, if it is less than the threshold value, then the selected route is not a trusted route. Therefore source select the alternate trusted path for transferring the data packets.

Algorithm 1

Algorithm for Selecting Trusted Node for Routing

1. Initiate α,γ,ε.
2. Assign Q(s, a)=0 Where s ∈ S, a ∈ A
3. Monitor the initial state s
4. Repeat
5. //Select the one-hop neighbor(action a) for transferring the data packets
6. If(node has no DT or IT) then

7. Node randomly selects the one-hop neighbor to transfer the data. //Exploration

8. Compute DT based on the interaction. Calculate CT for the selected node and which is given as a reward in equation 11.

9. Compute Q(s,a) using equation 12.

10. Else

11. choose the node with the maximum Q(s, a)

$$\text{If}(Q(s,a)>=\text{Threshold} \;\&\&\; b= \underset{a \in A}{\text{argmax}}\; Q(s, a)$$

12. Transfer the data packet to the selected node.//action performed

End if

13. // observe next state s' and reward r

14. /*Based on the communication, the node observes new trust metrics and computes Composite Trust for the selected node which is given as a Reward in equation (11). If node performs well then the composite trust will high, the node will get positive reward otherwise node will get negative reward*/

15. //update Q-value

$$\text{Compute } Q[s, a] = Q[s, a] + \alpha*(R + \gamma*\text{Max}[Q(s$$

16. Assign s=s'

17. End if

18. Until termination//To reach the destination

Algorithm 2(Algorithm for selecting trusted route path)

1. Initialize

2. Repeat

3. Select the node with maximum Q-value

4. Until to reach the destination node

5. Estimate RT from the nodes with maximum Q-value

6. If RT> Threshold value

7. Select the path for routing

8. Else

9. Select another path for routing

10. End if

Algorithm 1 explains Q learning-based trust routing methods. At the initial stage, every state-action pair Q(s, a) at the Q- table is assigned with 0s. over the period the methods enter into the learning phase, which iteratesuntil to arrive at the target node or termination. In the initial stage, the node does not have any information about one-hop neighbors, so the node randomly selects the neighbor node with the probability of ε(Exploration). when the node gets the experience it selects the node with maximum Q-value(a=argmax(Q(s, a')) with the possibility of 1−ε(Exploitation). The nodes are chosen from the node-set. Algorithm 2 explains source nodes collecting RT from the nodes with maximum Q-values. If Routing Trust(RT) is larger than the predetermined threshold value then the origin node selects the path otherwise select the alternate path.

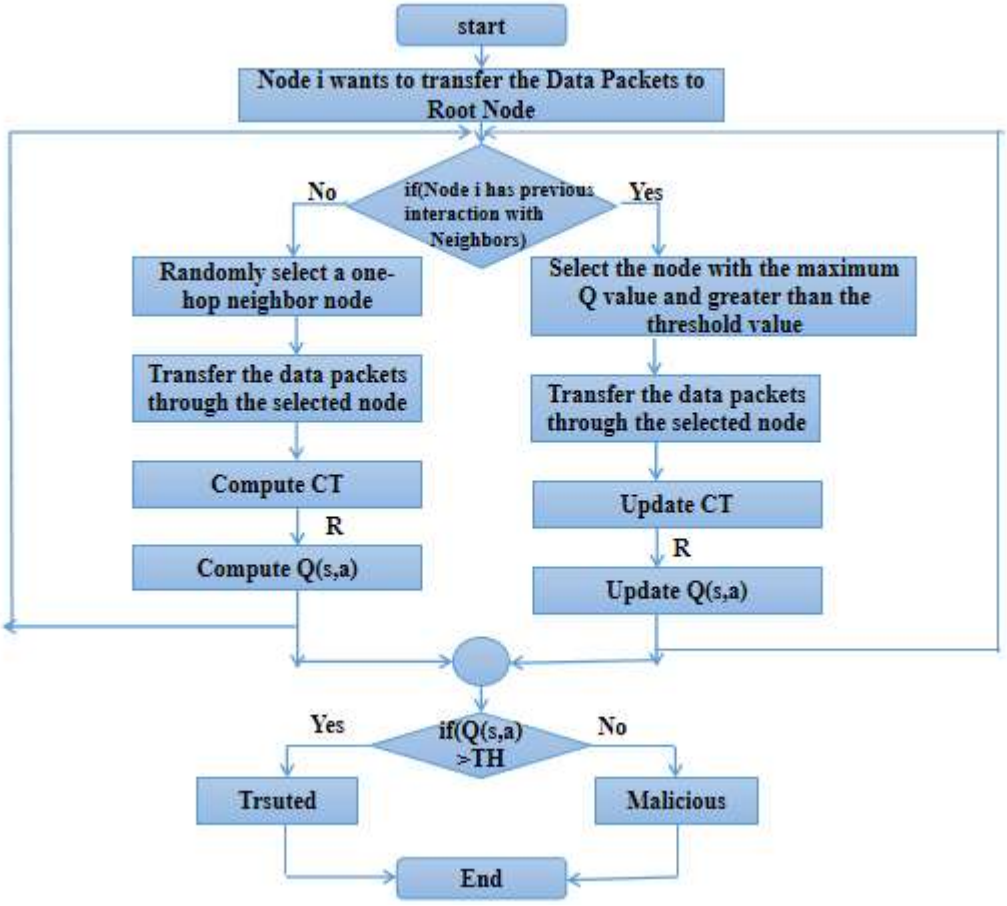

Figure 5. Overall Structure of RLTrust Model

5. MATHEMATICAL ANALYSIS

For simplicity, the sample network only consists of a unidirectional path. The network contains the 8 Nodes, at any given time any one of them can act as the agent. Here, Source node N1 transfers its packets to the target node N8 through the intermediate nodes.

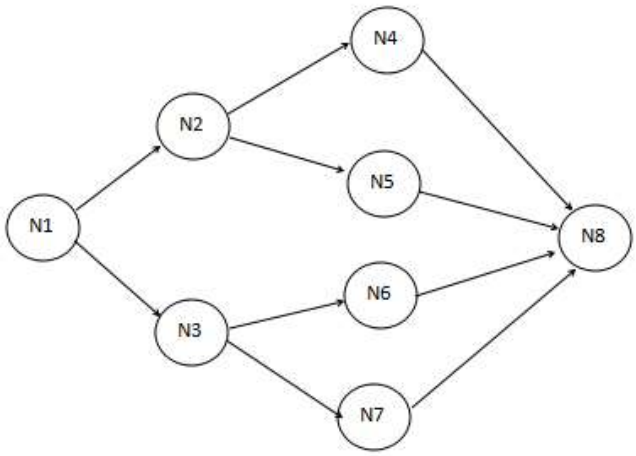

Figure 6. Sample Network Model

Initially, Q-table values will be initialized as 0.

S/A	N1	N2	N3	N4	N5	N6	N7	N8
N1	-	0	0	-	-	-	-	-
N2	-	-	-	0	0	-	-	-
N3	-	-	-	-	-	0	0	-
N4	-	-	-	-	-	-	-	0
N5	-	-	-	-	-	-	-	0
N6	-	-	-	-	-	-	-	0
N7	-	-	-	-	-	-	-	0
N8	-	-	-	-	-	-	-	-

Table 2. Initial stage of Q-value

Sample mathematical calculation for updating Q-value is given below

Assuming α=0.6 and γ=0.8

Exploration

Initial Stage(Exploration with the probability of ε)

The current state is N1 and it can take any 2 different actions. Either it can select N2 or N3. State s ∈ S(N1,N2,N3,N4,N5,N6,N7,N8), a ∈ A{N1,N2}

Initially, all Q-values in the Q-table are initialized as 0.

N1 randomly selects the node N2. The new Q-value is calculated as follows.

Reward Calculation

In the initial episodes, the node does not have a recommendation, honesty trust metric, and recommendation trust. So, In this exploration stage PDR only considers the reward. Suppose N2's PDR of a particular time is 0.8 then the reward is also given 0.8

Q[N1,N2]=Q(N1,N2)+0.6*(0.8+0.8*Max[Q(N2,N4),Q(N2,N5)]-Q(N1,N2))

Q[N1,N2]= 0+0.6*(0.8+0.8*Max[0,0]-0)

Q[N1,N2]= **0.48**

Similarly

Q[N1,N3]=**0.24**

In the same way, Q-value calculated for all the remaining nodes.

Table 2 shows the updated Q-value.

S/A	N1	N2	N3	N4	N5	N6	N7	N8
N1	-	0.48	0.24	-	-	-	-	-
N2	-	-	-	0.54	0.36	-	-	-
N3	-	-	-	-	-	0.48	0.18	-
N4	-	-	-	-	-	-	-	0.54
N5	-	-	-	-	-	-	-	0.48
N6	-	-	-	-	-	-	-	0.36
N7	-	-	-	-	-	-	-	0.48
N8	-	-	-	-	-	-	-	-

Table 3. Q-valueAfter some time

After some Interaction(Exploitation)

N1 selects the node with the highest Q-value. In our example, N2 has the highest Q-value which will be selected for routing. The new Q-value is calculated as follows.

Reward Calculation

Over the period, the node will get all the trust metrics and recommendation trust then the reward is computed as follows.

N1 evaluate N2 trust value as follows

Suppose N1 transfers 100 data packets, N2 forwarded only 80 data packets then the PDR calculated as follows.

PDR (t)= 80 /100=0.8

Assuming N1 has 2 successful and 1 failure interaction with N2 then the honesty metric is measured as follows.

H(t)= 2+1/2+1+2=0.6

Suppose N2 gives 2 correct opinions to N1. Total recommendation 2 then the opinion metric is measured as follows.

O(t) = 2/2=1

Direct Trust

The following weights are assigned for trust metrics

$DT_{N1,N2}(t)=0.6*0.8+0.2*0.6+0.2*1=$**0.8**

Assuming,N1 receives 2 recommendation from its common friend as 0.9 and 0.8 then the RT value is

$IT_{N1,N2}(t)=(0.9+0.8)/2=0.85$

TT is calculated as follows

Assigning w1=0.6 and w2=0.4

$CT_{N1,N2}(t)=0.6*0.8+0.4*0.85=0.82$

The reward values between 0.0 to 1.0.

The trust metric computation based on ratio, so the CT will be the range between 0.0 to 1.0. Here, the Reward is given as 0.82.

Q[N1,N2]=0.48+0.6*(0.82+0.8*Max[Q(0.54,0.36)]-0.48)

Q[N1,N2]=0.48+0.6*(0.82+0.8*0.54-0.48)

Q[N1,N2]=0.9432

In the same way, the remaining Q- values will be updated.

If N1 selects the N3 then the new Q-value as follows

Q[N1,N3]=0.36+0.6*(0.488+0.8*Max[Q(0.48, 0.18)]-0.36)

Q[N1,N3]=0.36+0.6*(0.488+0.8*0.48-0.36)

Q[N1,N3]=0.6672

When an agent(node) selects the maximum Q-value then the node is trusted, because in the proposed model reward is given based on its trust metric. So the agent can select the trusted node and can avoid the misbehaving node. If Q-value is below the threshold value then the node is a malicious agent avoided from further interaction.

6. SIMULATION RESULTS AND DISCUSSION

6.1. Performance Evaluation Metrics

The trust model is evaluated in the Contiki 3.0 OS and the Cooja simulator. The trust 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 byte
Traffic	UDP
Mac Layer	IEEE 802.15.4
Communication Range	50m
RPL Parameter	MinHopRankIncrease=256
Routing Protocol	NBDSTrust, Trust-based RPL, RPL

Table 3. The Simulation Parameters of the Proposed RLTrust Model

6.2 Simulation Results

The evaluation of the RLTrust 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 these attacks on the network protocol. In this regard, increase the percentage of misbehaving nodes and measure the proportion of dropping ratio in RPL.

2. The performance of the RLTrust model is compared with the (Alnasser, A et al., 2017)[27] in terms of Packet Delivery Ratio, End to End Delay, and Throughput.

3. Increase the percentage of malicious nodes and compare the detection accuracy of RLTrust and (Alnasser, A et al., 2017).

Scenario 1: The analysis is performed with the varying number of malicious nodes under normal RPL routing protocol. The observations in figure.4 show that when the number of malicious increases, data dropping also increases in RPL.

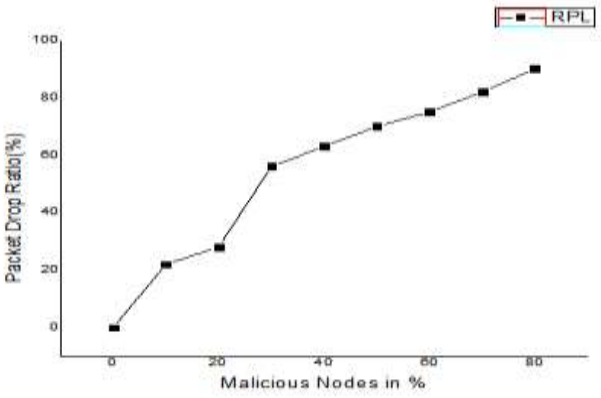

Figure 7. Impact of Blackhole attacks under normal RPL routing protocol

Scenario 2: In this simulation, the performance evaluation of the proposed RLTrust model is compared with the (Alnasser, A et al., 2017) model in terms of delivery ratio, average delay, and throughput

In this simulation, performance metrics such as packet delivery ratio and the end to end delay and throughput are compared (Alnasser, A et al., 2017).

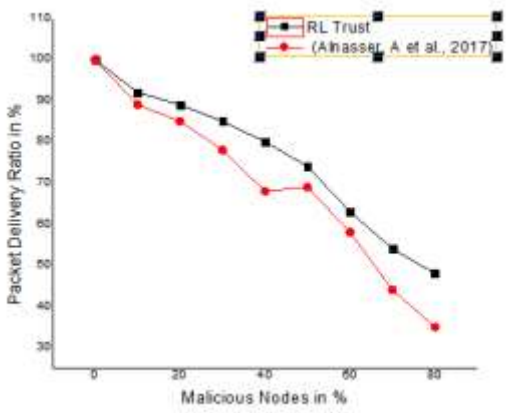

Figure 8. Malicious Nodes vs Packet Delivery Ratio

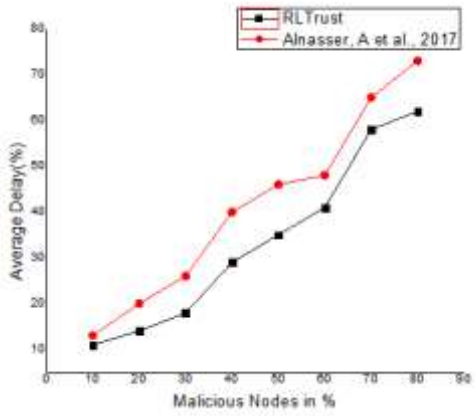

Figure 9. Malicious Nodes vs Average Delay

Delivery Ratio: It is a proportion between the volume of packets forwarded by the source node and the volume of packets received by the destination node. It has significant trust properties to assess the functionality of the RLTrust 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 8. demonstrate the delivery ratio of RLTrust and (Alnasser, A et al., 2017). Results depict the proposed model has the greatest delivery ratio when compared to (Alnasser, A et al., 2017) model. The reason is, (Alnasser, A et al., 2017) model considers a single trust property such as a forwarding ratio to evaluate the trustworthiness of the node, but the RLTrust model considers multiple trust metrics (PDR, H, R) to evaluate the trustworthiness of the node. Due to this multiple trust metric, the proposed model easily detects and removes the malicious nodes which perform the data drop attack. The malicious nodes are not selected for routing, only trusted nodes involved for routing the data packets, thus increasing the packet delivery ratio.

Average delay: It is measured as a mean time needed to transfer a packet from the source to the target node. It is another important metric to measure the functionality of the proposed protocols. The existence of misbehaving nodes in the IoT network increases the delay. Figure 9. depicts the impact on the delay of the two models (RLTrust and (Alnasser, A et al., 2017)) with the varying percentage of the malicious nodes. It shows that the RLTrust model delay is lesser than the (Alnasser, A et al., 2017) because the proposed model selects the trusted nodes accurately for routing than the Trust-based RPL, thus avoiding malicious nodes and decreasing the average delay.

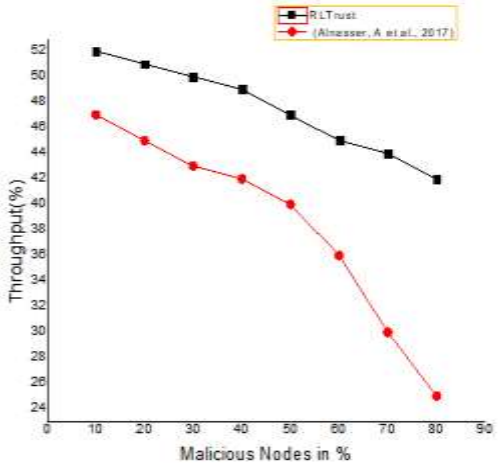

Figure 10. Malicious Nodes vs Throughput

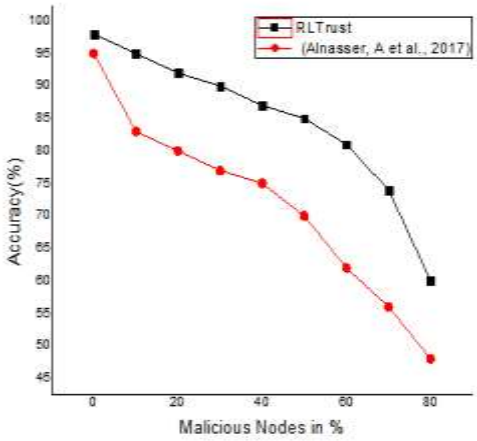

Figure 8. Malicious Nodes vs Accuracy

Average Throughput: The total volume of packets transferred in a certain time or an average number of successful information transferred per second over a communicating transmission channel is called throughput. In general, it is represented in bits per second (bits/s or bps).

Throughput= (Total amount of Packets Received) / ((End Time ±begin Time))

As in Figure 10, when compared to average throughput, the proposed model is greater than the RPL and trust-based RPL. RLTrust model effectively identifies the misbehaving nodes in the initial stage than the (Alnasser, A et al., 2017). So average throughput is increased in the proposed system compared to the (Alnasser, A et al., 2017).

Scenario 3: Figure 11. depicts the detection accuracy between the RLTrust model and (Alnasser, A et al., 2017) model with the varying percentage of misbehaving nodes.

when the percentage of misbehaving nodes increases, the accuracy of both RLTrustand (Alnasser, A et al., 2017) model degrades. When compared to the (Alnasser, A et al., 2017) model, the accuracy of the RLTrust mode is high. Because the proposed model uses the Q-learning algorithm to select the trusted node with the high trust value. It effectively identifies the malicious nodes, thus increasing the detection accuracy.

7. CONCLUSION

The proposed trust-based routing model using Q-learning. It is the model-free technique of reinforcement learning. Because of dynamic topology and resource constraints, IoT applications are susceptible to many attacks. Q-learning based routing ensures security in the IoT environment by implementing a malicious node identifying mechanism. In the proposed model, Q-value calculated for each neighbor node. The nodes with the maximum Q-value and greater than the threshold value are trusted nodes and the nodes those Q-values less than the threshold values are malicious nodes. This can be achieved through the reward. The proposed model uses the trust metric to provide a reward. Each node calculates its one-hop neighbor nodes DT using packet delivery ratio, honesty, and recommendation trust metrics. Node also receives a recommendation from common friends and computes IT. CT is calculated from the DT and IT. This CT value is given as a reward for each node after an action taken. So this Q-learning based routing algorithm selects the most trusted node which ensures the security and it avoids the misbehaving nodes for routing. Packet Delivery ratio is one of the trust metrics in trust computation, so the system can identify the black hole attacks. The proposed trust model has been embedded into RPL and the performance of the RLTrust is evaluated using a cooja simulator. The performance evaluation shows the effectiveness of the RLTrust with varying percentages of malicious nodes as compared to existing one.

REFERENCES

[1] Medaglia, C. M., &Serbanati, A. (2010),."An Overview of Privacy and Security Issues in the Internet of Things" *The Internet of Things*, 389–395. doi:10.1007/978-1-4419-1674-7_38 2.

[2] Basheer, Shakila; Bivi, S Mariyam Aysha; Jayakumar, S; Rathore, Arpit; Jeyakumar, Balajee. Machine Learning Based Classification of Cervical Cancer Using K-Nearest Neighbour, Random Forest and Multilayer Perceptron Algorithms,Journal of Computational and Theoretical Nanoscience, Volume 16, Numbers 5-6, May 2019, pp. 2523-2527(5).

[3] Lu Tan, &Neng Wang. (2010).," Future internet: The Internet of Things." *2010 3rd International Conference on Advanced Computer Theory and Engineering(ICACTE)*. doi:10.1109/icacte.2010.5579543.

[4] Sharma, A., Pilli, E. S., Mazumdar, A. P., &Govil, M. C. (2016).," A framework to manage Trust in Internet of Things," *2016 International Conference on Emerging Trends in Communication Technologies (ETCT)*. doi:10.1109/etct.2016.7882970.

[5] Deogirikar, J., &Vidhate, A. (2017).."Security attacks in IoT: A survey," *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*. doi:10.1109/i-smac.2017.8058363

[6] Sultana, H Parveen; Shrivastava, Nirvishi; Dominic, Dhanapal Durai; Nalini, N; Balajee, J.M. Comparison of Machine Learning Algorithms to Build Optimized Network Intrusion Detection System, Journal of Computational and Theoretical Nanoscience, Volume 16, Numbers 5-6, May 2019, pp. 2541-2549(9).

[7] lshehri, M. D., & Hussain, F. K. (2015).," A Comparative Analysis of Scalable and Context-Aware Trust Management Approaches for Internet of Things," *Lecture Notes in Computer Science, 596–605*. doi:10.1007/978-3-319-26561-2_70.

[8] T. Winter, "RPL: IPv6 Routing Protocol for Low-Power and Lossy Networks",https://tools.ietf.org/html/rfc6550, 2012.

[9] Basheer, Shakila; Mathew, Rincy Merlin; Ranjith, D; Sathish Kumar, M; Praveen Sundar, P. V; Balajee, J. M. An Analysis on Barrier Coverage in Wireless Sensor Networks, Journal of Computational and Theoretical Nanoscience, Volume 16, Numbers 5-6, May 2019, pp. 2599-2603(5).

[10] Wen, T.-H., Lee, H., Su, P., & Lee, L.-S. (2013).,"Interactive spoken content retrieval by extended query model and continuous state space Markov Decision Process," *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, doi:10.1109/icassp.2013.6639326 .

[11] Vinoth Kumar V, Karthikeyan T, Praveen Sundar P V, Magesh G, Balajee J.M. (2020). A Quantum Approach in LiFi Security using Quantum Key Distribution. International Journal of Advanced Science and Technology, 29(6s), 2345-2354.

[12] Dias, G. M., Nurchis, M., &Bellalta, B. (2016)," Adapting sampling interval of sensor networks using on-line reinforcement learning," *2016 IEEE 3rd World Forum on Internet of Things (WF-IoT)*. doi:10.1109/wf-iot.2016.7845391.

[13] Lahbib, A., Toumi, K., Elleuch, S., Laouiti, A., & Martin, S. (2017). Link reliable and trust aware RPL routing protocol for Internet of Things. 2017 IEEE 16th International Symposium on Network Computing and Applications (NCA). doi:10.1109/nca.2017.8171360.

[14] Ben Saied, Y., Olivereau, A., Zeghlache, D., & Laurent, M. (2013). *Trust management system design for the Internet of Things: A context-aware and multi-service approach. Computers & Security, 39, 351–365*. doi:10.1016/j.cose.2013.09.001 .

[15] Hellaoui, H., Bouabdallah, A., &Koudil, M. (2016). TAS-IoT: Trust-Based Adaptive Security in the IoT. 2016 IEEE 41st Conference on Local Computer Networks (LCN). doi:10.1109/lcn.2016.101 .

[16] Glowacka, J., Krygier, J., &Amanowicz, M. (2015). *A trust-based situation awareness system for military applications of the internet of things. 2015 IEEE 2nd*

World Forum on Internet of Things (WF-IoT). doi:10.1109/wf-iot.2015.7389103

[17] Djedjig, N., Tandjaoui, D., Medjek, F., &Romdhani, I. (2017). New trust metric for the RPL routing protocol. 2017 8th International Conference on Information and Communication Systems (ICICS). doi:10.1109/iacs.2017.7921993 .

[18] Chen D, Chang G, Sun D, Li J, Jia J, Wang X (2011) TRM-IoT: a trust management model based on fuzzy reputation for internet of things. Comput Sci Inf Syst 8(4):1207–1228. https://doi.org/10.2298/CSIS110303056C.

[19] Mathur, A., Newe, T., & Rao, M. (2016). Defence against Black Hole and Selective Forwarding Attacks for Medical WSNs in the IoT. Sensors, 16(1), 118. doi:10.3390/s16010118

[20] A. Dvir, T. Holczer, and L. Buttyan. Vera - version number and rank authentication in rpl. In 2011 IEEE Eighth International Conference on Mobile Ad-Hoc and Sensor Systems.

[21] H. Perrey, M. Landsmann, O. Ugus, M. Wahlisch, and T. C. Schmidt. ¨ TRAIL: Topology authentication in RPL. In Proceedings of the 2016 International Conference on Embedded Wireless Systems and Networks, EWSN ’16.

[22] Kamble, A., Malemath, V. S., & Patil, D. (2017). Security attacks and secure routing protocols in RPL-based Internet of Things: Survey. 2017 International Conference on Emerging Trends & Innovation in ICT (ICEI). doi:10.1109/etiict.2017.7977006.

[23] Xia, H., Jia, Z., Li, X., Ju, L., & Sha, E. H.-M. (2013). Trust prediction and trust-based source routing in mobile ad hoc networks. Ad Hoc Networks, 11(7), 2096–2114. doi:10.1016/j.adhoc.2012.02.009.

[24] J. H. Cho, A. Swami, and I. R. Chen, “A Survey on Trust Management for Mobile Ad Hoc Networks,” IEEE Communications Surveys & Tutorials, vol. 13, no. 4, 2011, pp. 562-583.

[25] Guo, J., Chen, I.-R., & Tsai, J. J. P. (2017). A survey of trust computation models for service management in internet of things systems. Computer Communications, 97, 1–14. doi:10.1016/j.comcom.2016.10.012.

[26] Labraoui, N. (2015). A reliable trust management scheme in wireless sensor networks. 2015 12th International Symposium on Programming and Systems (ISPS). doi:10.1109/isps.2015.7244958.

[27] Alnasser, A., & Sun, H. (2017). A Fuzzy Logic Trust Model for Secure Routing in Smart Grid Networks. IEEE Access, 5, 17896–17903. doi:10.1109/access.2017.2740219 .