Trust-based Routing Optimization using Learning Automata in Wireless Sensor Network

Maryam Hajiee[1], Mehdi Fartash[2*], Nafiseh Osati Eraghi[3]
1-Department of Computer Engineering, Arak Branch, Islamic Azad University, Arak, Iran.
Email:mhajiee2016@gmail.com
2-Department of Computer Engineering, Arak Branch, Islamic Azad University, Arak, Iran.
Email: m-fartash@iau-arak.ac.ir (Corresponding author)
3-Department of Computer Engineering, Arak Branch, Islamic Azad University, Arak, Iran.
Email: n-osati@iau-arak.ac.ir

Received: March 2021 Revised: June 2021 Accepted: August 2021

ABSTRACT:
The use of wireless sensor networks is becoming more and more important due to the COVID-19 pandemic and the living conditions of human beings today. The three main goals in designing this type of network are to reduce energy consumption, choose the shortest route and choose a reliable route for data transmission. In this paper, these three goals are considered in routing. Due to the fact that this type of network is exposed to many attacks, identifying malicious nodes and removing them creates security in this type of network. This paper presents an energy-aware and trusted-based routing method using learning automata and an evaluation function. Learning automata identifies trusted nodes (to send data) and malicious nodes using the corresponding evaluation function. The evaluation function considers the residual energy, the node's trust and the number of hops to the sink parameters. Thus, the data reaches its destination in a safe and reliable way. The evaluation results of the proposed method show an improvement in the performance of this method compared to other relevant methods.
.

KEYWORDS: Wireless Sensor Network, Trust, Learning Automata, Security.

1. INTRODUCTION

WSN is a network composed of a set of nodes that can vary from a few to several hundred sensors. In these networks, each node is connected to another node (or several other nodes). The wireless sensor network interacts strongly with the physical environment. It receives environmental information through sensors and reacts through actuators. Communication between nodes is wireless. Each node operates independently and without human intervention and it is typically very small physically and has limitations in processing power, memory capacity and power supply. These limitations create problems that are the source of much of the research challenge in this area. Wireless sensor network is used in various fields such as emergency response [1], healthcare monitoring [2], military and agriculture [3], and environment monitoring and smart power grid [4]. Another important application of the wireless sensor network, common during the COVID-19 pandemic, is to support patients and the elderly when hospitals are full of patients with COVID-19 and their hospitalization increases their risk of developing the virus. Using sensors and wireless sensor networks can help a great deal in treating them preventing the spread of COVID-19. Various sensors such as those that monitor blood pressure and temperature are designed to detect medical signals [5]. These wireless sensors can be implanted in a patient's body or worn on body. In cases of emergency, such as when receiving abnormal information from the ECCG, a warning will be sent to the care team and appropriate action will be taken according to the severity of the alert. These networks have been proven to be suitable for emergencies because it sends information to physicians who are ready for immediate treatment of the patient [6-8]. Patients with WBANs do not normally need to see a doctor physically, reducing the number of patients in hospitals. The sensors must be able to provide real-time and accurate patient information. Incorrect timing and information can lead to patient death. As a result, creating security in this type of network so that the data sent and received by this type of network is safe and reliable is of vital importance and necessary. Due to the openness and insecurity of the communication channel between the sensor nodes, the network is prone to many attacks. In addition, central communications are quite

Paper type: Research paper
DOI: https://doi.org/10.52547/mjee.15.4.87
How to cite this paper: M. Hajiee, M. Fartash and N. Osati Eraghi, **"Trust-based Routing Optimization using Learning Automata in Wireless Sensor Network"**, *Majlesi Journal of Electrical Engineering*, Vol. 15, No. 4, pp. 87-98, 2021

complex due to dynamic topological structure [9]. All kinds of internal or external attacks such as Black hole, Gray hole, Node capture, Eavesdropping, Worm-hole, Sybil attack, Sink-hole and Denial of service threaten the wireless sensor network. Therefore, creating security in this type of network is very important. In the past, methods such as Cryptography, Authentication and Hash functions [10-12] were used to create security in these networks. However, these techniques are not very effective and cannot separate Selfish and Malicious nodes from the honest nodes due to their low computational capacity, memory and power. Thus, today concepts such as Trust and Reputation are used to increase security in this type of network. Humans use the concept of trust in human relationships. Similarly, two nodes communicate with each other based on the degree of trust they have in each other. In [13], trust is defined as the degree of trust that we can consider for a node's future behavior, which is based on the node's past behaviors or as the degree of trust that exists between two entities. The trust management system has been implemented in various security applications including secure protocol [14], secure data aggregation [15], trusted routing [16], and intrusion detection system [17].

One way of increasing security in wireless sensor networks is to design a secure mechanism for routing in this network. Selecting an optimal and secure path to transfer sensed data to the sink node will both increase network security and extend network life time. In this paper, in order to achieve the goal of secure routing, a novel solution for detection of malicious nodes in the wireless sensor network, using trust evaluation method and learning automata is presented. Thus, when a malicious node is detected in the network, the data is neither sent to that node nor received from that node, increasing network security. The learning automata is a single model suitable for solving learning problems in random and unknown environments. The automaton selects an action from its set of authorized actions based on probability distribution and updates its probability distribution by receiving feedback from the environment. Different types of learning automata are presented for different applications. In this paper, learning automata with variable structure is used. [18].

The rest of the article is outlined as follows: the second section expounds on related research. The third section covers the concept of trust and trust evaluation method, learning automata and network structure and assumptions as well as the proposed method. The simulation results and conclusions are respectively presented in the fourth and fifth sections.

2. LITERATURE REVIEW

Wireless sensor networks have many advantages in human life, nevertheless providing security in this network is faced by challenges such as the distributed nature of these sensors, the limited memory and energy, and the physical attacks that threaten this network. Various methods such as data encryption, key management, IDS methods and trust and reputation-based methods have been proposed to remove these challenges. One of the best ways to put security in a wireless sensor network is to use the concept of trust in establishing security in the network. Due to the importance of trust in establishing security in the sensor network, many researchers have focused on this issue, each trying to improve on previous work. Some articles on the topic of securing wireless sensor network are outlined below.

Ganeriwal et al. [19] provide a reputation-based framework for sensor networks (RFSNs). RFNS uses a watchdog system to observe neighboring node behaviors and uses a beta distribution to distribute reputation values.

In [20], a trust and friendship routing scheme based AODV (Fr-AODV) which detects black hole attacks is presented. Trust evaluation is based on features such as node reputation and node identity. Each attribute is numerically stated and changes during the sending of packets. Fr-AODV also uses the path maintenance mechanism.

Song et al. [21] introduced a dynamic trust calculation method based on several factors. Trust is obtained by combining direct and indirect trusts. This method does not consider the process of updating trusts.

Adnan et al. [22] introduced a protocol called TERP. TERP was developed to support security and intrusion detection based on direct surveillance, indirect surveillance and a factor called increasing the accuracy of security calculations. Challenges associated with TERP include increased overheads due to recommendation sharing, vulnerability to intrusive nodes, and inability to detect some attacks.

Adnan et al. in [23] introduced another protocol called TESRP aimed at increasing protocol performance. The performance of this protocol was developed to improve the challenges and vulnerabilities of the TERP protocol based on beta distribution. Challenges of the TESRP protocol include security vulnerabilities against intruders with the intention of deceiving the intrusion detection system, inconsistencies in reliable assessments, increased overheads and delays in the routing process.

In another study, Adnan et al. [24] introduced a protocol called ESRT with the aim of improving the performance of the TERP and TESRP protocols. In order to achieve the set goals and improve the issues of the two protocols, the performance of the mentioned two protocols was developed by increasing the transaction time interval criterion in intrusion detection

assessments. Interaction paths associated with poor TESRP performance in addition to support for reliability use the single-propagation of path failure packets only on paths to the source and multi-path routing with the aim of using backup paths.

Datta [25] proposes the TLB-AODV protocol to defend against Black hole and Gray hole attacks. In the proposed method, intrusion detection system has been used to estimate the amount of trust. The amount of trust is based on the behavior of the nodes in sending packets. Each node must calculate the amount of trust its neighbors have. It also uses the amount of indirect trust it receives from other neighbors to calculate trust. One of the main limitations of this method is the use of intrusion detection systems to calculate trust.

In [26], R-AODV protocol is proposed for identifying malicious and defective nodes in packet transmitting. In this method, the trust of a node is obtained based on statistical data concerning the rate of packets that are sent correctly. Each node gains the trust of its neighbors only on the basis of its direct observations. Routing in this method is based on the degree of trust of nodes and the amount of end-to-end delay between nodes. A rout maintenance routine is called when a faulty or malicious node is detected in an active path.

Feng et al. [27] introduced a protocol called BTRES to investigate and identify internal attacks and provide a method for preventing these attacks. BTRES was developed based on the performance of the intrusion detection system and expanded its performance in wireless sensor networks based on the advantages of this system and the use of the capabilities of the beta distribution function.

In order to learn and detect unknown attacks, a neural network method based IDS is presented in [28]. Here, the Markov model is used to learn and analyze time-related changes.

In this study [29], Feng et al. introduced a protocol called EDTM. EDTM extends intrusion detection based on direct agents and develops calculations of how sensors operate based on direct monitoring.

In [30], an ant algorithm is used and two models (BTRM-WSN) are presented combining the Peer Trust System to increase performance. The results show that this model is more accurate in detecting reliable nodes and thus increases the level of security.

In [31], the authors introduced a protocol data-based learning automata (CADA) that not only ensures secure data transmission but also tracks malicious movements.

Moreover, in [32], ant algorithm in the Mobile network Ad hoc was used to generate a QOS Mobility Aware ACO Routing Protocol (QMAA) to improve the performance of QOS.

In [33], an intrusion detection system in wireless sensor network with a learning automata approach is presented. This approach is based on three components including automata, environment, and update of the action selection probability function. The S-Model approach was also used to solve the problem.

In [34], an approach to detect selfish nodes is used. In the proposed approach, a control data packet is used to identify selfish nodes. Thus, if a packet reaches the middle node during its transmission from a source node to a destination node, it is selfish. However, if that node does not send the packet and the packet does not reach its destination, the source node must re-send the packet. Finally, if the number of re-sent packets exceeds a threshold, it indicates that there are selfish nodes in the network.

In [35], a combination of Deterministic Finite Automata (DFA) and Particle Swarm Optimization is used to detect data intrusion and send data in a secure rout. In this article, LD 2 FA (Learning Dynamic Deterministic Finite Automatic) is introduced, which detects intrusion by examining the packet, data and path, and as a result, the data is transferred in a convenient and secure path.

In [36], DSR is designed to secure the routing protocol mechanism and uses "path rater" and "watchdog" modules. This method can be used in routing protocols in which the origin determines the path of the packets. In the proposed trust management method, a credit system is added to the watchdog and path rater method and this credit system maintains a blacklist in the nodes and shares it with friend nodes. DSR is a combination of direct and indirect reputation and operational reputation based on observation of behaviors.

[37] Provides a secure on-demand routing protocol in case networks which prevents the manipulation of secure paths involving healthy nodes as well as many denial-of-service attacks.

3. THE PROPOSED METHOD

This paper presents a method for secure routing in wireless sensor networks with the approach of identifying malicious nodes. The presence of malicious nodes in the network reduces the throughput of the network. As a result, by identifying these malicious nodes, the other nodes neither send nor receive data from them. In this paper, we have used a method to detect trusted nodes with the help of a learning automata and present an evaluation function to select a trusted neighbor to send data to. In this section, first the concept of learning automata is expressed, and then the concept of trust and the method of evaluating the trust of sensor nodes are expressed, and finally the proposed method is described.

3.1. Learning Automata

Learning is a process that is necessary to change the behavior of organisms to adapt to the environment. Thus, with the advancement of technology to improve the performance of technologies, understanding the principles of learning of living organisms and its stages and providing a methodology to put these principles in a system is necessary. To date, the learning automata model has been used in numerous studies such as environmental monitoring, firefighting and rescue, distribution of sensor nodes in the network environment, detection and response to network attacks. A learning automata is a decision maker that operates in a [random] environment and, based on the response it receives, updates its strategy for action selection. The purpose of designing a learning automata is to identify how to choose action based on past experiences (actions and responses). Learning occurs when the environment changes over time and there is very little knowledge regarding that environment. At each stage, the learning automata selects an action from its set of actions and applies it to the environment. Each action is selected based on the probability assigned to it. The selected action is evaluated by a random environment and the evaluation result is delivered to the learning automata in the form of a positive or negative signal with a fixed indefinite probability distribution. The learning automata, based on the feedback of the environment, updates the probability vector of its set of actions and thus is affected by the feedback of the environment to select its next action. During this process, the automata learns to choose the optimal action. The automata learning algorithm determines how to use the environment feedback to the automata selective action. A learning automata consists of the following two parts: a random automata with a limited number of operations and a random environment, and a learning algorithm by which the automata learns the optimal operation. A random automata is defined as five SA≡ {α, β, F, G,Q} where α = {α_1, α_2… α_r} is set of automata operations (r is number of automata operations), β = {β_1, β_2,…, β_m} is set of automata inputs, F≡∅ × β is new state output function, G≡∅ → α is Output function that maps the current state to the next output, and ∅ (n) = { $\emptyset_1$, $\emptyset_2$,…, ∅k} is the set of internal states of the automata at moment n. Functions F and G map the current state of the input to the output of the next (next operation) automata. If the maps F and G are deterministic, it is called a deterministic automata. If the maps F and G are random, the automata is called a random automata [38]. Similarly, random learning automata can be represented by the quadratic LA≡ {α, β, P, T} where α ≡ {α_1, α_2,…, αr}is the set of automata operations (r is the number of automaton operations), β ≡ {β_1, β_2,…, β_r} is the set of automata inputs, p≡ {p_1, p_2,…, p_r}is the vector of probability of automatic operations and T≡p (n + 1) = T [α (n), β (n), p (n)] is a learning algorithm [39]. The main idea in learning algorithms is that if the automata selects the operation αi at time n and receives the desired response from the environment, the probability of p_i (n) increases and the probability of selecting other operations decreases slightly. For an unfavorable response, p_i (n) decreases and other probabilities increase. In a random learning automata with a fixed structure, the probability of selecting operations is fixed, while in a random learning automata with a variable structure, the probability of selecting operations is updated at each iteration. In a learning automata with a variable structure, the change of action probabilities is carried out based on the learning algorithm.

3.2. Trust Model

Designing a trust evaluation method is very useful for detecting different types of attacks and malicious nodes. To obtain the trust model, the method used was as per reference [24] with some modifications. In this paper, Equation (1) is used to evaluate the trust and identify the malicious nodes which is described in full below. In relation (1), $T_{i,j}(t)$measure the validity and detects the influence of node i in relation to node j, and w1 and w2 are the relative valuation coefficients ($w_1 > w_2$ and $w_1 + w_2 = 1$) where $DT_{i,j}(t)$ is used to evaluate direct observations, $\frac{IT_{i,j}^{k}(t)}{N_j}$ is used to evaluate indirect observations.

$$T_{i,j}(t) = w_1 DT_{i,j}(t) + w_2 \frac{IT_{i,j}^{k}(t)}{N_j} \quad (1)$$

Initially, the network setup to each node in the network is given an initial level of security (0.5) as the average of the initial trust, and then based on the behavioral performance of the node, this rate will increase and decrease according to Equation 1. If the security level of the node decreases below the malicious node detection threshold value, the node will be identified as a malicious node, added to the blacklist and placed in the quarantine network (the package will not be received from the node and will not be sent to that node). The degree of trust of the nodes in the proposed protocol varies from zero to 1 (zero meaning complete distrust and 1 meaning complete trust).

3.2.1. Evaluation of Direct Observations

Node self-observations should be used to calculate direct trust. When node i wants to calculate the direct trust of node j, the amount of direct trust of node j is obtained by monitoring the behaviors of node j. To calculate the direct trust of a node, the concepts of variable credit over time and variable credit with the position of interaction and variable credit with the type of behavior should be used. The variable credit over

time indicates the value of the concept of time in detecting erroneous nodes. In fact, the value of each interaction will vary according to the time of its occurrence as interactions in relation to proximity to the present time will be of greater value and importance. Equation (2) presents how to evaluate and calculate the validity of a variable over time. The relation $Change\ Credit_{time(i,j)}$ shows the validity of the variable over time from the evaluator node i to the evaluated node j, n shows nth interaction between the two nodes i and j, $F(n)$, shows the type of behavior of node j in the nth interaction with node i (in case of correct behavior the value of 1 and in case of incorrect behavior the value of zero is assigned to this variable), and z is the number of interactions between the two nodes of i and j.

$$Change\ Credit_{time(i,j)} = \frac{\sum_{n=1}^{z} F(n)\cdot t_n}{\sum_{n=1}^{z} t_n} \quad (2)$$

Variable Credit with Interaction position refers to the position of interaction and its effect on malicious node detection. For instance, control packet and information packet or military data and ordinary data have different values. Therefore, it is necessary to value each interaction in relation to its position. Equation (3) is the developed Equation (2) for evaluating calculations of the variable's credit with the interaction position. Thus, in the provided equation, $Change\ Credit_{time(i,j)}$ is variable valuation over time and the desired interaction position is between evaluator node i and the evaluated node j, σ and τ are relative valuation coefficients (so that $\tau + \sigma = 1$) and Q is the value of the nth interaction.

$$Change\ Credit_{Time,Data(i,j)} = \frac{\sum_{n=1}^{z} F(n).\left[\left(\sigma.1-\delta^{-E.t_n}\right)+\left(\tau.1-\delta^{-Q(n)}\right)\right]}{\sum_{n=1}^{z}\left[\left(\sigma.1-\delta^{-E.t_n}\right)+\left(\tau.1-\delta^{-Q(n)}\right)\right]} \quad (3)$$

Variable credit with the type of behavior refers to type of behavior of interaction and its effect on malicious node detection. Therefore, it is necessary to give a variable value to any behavior in relation to its position. As a result the capability of applying destructive behaviors based on the validity obtained from the positive behaviors is discarded from the faulty nodes.

Equation (4) is the developed Equation (3) which provides a variable with the position of behavior with respect to the type of behavior and calculations. Therefore, in the equation provided, $Change\ Credit_{Time,Data,Behavior(i,j)}$, variable valuation over time, the position of the interaction and the type of desired behavior between the evaluator node i and the evaluated node j, and σ, τ, φ are the relative valuation coefficients (so that $\tau + \sigma + \varphi = 1$) where according to the value of each factor in the malicious node detection, a higher or lower value can be given. M is the value of the type of nth behavior (for positive behaviors it has a value larger than negative behaviors and will vary between zero to one) stimulating negative behavior in the credit (if negative behavior occurs, value will be one and otherwise zero), ∂ encourages negative behavior in the credit (in the case of negative behavior value will be 1, otherwise 0), and p is the relative valuation variable of the behavior encouragement in the amount of credit, ω is the fine factor control and k is the number of inappropriate behaviors.

$$Change\ Credit_{Time,Data,Behavior(i,j)} = \frac{\sum_{n=1}^{z} F(n)\cdot\left[(\sigma\cdot t_n)+(\tau\cdot Q(n))+(\varphi\cdot M)\right]}{\sum_{n=1}^{z}\left[(\sigma\cdot t_n)+(\tau\cdot Q(n))+(\varphi\cdot M)\right]} - \left(\omega\cdot\left(\partial\frac{1}{1+p^{-k}}\right)\right) \quad (4)$$

As mentioned above, the evaluation of the direct observations in Equation (1) is carried out based on Equation (4) in order to increase the accuracy in the calculations and the malicious nodes detected in a desirable form during direct observations. The following Equation (5) shows evaluation of direct observations by the evaluator node i for the evaluated node j.

$$\boldsymbol{DT_{i,j}(t)} = \frac{\sum_{n=1}^{z} \boldsymbol{F(n).\left[\left(\sigma.1-\delta^{-E.t_n}\right)+\left(\tau.1-\delta^{-Q(n)}\right)+(\varphi.M)\right]}}{\sum_{n=1}^{z}\left[\left(\sigma.1-\delta^{-E.t_n}\right)+\left(\tau.1-\delta^{-Q(n)}\right)+(\varphi.M)\right]} - \boldsymbol{\omega.\left(\partial\frac{1}{1+p^{-k}}\right)} \quad (5)$$

3.2.2. Evaluation of Indirect Observation

Due to the distributed nature of wireless sensor networks, indirect observation and its sharing between nodes is one of the necessities in the field of malicious node detection and is a complement to direct observations in order to better evaluate malicious node detection. In the proposed protocol, based on the confidence levels presented below, if necessary, the evaluator node will ask for advice regarding the node to be evaluated. Then, while receiving the submitted recommendations, the node first validates the received recommendations. This authentication is embedded in the proposed protocol in order to prevent malicious nodes and malicious recommendations. If the verification indicates that the received recommendation is malicious, the received recommendation will be deleted and will not be included in the calculations, and the recommendation sending node will be added to the list of suspicious nodes. Indirect observations will then be evaluated using validated recommendations. Equation (6) provides how to measure and calculate the authentication of the credit in relation to indirect observations received, so that in the equation provided, $AC_{i,j}$ indicates the credit authentication authority of the recommendation received, $Rec\,(f)_j$ is fth recommendation received from jth node, k is the total

number of recommendations, and $\frac{\sum_{f=1}^{k} Rec\ (f)}{k}$ is the average of all received recommendations. If $AC_{i,j}$ for node j exceeds the threshold of credit, the recommendation received is not considered in the calculations and the node will be added to the list of suspicious nodes.

$$AC_{i,j} = \left| \frac{\sum_{f=1}^{k} Rec\ (f)}{k} - Rec\ (f)_j \right| \quad (6)$$

If based on Equation (6), jth node is identified as a suspicious node in the validation of indirect observations, the node which receives the recommendation evaluates the suspicious node in order to detect malicious nodes. In order to identify and detect malicious nodes in this process, the ith evaluator node asks recommendation from ith evaluated node for the nodes with enough knowledge and trusted nodes. Subsequently, the recommendations received by the ith evaluator node are evaluated according to Equation (7) related to indirect observations and if the deviation of the received recommendations exceeds one value of threshold, the node will be added to the blacklist as malicious node. In Equation (7) presented, $ER_{i,j}$ is Detection Indicator, $T(a)_i$ is the value of the trust of the node i to the ath node, m is the total number of nodes for which the recommendation is asked, $RT(a)_j$ is the recommendation received from the jth node under evaluation in relation to the ath node.

$$ER_{i,j} = \left| \frac{\sum_{a=1}^{m} T(a)_i}{m} - \frac{\sum_{a=1}^{m} RT(a)_j}{m} \right| \quad (7)$$

Furthermore, based on the validated recommendations of the indirect factor, evaluation will be calculated using Equation (8) and considered in malicious nodes detection of Equation (1). As in Equation (8), $DT_{i,k}$ is the credit of node k for node i and $DT_{k,j}$ j is the credit of node j for node k.

$$IT_{i,j}^{k}(t) = \begin{bmatrix} \sum_{k \in N_j, k \neq i} DT_{i,k}(t) * DT_{k,j}(t) & if\ AC_{I,J} < T_{Re} \\ Discard\ Recommendation & Else \end{bmatrix} \quad (8)$$

3.2.3. Trust Update

The main nature of the trust is its dynamics. This means that the amount of trust increases or decreases over time, and it does not always have a fixed value. In fact, the amount of trust varies according to the number of transactions and the type of transactions [40]. Due to the rapid and unpredictable behavior of a node, a node that was previously a faulty node might behave as a malicious node or a node that was previously normal becomes a faulty or malicious node after a while. Therefore, in order to have a secure network, the trust has to be updated after an event or over time. The direct trust of a node is updated by observing the direct behavior of nodes that are one-hop neighbor of nodes. Indirect trust is updated by recommendations from other nodes that observe the desired node behaviors and the update value of the beta distribution function is also carried out based on the data rate that is sent. Updating the trust value at time intervals Δt is undertaken as follows [24]:

$$T_{updateij}\ (t+\Delta t) = T_{ij} + T_{ij}\ (t+\Delta t) \quad (9)$$

$T_{updateij}$ (t+Δt) represents the trust value at the time t+Δt and T_{ij} (t+Δt) indicates the updated value trust of node i to node j at time t + Δt.

3.3. The Proposed Protocol

3.3.1. Assumptions

In this section, the present research details are described.

1. All nodes in the network have the same resources in terms of energy, processing power and communication power.
2. The nodes are randomly distributed in the region and the location of the nodes and the location of the base station are fixed.
3. The destination of the transmitted data is the sink.

There are a small number of malicious nodes in the network that divert incoming packets from the path.

4. Malicious nodes do not collude with each other.

Each node maintains a list of its neighbors.

5. Each node can communicate with nodes that are in their radio range and are neighbors of the node.

3.3.2. Goals

The proposed protocol pursues the following goals:

1. Enhancing security in wireless sensor networks using learning automata and a concept called trust.
2. Increasing the life time of the network as our protocol tries to select nodes that have more residual energy as well as selecting the shortest path, thus increasing the life of the network.
3. Choosing the path with the least number of hop to destination.
4. Increasing network throughput.

3.3.3. Network Model

A wireless sensor network consists of a large number of nodes that are distributed in different places and work together to send data and information they receive from the environment to the base station. Information on all sensor nodes, including the base node, is stored in the Init_Table. Information concerning nodes, paths, and packets is dynamically learned by the automata placed on each sensor node, and the information is updated at each period T. The

proposed method investigates the data packet and the behavior of the nodes in relation to the received packet and using the learning automata identifies the malicious nodes and avoids communication with them, thus preventing the occurrence of attacks such as black hole and increasing the security of the wireless sensor network. The network consists of N nodes which are distributed in different places. Each node has its own characteristics such as initial energy, maximum packet size, and maximum control packet size which is shown in Table 1. In this network, the amount of initial energy for each node is 50 J, the size of the data packet is 1500 bits, the communication radius is 250 meters and the sensing radius is considered to be half the communication radius.

Table 1. Initial values assigned for nodes.

Input Parameters	Value initialization
Initial energy	50 J
Packet size	1500 bit
Control radius	250m
Sense radius	125m

Due to the fact that wireless sensor networks are widely used in unpredictable and dynamic environments, the use of this type of network is very useful at current times. However, this type of network is exposed to all types of attacks due to its open media. Therefore, providing a solution to identify malicious nodes in this type of network and designing a secure routing method will increase throughput of this type of network. In this paper, a secure routing method based on the detection of malicious nodes using learning automata and an appropriate evaluation function (EF) in a wireless sensor network is presented. Fig. 1 illustrates the proposed framework. A multi-step scenario was considered for the present research. In the first phase (network setup and initialization), the wireless sensor nodes are randomly distributed in the desired environment. Forming a network and identifying the neighbors provide each node with a list of neighbors. In the second phase (selecting the best path based on the evaluation function and the learning automata), the best path is selected to send data to the sink. Each node selects from the neighbor list; the node that has more trust and residual energy and fewer steps to the sink using the learning automata and the appropriate evaluation function to send data is selected. In the third phase (learning and updating), when a node sends data to one of its neighbors, according to the behavior of the neighboring node towards the received data (environmental feedback), the learning automata that is on the source node will be rewarded or punished, and the probability of selecting other neighbors will be updated in proportion to the feedback from the environment. Fig. 2 demonstrates the proposed mechanism that provides a secure routing for the wireless sensor network.

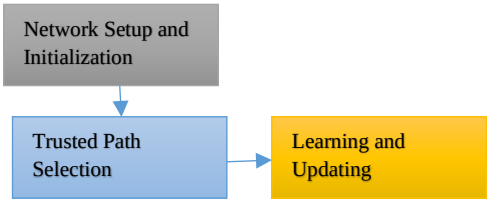

Fig. 1. Proposed Mechanism.

3.3.4. Network setup and initialization phase

In the first phase, the nodes are randomly distributed in the desired environment for different applications.

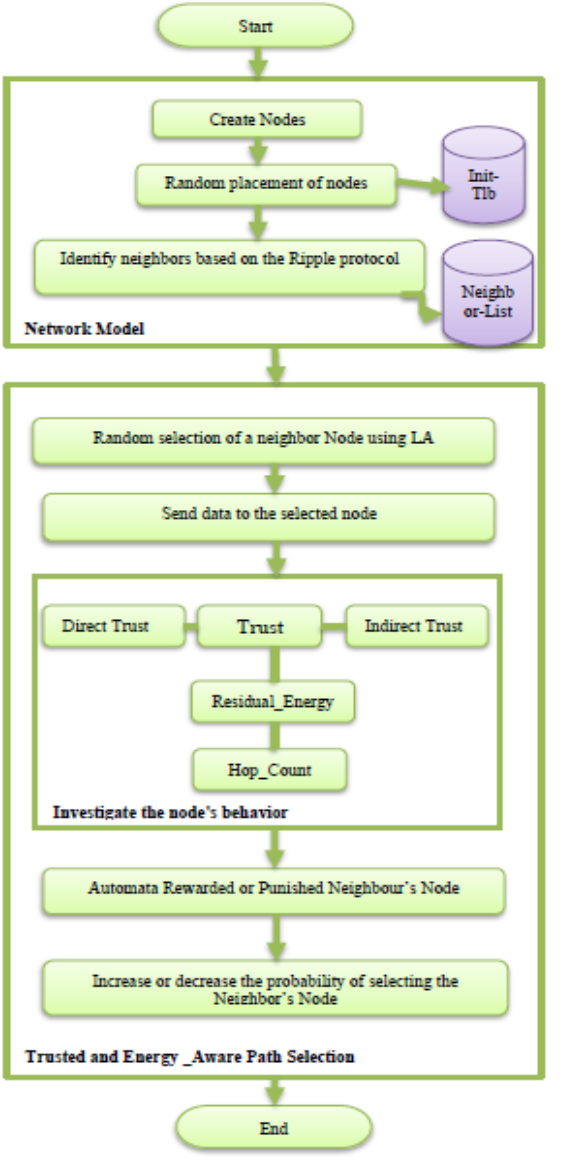

Fig. 2. The proposed framework.

The network is formed after all the nodes distributed in the desired locations. Information concerning all sensor nodes, including the base node, is stored in the Init_Table. The Ripple protocol was used to form the network, so that each node identifies its neighbors and stores information on neighbors in the Neghbour_List table.

Each node i maintains a Neighbour_List containing 6 fields. Fig. 3 displays the details.

Node_Id	Residual_Energy	Coordinate(x,y)	Hop_count	T	P

Fig. 3. Fields of a record in the Neighbour_List table.

Below greater details regarding each field in Fig. 3.

Node_Id: shows the node's ID of the desired neighbor.

Residual_Energy: shows the residual energy of the neighboring node that is updated at each period T.

Coordinate (x, y): shows the coordinates of the node's location.

Hop_Count: indicates the number of node hops to the sink.

T: indicates the degree of trust of the node that is updated at each period T.

P: specifies the initial probability of selecting this node to send data to and this probability is updated by the learning automata upon receiving feedback from the environment.

3.3.5. Selecting the best path based on the evaluation function and the learning automata phase

In the proposed method, a learning automata LA_k is placed on each node k. This automaton helps the relevant node to select the most trusted and suitable neighbor for sending data to. In the first stage, the automata consider the same probability for the neighbors of node k. The amount of this probability is determined according to the number of neighbors. If node k has r neighbors, the probability of selecting each of these neighbors is $\frac{1}{r}$. Each element in the Neighbour_List table corresponds to a learning automata's action. Therefore, each time a sensor node senses data from the environment or receives data from a neighboring node, for the first time, the automata randomly selects one of these neighbors and then waits for the feedback of the environment. If the environment has positive feedback on the performance of the node that the data was sent to, the automata automatically rewards the action of its choice and increases the probability of choosing this neighbor at later stages, reducing the probability of choosing other neighbors (learning and updating the probability phase). In later stages when that node senses or receives data from one of his neighbors, the selection procedure is such that the node with the highest probability is always selected to send data to. Environmental feedback is based on the value of the evaluation function, which is obtained according to the criteria of the node's trust, the residual energy of the node and the number of hops that node needs to reach its destination. To avoid high energy consumption of nodes in the network and increase the life of the network, the environmental feedback is not calculated after sending each packet. To reduce the calculations and reduce the energy consumption of the sensor nodes, the evaluation function is calculated after each period T. Evaluation function and its parameters are stated in Equation 10. The automata placed on each node in the wireless sensor network considers three criteria for calculating these probabilities: (1) residual energy level of the node, (2) number of hops of the node to the sink and (3) the level of trust of the desired node.

$$EF=w1\ Residual\_Energy_i\ (t) +w2\ T_i\ (t) +w3\ Hop_i \quad (10)$$

Details of the evaluation function are given below:

EF: shows the value of the evaluation function for node i

$Residual\_Energy_i$ (t): the residual energy of node i at time (t)

T_i (t): the amount of trust of node i at time (t)

Hop_i: indicates the number of hops that node i need to reach the sink

w1, w2 and w3 are coefficients that represent the residual energy weight of the node, the weight of the node's trust, and the weight of the number of node's hops to the destination, respectively. In this article, the value of these coefficients must be the same so that the importance of the three parameters is the same, but in different applications, these coefficients can be different and also w1 + w2 + w3 = 1.

3.3.6. Learning and updating phase

As described in the previous section, a learning automata LA_k is placed on each node K in the wireless sensor network, so whenever each node receives data from another node or senses data itself, one of its neighbors chooses to send data to it and then waits for feedback and the function of the node which received data. Depending on the function of the node in question, the automata is either rewarded or punished. Given that the main purpose of this article is to select the trusted nodes, then the feedback of the node should be considered in addition to considering the remaining energy parameters of the node and the number of hops needed to reach the destination node indicating the degree of trust of that node. To calculate the trust of that node, a combination of two parameters, direct trust and indirect trust, is used. At the end of each period, the

value of the EF function is calculated for each neighboring node, which occurs according to the Table 2 in 4 cases.

If, according to the above table, the desired node is in class 1, it means that the desired node is reliable in terms of trust and has enough energy to send the packet and the number of hops to the destination is small. So the automata is rewarded and increases the probability of selecting this node in the next step corresponding to the a parameter (a).

If, according to the above table, the node in question is in class 2, it means that this node is a good node, but not as good as the nodes in class 1, so the reward that is given to the automata is half to that at the previous stage. In the next step, the probability of selecting this node increases by half of the a parameter $\left(\frac{\alpha}{2}\right)$.

If, according to the above table, the node in question is in class 3, the status of the node is unknown and automata will not be rewarded or punished, and the probability of selecting this node will not change.

If, according to the above table, the node in question is in class 4, it means the selected node is not a suitable node (either its trust level and residual energy are low or it has a long way to the destination) and the relevant automata is punished. In the next step, the probability of selecting this node decreases as much as the b parameter.

Table 2. Ef values and class of nodes.

Level	EF value	Class of Node
1	$(\varepsilon, 1]$	good
2	$(0.5, \varepsilon]$	Less good
3	0.5	Indecisive
4	(0,0.5)	bad

In general, if the action α_i is selected in step n and this action receives the desired response from the environment, the probability of selecting this action, hence pi (n), increases and other probabilities decrease. If an unfavorable response is received from the environment, the probability of selecting action α_i, hence pi (n), decreases and other probabilities increase. These changes are always applied in such a way that the sum of all probabilities is equal to one. That is $\sum p_i(n) = 1, i = 1,2,3, \dots r$.

And r is equal to the number of node's neighbors. Increasing or decreasing the probabilities under different conditions in a learning automata with variable structure is undertaken according to equations (11) and (12).

In case of receiving a favorable response from the environment, the automata receive a reward:

$$p_i(n+1) = p_i(n) + a[1 - p_i(n)]$$

$$p_j(n+1) = (1-a)p_j(n) \qquad \forall j, j \neq i \qquad (11)$$

In case of receiving an unfavorable response from the environment, automata are punished:

$$p_i(n+1) = (1-b)p_i(n)$$
$$p_j(n+1) = \frac{b}{r-1} + (1-b)p_j(n) \qquad \forall j, j \neq i \qquad (12)$$

In these equations, r represents the number of automata operations (number of node's neighbors), a represents the reward parameter, and b indicates the penalty parameter. Fig. 4 shows the pseudo-code of the automata learning section and updating of the probability parameters.

Algorithm. Learning phase

1) In Step n, each Node N_i Randomaly Selects Node N_j to Send Data Packet $N_j \in$Neghbour N_i

2)After the time period T, node N_i checks function EF for node N_j

3)if Ef(N_j) $> \varepsilon$ then

status(N_j)=good, It means Rewarded and incremaent $P_j(n+1)$ with parameter a

4) if Ef(N_j) > 0.5 and Ef(N_j) $< \varepsilon$ then

status(N_j=less good, It means Rewarded and incremaent $P_j(n+1)$with parameter $\frac{a}{2}$

5) if Ef(N_j) $== 0.5$ then

status(N_j)=Indecisive, It means No change and Not incremaent $P_j(n+1)$ and Not Punishment $P_j(n+1)$

6) if Ef(N_j) $< \varepsilon$ then

status(N_j)=bad , It means Punishment and decrese Pj(n + 1) with parameter b

Fig. 4. Pseudo-code for learning and updating phase.

6. SIMULATION RESULTS

Performance evaluation of the proposed method is described in this section. The simulation of these experiments was implemented in MATLAB R2018b software installed on a system with the following specifications: Intel® coreTM i5-7200CPU @ 2.50 GHz processor with 8 GB of RAM and 64-bit operating system and x64-based processor.

The number of sensor nodes was considered to be 100 and they were randomly distributed in a space with dimensions of 1400 * 800 m2. The initial energy of all nodes was 50 J.

Transmit and Receiver Electronic (E_{elect}) was set to 50 nJ/bit, Transmitter Amplifier(E_{amp}) was set to 100 PJ/bit /m2, EDA was set to 5nJ/b, the size of Packet was set to 1500 bits, d0 was set at 87.0 and

communication range was set at 250.0 m (see Table 3).

Different criteria were used to measure the proposed method. The simulation was performed in 1000 seconds and the criteria of throughput, network life time, average end-to-end latency and normalized routing load in the proposed method were compared to the three methods of ESRT [24], TLB-AODV [25] and R-AODV [26].

6.3. Throughput Analysis

Fig. 5 shows the efficiency of the proposed method compared to the three protocols of ESRT, R-AODV and TLB-AODV. As can be observed, the proposed method demonstrates better results compared to the three mentioned protocols, and this is due to the efficiency of the proposed method in calculating the trust of nodes and detecting malicious nodes. After identifying malicious nodes, they are isolated and no data is transmitted to them. This use of trusted nodes increases the efficiency of the entire network in routing.

Table 3. Network Simulation Parameters.

Input Parameters	Value Initialized
Area	1400*800 m
Number of Sensor Nodes	100
Initial Energy of SNs(E_0)	50 J
Simulation Times	1000 s
Communication range	250m
Packet Size	1500 bit
Transmitter and Receiver Electronics(E_{elect})	50nJ/bit
d_0	87m
EDA	5nJ/bit
Transmit Amplifier(E_{amp})	100pJ/bit/m2

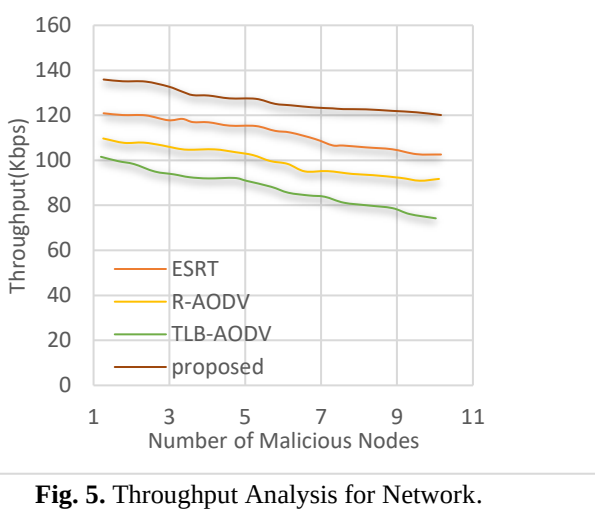

Fig. 5. Throughput Analysis for Network.

6.4. Network Lifetime Analysis

This analysis refers to the life time of the network, where the life of the network is considered until the first node in the network is shut down due to the termination of its energy and so called death. Fig. 6 indicates the life time of the network in the presence of malicious nodes. As previously explained, in the proposed method, when a node wants to send its data to one of its neighbors, one of the main criteria for selecting a neighboring node is its energy level in addition to the number of hops from neighbor to destination. Thus, by always attempting to select a node with a higher energy level and fewer hops to the destination, the life of the network is lengthened.

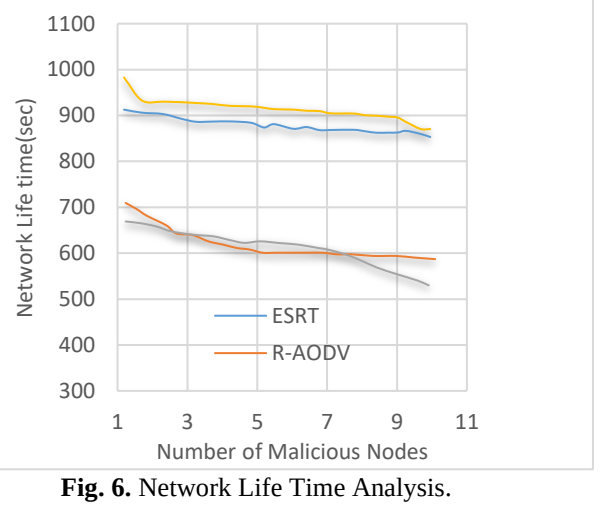

Fig. 6. Network Life Time Analysis.

6.5. Average end-to-end Delay Analysis

This analysis refers to the average arrival time of all packets from different sources to the destination node. It is clear that as the number of malicious nodes in the network increases, the average end-to-end delay increases also. The results of average end-to-end delay for the proposed protocol and the three protocols of ESRT, R-AODV and TLB-AODV are shown in Fig. 7. Due to the solution that the proposed protocol has in choosing the path that is more efficient and considers the three parameters of trust, energy and the shortest path, the average end-to-end delay is less than other protocols.

6.6. Normalized Routing Load (NRL) Analysis

NRL refers to the ratio of total number of transmitted control packets to the total number of received data packets. As the number of malicious nodes in the network increases, so does NRL because malicious nodes throw packets out of the way and have to be re-sent. In the proposed method, according to the approach used, timely detection of malicious nodes prevents data from being sent to it, and as a result data packets are sent by reliable and correct nodes and there

is no need to re-send data packets. Fig. 8 displays NRL analysis results.

7. CONCLUSION

Due to the importance of security in wireless sensor networks, the main purpose of this paper was to provide a trust-based routing method using a learning automata and an appropriate evaluation function. The evaluation function uses the three parameters of node's residual energy, its trust and the number of hops to the sink after each period to obtain a value for evaluating the node, which was compared with the threshold. Based on this, the automata's selection action must be rewarded or punished. To obtain the amount of trust in the evaluation function, a combination of two parameters, direct trust and indirect trust of each node, was used. The evaluation results of the proposed method show that network throughput, network life, average end-to-end latency and normalized routing load in this method are significantly improved compared to those of ESRT, TLB-AOVD and R-AODV methods.

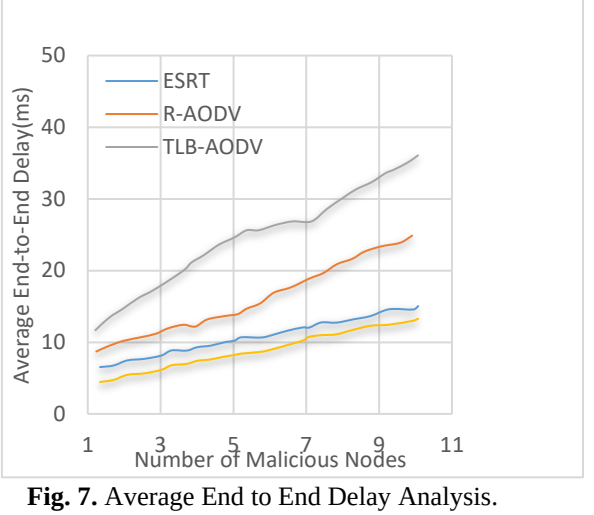

Fig. 7. Average End to End Delay Analysis.

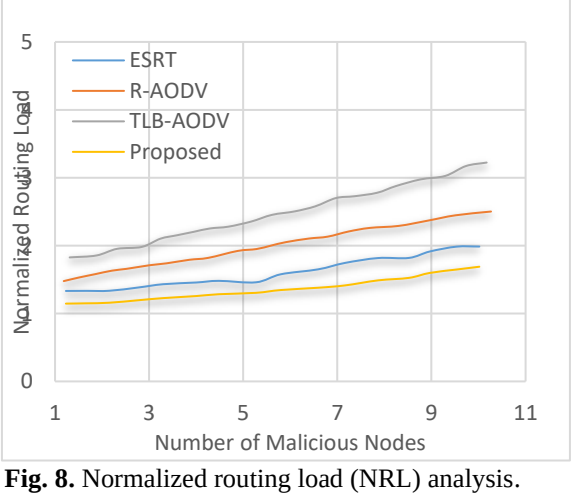

Fig. 8. Normalized routing load (NRL) analysis.

REFERENCES

[1] Bhuiyan, M.Z.A., et al., **"Dependable structural health monitoring using wireless sensor networks,"** *IEEE Transactions on Dependable and Secure Computing*, Vol. 14(4), pp. 363-376, 2015.

[2] Alam, M.M., D.B. Arbia, and E.B. **"Hamida, Wearable wireless sensor networks for emergency response in public safety networks,"** *in Wireless Public Safety Networks 2.*, Elsevier. pp. 63-94, 2016.

[3] Ojha, T., S. Misra, and N.S. Raghuwanshi**," Wireless sensor networks for agriculture: The state-of-the-art in practice and future challenges,"** *Computers and Electronics in Agriculture*, Vol. 118: pp. 66-84, 2015.

[4] Gomes, R.D., et al., **"Application of Wireless Sensor Networks Technology for Induction Motor Monitoring in Industrial Environments,"** in *Intelligent Environmental Sensing*., Springer .pp. 227-277, 2015.

[5] Kuorilehto, M., M. Hännikäinen, and T.D. Hämäläinen, **"A survey of application distribution in wireless sensor networks",** *EURASIP Journal on Wireless Communications and Networking*, Vol. 2005(5), pp. 859712, 2005.

[6] Almudevar, A., A. Leibovici, and C. Horwitz. **"Electronic motion monitoring in the assessment of non-cognitive symptoms of dementia,"** *2005. INTERNATIONAL CONGRESS OF THE INTERNATIONAL PSYCHOGERIATRIC* ASSOCIATION.

[7] Nurmi, P., et al**."A Framework for Distributed Activity Recognition in Ubiquitous Systems,"** *in IC-AI*. 2005.

[8] Jafari, R., et al. **"Wireless sensor networks for health monitoring,"** *In the Second Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services*. 2005. IEEE.

[9] Butun, I., S.D. Morgera, and R. Sankar, **"A survey of intrusion detection systems in wireless sensor networks,"** *IEEE communications surveys & tutorials*, Vol. 16(1), pp. 266-282, 2013.

[10] Shaikh, R.A., et al. **"LSec: lightweight security protocol for distributed wireless sensor network,"** *in IFIP International Conference on Personal Wireless Communications*. 2006. Springer.

[11] Perrig, A., and et al., SPINS: **"Security protocols for sensor networks. Wireless networks,"**, Vol. 8(5), pp. 521-534, 2002.

[12] Karlof, C. **"Secure routing in sensor networks: Attacks and countermeasures,"** *in Proc. of 1st IEEE International Workshop on Sensor Network Protocols and Applications (SNPA2003)*. 2003.

[13] Boukerch, A., L. Xu, and K. El-Khatib, **"Trust-based security for wireless ad hoc and sensor networks,"** *Computer Communications*, Vol. 30(11-12), pp. 2413-2427, 2007.

[14] Lacuesta, R., et al., **"A secure protocol for**

spontaneous wireless ad hoc networks creation," *IEEE transactions on parallel and distributed systems*, Vol. 24(4), pp. 629-641, 2012.

[15] Liu, Y., C.-x. Liu, and Q.-A. Zeng**, "Improved trust management based on the strength of ties for secure data aggregation in wireless sensor networks Telecommunication Systems,"** 2016. Vol. 62(2), pp. 319-325, 2016.

[16] Anita, X., M.A. Bhagyaveni, and J.M.L**." Manickam, Collaborative lightweight trust management scheme for wireless sensor networks,"** *Wireless Personal Communications*, Vol. 80(1), pp. 117-140, 2015.

[17] Rajeshkumar, G. and K. Valluvan, **"An energy aware trust based intrusion detection system with adaptive acknowledgement for wireless sensor network,"** *Wireless Personal Communications*, Vol. 94(4), pp. 1993-2007, 2017.

[18] Narendra, K.S. and M.A. Thathachar, **"Learning automata-a survey,"** *IEEE Transactions on systems, man, and cybernetics*, Vol. 1974(4), pp. 323-334.

[19] Ganeriwal, S., L.K. Balzano, and M.B. Srivastava, **"Reputation-based framework for high integrity sensor networks,"** *ACM Transactions on Sensor Networks (TOSN)*, Vol. 4(3), pp. 15, 2008.

[20] Eissa, T., et al., **"Trust-based routing mechanism in MANET: Design and implementation,"** *Mobile Networks and Applications*, Vol. 18(5): p. 666-6, 2013.

[21] Song, J., et al. **"Dynamic trust evaluation of wireless sensor networks based on multi-factor,"** in *2015 IEEE Trustcom/BigDataSE/ISPA. 2015. IEEE*.

[22] Ahmed, A., et al., **"A trust aware routing protocol for energy constrained wireless sensor network,"** *Telecommunication Systems*, Vol. 61(1), pp. 123-140, 2016.

[23] Ahmed, A., et al., **"A secure routing protocol with trust and energy awareness for wireless sensor network,"** *Mobile Networks and Applications*, Vol. 21(2), pp. 272-285, 2016.

[24] Ahmed, A., et al., **"Energy-aware and secure routing with trust for disaster response wireless sensor network,"** *Peer-to-Peer Networking and Applications*, Vol. 10(1), pp. 216-237, 2017.

[25] Marchang, N. and R. Datta, **"Light-weight trust-based routing protocol for mobile ad hoc networks,"** *IET information security*, Vol. 6(2), pp. 77-83, 2012.

[26] Channa, M.I. and K.M. Ahmed, **"A Reliable Routing Scheme for Post-Disaster Ad Hoc Communication Networks,"** *JCM*, Vol. 6(7), pp. 549-557, 2011.

[27] Fang, W., et al., **"BTRES: Beta-based Trust and Reputation Evaluation System for wireless sensor networks,"** *Journal of Network and Computer Applications*, Vol. 59, pp. 88-94, 2016.

[28] Li, Y. and L.E. Parker. **"Intruder detection using a wireless sensor network with an intelligent mobile robot response,"** *in IEEE SoutheastCon 2008*, 2008. IEEE.

[29] Jiang, J., et al., **"An efficient distributed trust model for wireless sensor networks,"** *IEEE transactions on parallel and distributed systems*, Vol. 26(5), pp. 1228-1237, 2014.

[30] Marzi, H. and M. Li, **"An enhanced bio-inspired trust and reputation model for wireless sensor network,**" *Procedia Computer Science*, Vol. 19, pp. 1159-1166, 2013.

[31] Chowdhury, A.R., S. Tripathy, and S. Nandi. **"Securing wireless sensor networks against spurious injections,"** *in 2007 2nd International Conference on Communication Systems Software and Middleware*. 2007. IEEE.

[32] Junnarkar, A. and A. Bagwan. **"Novel Quality of Service (QOS) Improvement Routing Protocol for MANET Using Ant Colony Optimization,"** *in 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)*. 2017. IEEE.

[33] Misra, S., et al**.," LAID: a learning automata-based scheme for intrusion detection in wireless sensor networks,"** *Security and Communication Networks*, Vol. 2(2), pp. 105-115, 2009.

[34] Das, S.K., P.S. Chatterjee, and M. Roy, **"Detecting and Punishing the Selfish Node and Its Behavior in WSN,"** *International Journal of Computer & Organization Trends*, Vol. 6(1), 2014.

[35] Prithi, S. and S. Sumathi, **"LD2FA-PSO: A novel Learning Dynamic Deterministic Finite Automata with PSO algorithm for secured energy efficient routing in Wireless Sensor Network,"** *Ad Hoc Networks*, 2020. 97: p. 102024.

[36] Zhan, G., W. Shi, and J. Deng, **"Design and implementation of TARF: A trust-aware routing framework for WSNs,**" *IEEE Transactions on dependable and secure computing*, Vol. 9, pp. 184-197, 2011.

[37] Pirzada, A.A. and C. McDonald**. "Trusted greedy perimeter stateless routing,"** *in 2007 15th IEEE International Conference on Networks.* 2007. IEEE.

[38] Narendra, K.S. and M.A. Thathachar**, "Learning automata: an introduction,"**, *Courier corporation*, 2012.

[39] Misra, S., P.V. Krishna, and K.I. Abraham, **"A simple learning automata-based solution for intrusion detection in wireless sensor networks,"** *Wireless Communications and Mobile Computing*, Vol. 11(3), pp. 426-441, 2011.