A Reinforcement Learning Based Secured Routing Protocol for Wireless Sensor Networks

FAGBOHUNMI Griffin Siji[1], ENEH I. I.[2]

[1]*Computer Engineering Department, Abia State Polytechnic Aba, Abia State, Nigeria*
[2]*Electrical and Electronics Engineering Department, Enugu State University of Science and Technology, Enugu, Nigeria*

***Abstract:*-Wireless sensor networks (WSNs) consist of spatial distribution of sensors which co-operatively monitor the environment for certain phenomenon of interest such as temperature, humidity, pressure etc, and send their sensed data through multi-hop route to the sink. These nodes may vary between hundreds to thousands depending on the size and nature of data (signals) to be detected. Wireless sensor networks are expected to operate over long periods without being attended to. The range of this period may span from some months to even years. However due to its resource constraints i.e. limited battery power, low bandwidth, limited sensing range and low memory, it is pertinent that its resources must be optimally utilized. This paper addresses a secured routing protocol to a specified sink in a multi-sink scenario. It is a subset of a novel algorithm required to securely route data to multiple mobile sinks in WSN. It employs reinforcement learning paradigm and in particular (Q-learning) while the transition (action) is modeled as a Partially Observable Markov Decision Process.**

***Index Terms*—reinforcement learning, Q-learning, Trust mechanism, computational intelligence, localization**

I. INTRODUCTION

WSNs are composed of spatially distributed sensor nodes that cooperatively monitor environmental changes over time. Sensors sense data and transmit it to the sink (gateway between sensor nodes and end users) through multi-hop routing. WSN has a number of constraints when compared with other data communication networks, such as wired and wireless networks. This includes (i) they are both energy and power constraint, [1] (ii) they have limited bandwidth (iii) they are deployed in the open. All these limitations suggest that a routing protocol designed for wireless sensor networks must not only optimize its limited resources optimally, but must also be secure so as to mitigate the effects of adversarial nodes. The unreliable wireless channels and unattended operations make it very easy to compromise/capture the nodes. In Nigeria the menace of crude oil pipeline vandalization has cost the federal government huge fortune, hence no cost should be spared in protecting this huge resource. This can be implemented through an appropriate deployment of a secured and energy efficient routing protocol using wireless sensor networks to monitor online environmental phenomenon such as, pressure, temperature and flow rate of the crude oil in the pipes. A lot of effort has gone into secured routing in Wireless sensor networks. The current approach is the combined use of cryptography and trust mechanism as proposed in RFSN [2] and TARP [3]. However this approach is not resilient to adversarial nodes capable of compromising the trust mechanism. These adversarial nodes can achieve this by giving false recommendation about neighbour nodes. Secondly these protocols require the explicit model of the network topology, a requirement that will be too much for the memory constrained wireless sensor network nodes. Thirdly in an attempt to isolate adversarial nodes using the trust mechanism, a lot of control information are included in the data packets which increases network overhead.

This paper employs Q-learning for the protocol design. Q-learning [4] is a reinforcement learning technique that models sequential decision making in a partially observable environment, making it an ideal choice for nodes in WSNs that need to choose a suitable next- hop neighbour to route packets with only limited information Its strength lies in the fact that it doesn't require an explicit model of the network topology, (It updates its Q-value based on the agent's interaction with the environment). It only stores the outcome of the agent's interaction with the environment, hence it can be easily deployed on the memory constrained WSN. It has been shown that Q-learning converges to the optimal action-value function [5] - [6]. However, it suffers from slow convergence, especially when the discount factor γ is close to one [7], [8]. The main reason for the slow convergence of Q-learning is the combination of the sample-based stochastic approximation (that makes use of a decaying learning rate) and the fact that the Bellman operator propagates information throughout the whole space (especially when γ is close to 1). This is taken care of in this protocol because the learning rate here is 1, i.e. the initial Q-value is a function of the number of nodes and neighbour to each nodes, unlike the random value used in the original Q-learning Hence the Q-value is bound to successively reduce and converge more quickly to the optimal value instead of oscillating as in the original Q-value model and secondly each node stores only the routing table of its neighbour nodes instead of all the nodes in the network. This gives the protocol its localized nature.

The use of the Partially Observable Markov Decision Process (POMDP) model for the transition parameter within the Q-learning model, [9] will help to simultaneously address security issues and energy constraints while routing in WSNs. POMDP model suffers from what is termed the curse of dimensionality because its state action steps increases exponentially with the number of horizon. Using factored representations as in [10], state-of-the-art off-line solution methods fail to achieve acceptable solutions. Even though on-line methods as in [11] can improve scalability, they are not applicable due to the energy constraints of WSNs. To overcome the above issues, the transition parameter (routing) in the Q-learning will be modeled using a hierarchical POMDP (called Secure Routing POMDP (SRP). Factored representation will be employed to address the complexity in solving each SRP component. The SRP hierarchy (Fig. 3 in section 4) consists of the routing POMDP for making routing decisions, the alarm POMDP for sending/receiving alarms about malicious nodes and the fitness POMDP to compute the fitness (suitability) of nodes to route packets. The contributions of this paper includes, (1) the SRP model can balance the energy consumption and secured routing required in a network involving several categories of adversarial nodes. (2) it demonstrates that SRP can mitigate the effects of different categories of adversarial nodes that target the trust systems. (3) Extensive evaluation was carried out in a simulated and a real-world test-bed, to validate the effectiveness of SRP against state-of-the-art trust based routing schemes. The above contributions greatly help to facilitate the deployment of WSNs in hostile environments.

The rest of the paper is organized as follows: Section 2 looks into related works, here current security enabled WSN routing algorithm are highlighted, section 3 provides a detailed description of the SRP model using Q-learning section 4 describes the protocol implementation, section 5 shows the results and analysis obtained through simulation and hardware test-bed, while section 6 concludes the paper and highlights areas for future research.

II. RELATED WORK

In RFSN: Reputation based Framework for High Integrity Sensor Networks. [2], the quality of a node was determined using the Beta distribution on the cooperation information collected from a watchdog [12] mechanism as well as from recommendations given by other nodes. In TARP [3] the authors use a trust mechanism which isolates routing through malicious nodes by assessing each node neighbour's forwarding ratio using both direct evaluation (RSSI) and recommendation information from other nodes. However, the above trust schemes are not resilient to sophisticated unfair rating attacks which target the trust systems and the size of their data packets due to the inclusion of many bytes of control information makes them infeasible to be deployed in a memory constrained WSN. In CONFIDANT [13] the authors use a broadcasting mechanism to send alarms about malicious nodes, however it is still susceptible to unfair ratings, where nodes can send false alarms in a sophisticated manner. The broadcast nature of the protocol also makes it memory intensive (i.e it doesn't employ the neighbourhood mechanism where the routing table comprises of routes to only neighbour nodes), and hence infeasible in WSN. In [14] the author proposed a POMDP based routing scheme that estimates its component states composed of neighbour nodes local parameters (selfishness and energy limitation). However, it uses gradient techniques (shortest path) to determine policies which (as is shown empirically), can be far from optimal. Also, it does not use recommendation information from other sensor nodes which leads to a poor packet delivery rate. This it does in other to reduce the overhead in memory requirements, taking into cognizance the limited memory capability in WSNs.

In this paper the hierarchical POMDP based approaches as in LEACH. [15] - [18], will be used. This is due to the large state space required to model its operation and because the routing problem can be easily partitioned into sub-problems based on the actions (see Fig. 3).

III. THE SECURE ROUTING POMDP MODEL USING Q-LEARNING (METHODOLOGY)

The SRP protocol preferred in this paper use the Q-learning model. Q-learning is a reinforcement learning technique in which an agent interacts with its environment in order to maximize cumulative reward in transversing from any given state to the goal state. Its description is given below:

A SRP can be described by the tuple (S, A, T, R, Ω, B): where:

Agent State (S): is defined as $\left(D_p, routes_{D_p}^{N}\right)$ where $D_p \subseteq D$ are the sinks the packets must reach and $routes_{D_p}^{N}$ is the routing information about all neighbouring nodes N with respect to the individual sinks.

Actions (A): This represents a routing decision through a neighbour node to a desired sink. This step is used to determine the sets of secured neighbour (route) to each sink from the sink announcement phase in the network. It is calculated as the number of hops to a desired sink.

$$A = \sum_{d \in D_i} hops_d^{n_i} \quad (1)$$

Where $hops_d^{n_i}$ are the number of hops to reach destination d ε D_i and | Di | is the number of sinks in D.

Transition (T) : This specifies probabilities Pr(s'|s, a) i.e. the probability of transiting from state s to s' given that a certain action 'a' has occurred. It is based on the Partially Observable Markov Decision Process (POMDP) model.

$$P(s'|s, a) = \sum_{s \in S} P(s'|s, s'). \; b(s) \quad (2)$$

Q-Values. This is used to determine the value of the number of hops from any source node s through any neighbour node to the destination. The purpose of the Q-learning algorithm in this thesis is to determine which series of neighbour nodes from the source node will lead to an optimal Q-value to any particular sink in the network. (i.e. minimum number of hops).the initial Q-value will be computed as a function of the available number of neighbours to a node. This is given as:

$p_i = (e_i , S_i)$ is:

$Q(p_i) = (\sum_{s \in Si} jumps_s^e) - 2\{(|S_i|) -2)\}$ - - - - - - - - (3)

where $jumps_s^{e_i}$ denotes the number of hops required to arrive at destination S using neighbour e_i .

Observation (O) : The agent also receives observations (O ε **Ω**) based on the observation model O, specifying the probabilities Pr(o|a, s') i.e. the probability of observing a certain reward given that the agent performs an action 'a' having transited to s'. The observation represents the probability distribution of the states. It is given by:

$P(o|a, s') = \sum_{s \in S} P(o|s'). P(s'|s, a)$ ---------- (4)

Reward R(s, a, s') : the reward that an action 'a' causes transition from s to s'. An infinite horizon problem is assumed . It is given by:

$R(s, a, s') = \sum_{s \in S} r(s, a). b(s)$ ----------(5)

Where $r(s, a) = C_{a_i} + {}^{min}_{a}Q(a))$ and $b(s) = P(s)$

Here C_{a_i} is the action's cost (always 1 in the hop count metric) and ${}^{min}_{a}Q(a)$ is the lowest (best) Q-value from the fit neighbours), b(s) is the probability distribution among the neighbour nodes.

Belief (B): This is a probability distribution over states via Bayes' rule. If b(s) specifies the probability of s (∀ s), the updated belief b' after taking action a and receiving observation o is given by,

$$b'(s') = \frac{\Pr(s', o|b, a)}{\Pr(o|b, a)} = \frac{\Pr(o|a, s')}{\Pr(o|b, a)} \sum_s \Pr(s'|s, a)\, b(s) \quad (6)$$

A SRP policy maps beliefs to actions and is associated with a value function Π (b) which evaluates the expected total reward of executing policy Π starting from b. The objective of a SRP agent is to find an optimal policy Π, which maximizes the expected total reward.

IV. SRP PROTOCOL IMPLEMENTATION

The following shows the flow of the protocol implementation

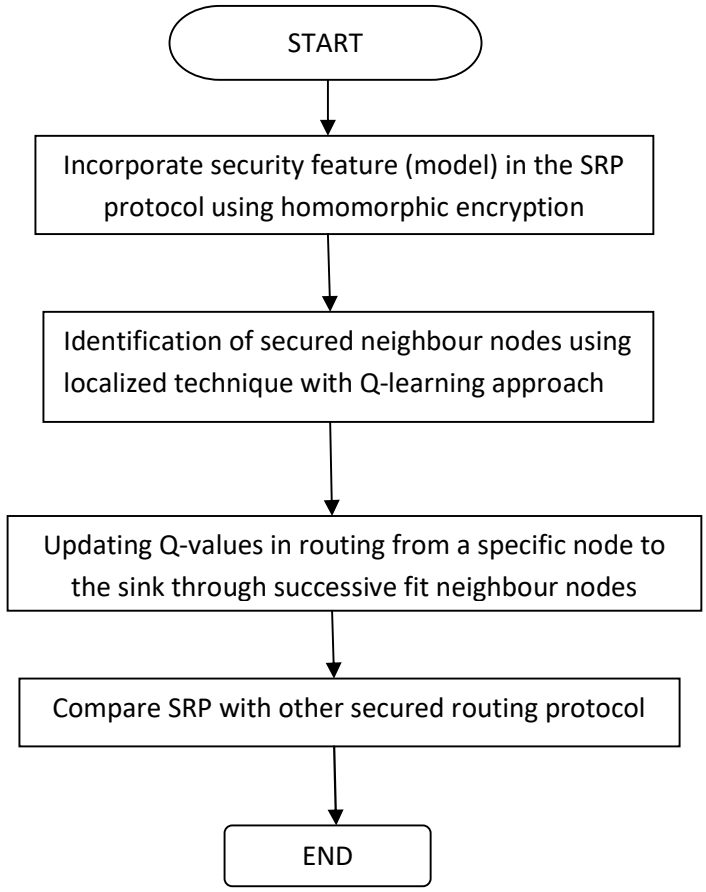


Fig1 Block diagram showing the flow of the SRP protocol

The first step in the protocol is the inclusion of the cryptographic mechanism. SRP employs the homomorphic encryption which allows arithmetic operation to be performed on ciphertexts with the result equivalent to performing same operation on plain texts. This enables nodes within a cluster to aggregate their sensor readings on the cluster without the need of decrypting the message along the path to the cluster head. This implies that all the nodes in a cluster require only the public key with the exception of the cluster heads which require both the public and private keys in order to securely transmit data in the network. The consequence of this is that the data packets for each nodes are considerably reduced leading to lower communication overhead.

The second step is the sink announcement phase, here the sink send the route request (RREQ) packet through its neighbours to all the nodes in the network. The purpose of this step is to compute the number of hop count from the sink to all the nodes in the network. Initially the hop count is nodes in a cluster require only the public key with the exception of the cluster heads which require both the public and private keys in order to securely transmit data in the network. The consequence of this is that the data packets.

The first step in the protocol is the inclusion of the cryptographic mechanism. SRP employs the homomorphic encryption which allows arithmetic operation to be performed on ciphertexts with the result equivalent to performing same operation on plain texts.

This enables nodes within a cluster to aggregate their sensor readings on the cluster without the need of decrypting the message along the path to the cluster head. This implies that all the for each nodes are considerably reduced leading to lower communication overhead.

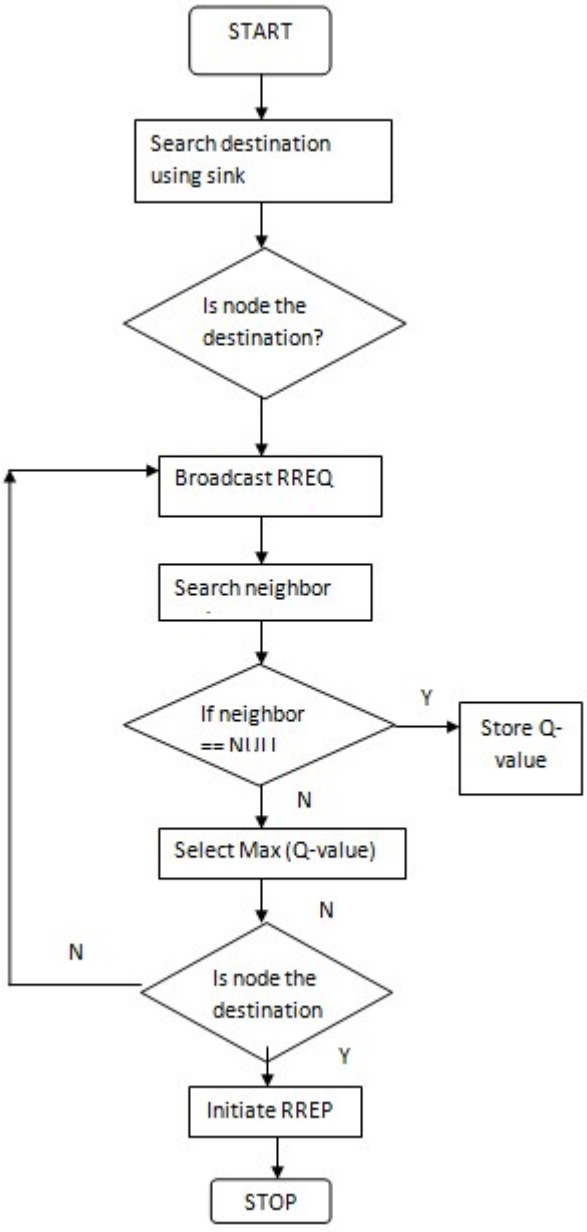


Fig 2 Flowchart for RREQ and RREP Procedure in SRP

The second step is the sink announcement phase, here the sink send the route request (RREQ) packet through its neighbours to all the nodes in the network. The purpose of this step is to compute the number of hop count from the sink to all the nodes in the network. Initially the hop count is initialized to 0 meaning that the hop count from the sink to itself is 0). This value is incremented by 1 through successive neighbour nodes. The nodes send a route reply (RREP) packets back to the sink, however this time through only the fit neighbour nodes. The parameter for determining a fit neighbour are (i) distance to sink, (ii) Percentage of remaining energy on node (iii) routing behavior (i.e. adversarial capability of nodes) and (iv) rating (i.e. the integrity of nodes to give correct recommendation about other nodes in the network. The flowchart for this is shown in fig 2

From the secured model proposed in section 3, a large state space will be required to model parameters needed in the network. This will result in an infinite convergence time for the protocol. In order to address this situation, a hierarchical formulation is proposed here (as shown in Fig 3).

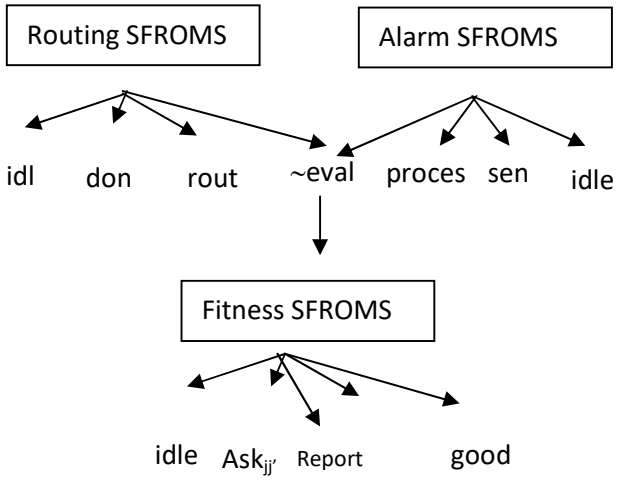


Figure 3: Secure Routing POMDP (SRP) procedure hierarchy

Hence the third step is the routing of data packets through a fit neighbour node, however this consists of three sub-function which is described as follows: anytime data packet is to be routed from a particular source node, to the sink the Routing SRP sub-function is activated. The Routing SFROMS sub-function then calls the fitness sub-function. The fitness sub-function determines the fitness of a neighbour node using the parameters stated earlier. The third sub-function the alarm sends alarm about unfit neighbbour nodes. The identity of such node is stored in the sub-function so that data packet will not be routed through the node in subsequent time.

The flowchart for the routing sub-function, fitness sub-function and alarm sub-function is shown in figure 4, 5, and 6 respectively.

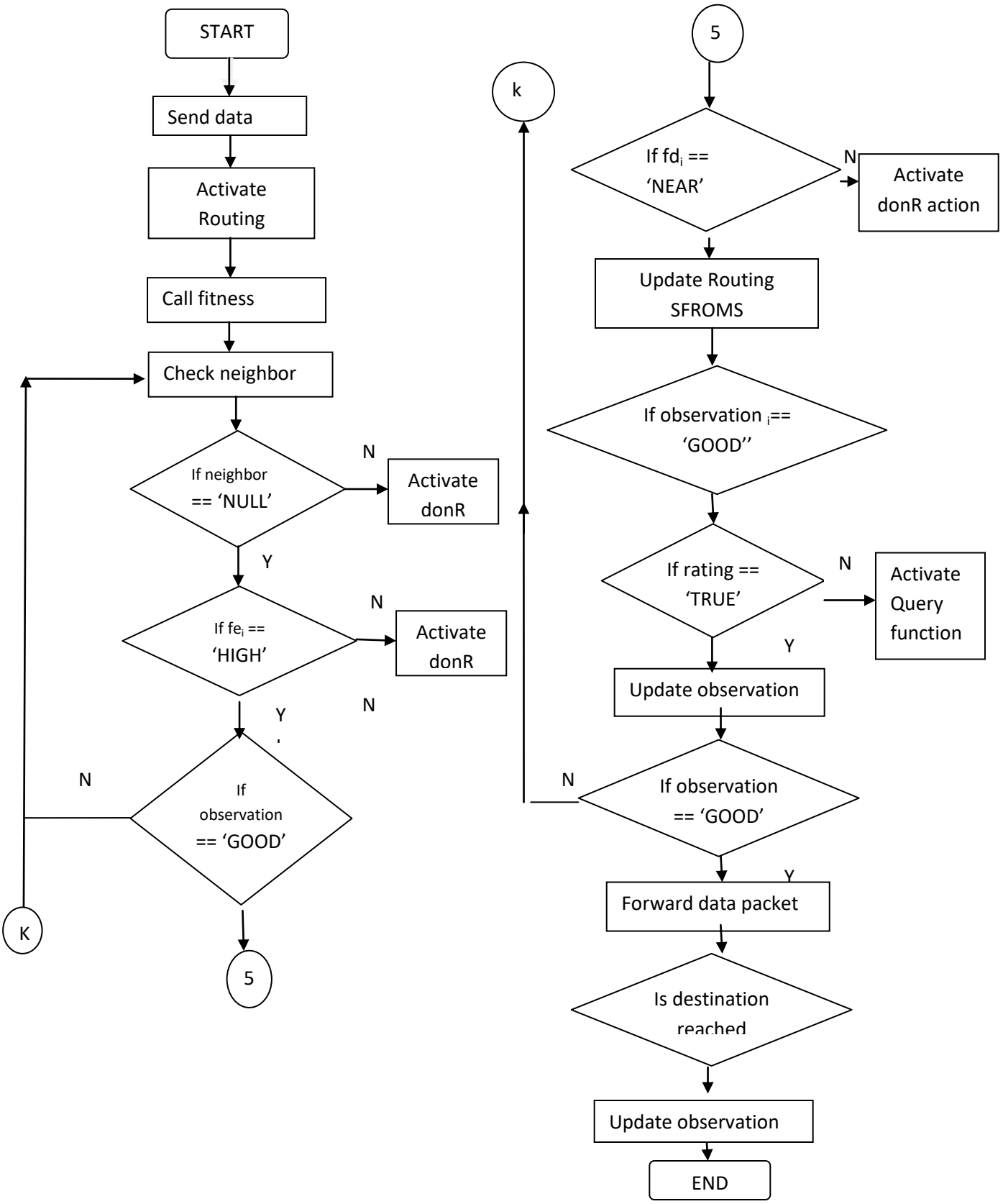


Fig 4 Flowchart for SRP routing Procedure

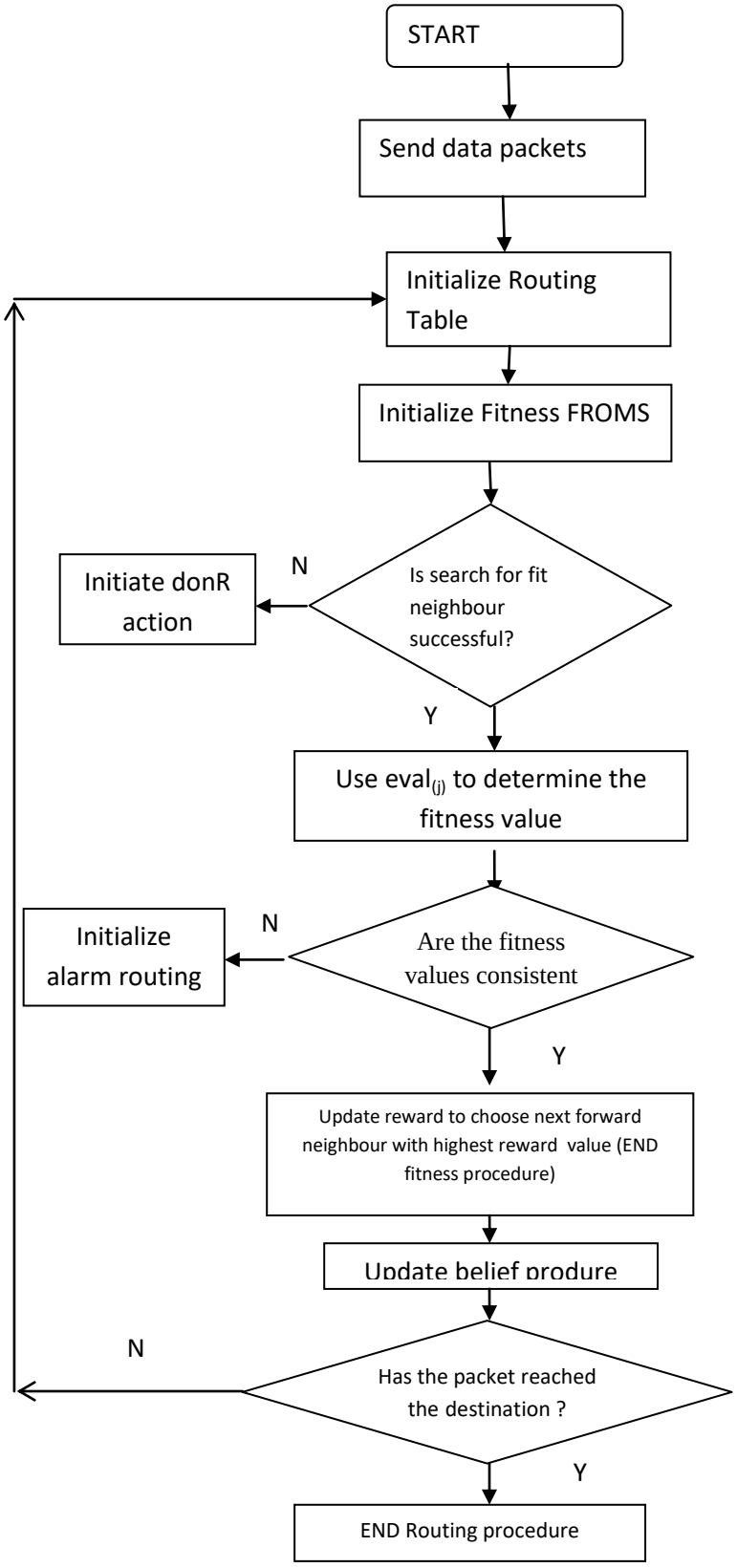
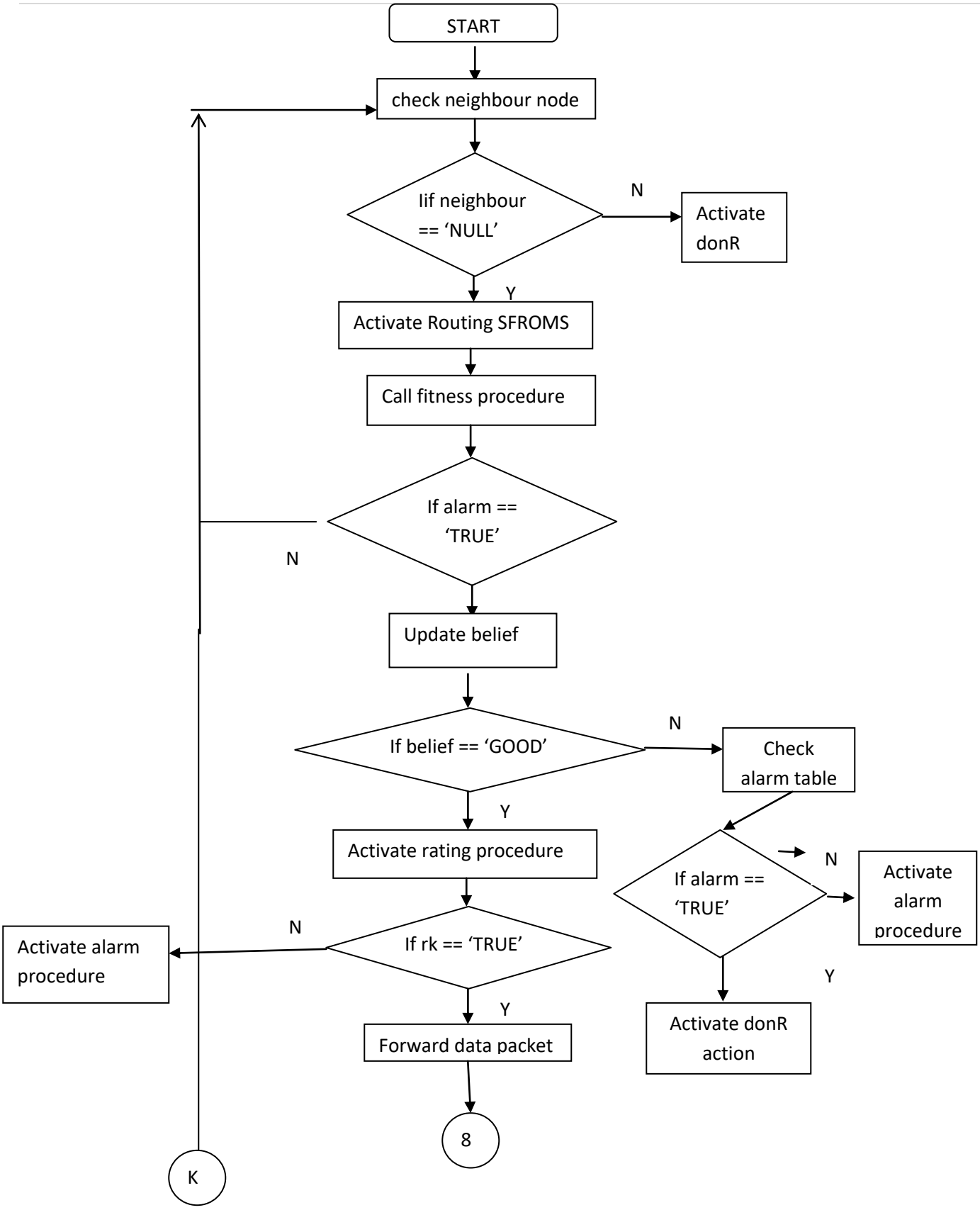


Fig 5 Flowchart for Fitness Procedure


START
check neighbour node
Iif neighbour == 'NULL'
N
Activate donR
Y
Activate Routing SFROMS
Call fitness procedure
If alarm == 'TRUE'
N
Update belief
If belief == 'GOOD'
N
Check alarm table
Y
Activate rating procedure
If alarm == 'TRUE'
N
Activate alarm procedure
Y
Activate donR action
If rk == 'TRUE'
N
Activate alarm procedure
Y
Forward data packet
8
K

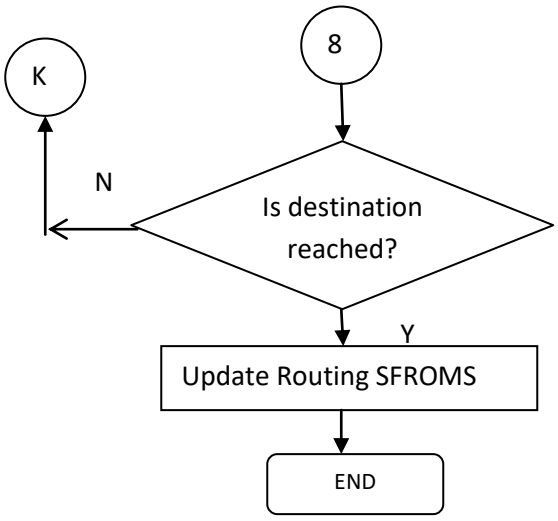
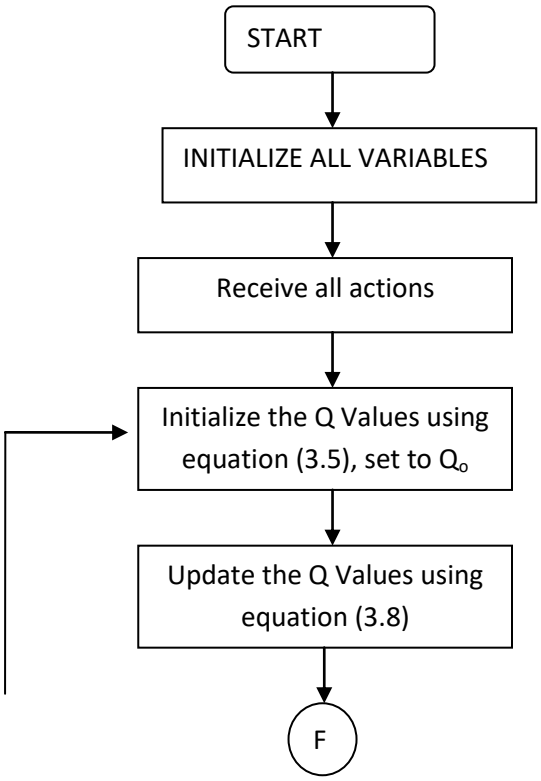


Fig 6 Flowchart for The alarm SRP procedure procedure

The fourth and final stage is the Q-value update, here the cost function is a weighted function of the number of hop count and the maximum remaining energy of the nodes on the path to the sink. This implies that data will be routed through nodes with high remaining energy even if it results in a higher hop count in preference to nodes with low remaining energy. This process continues until the Q-values converge to the optimal value for all the nodes in the network.. The flowchart for the Q-value update procedure is shown in figure 7.

Fig 7 Flowchart for updating Q-values in SRP



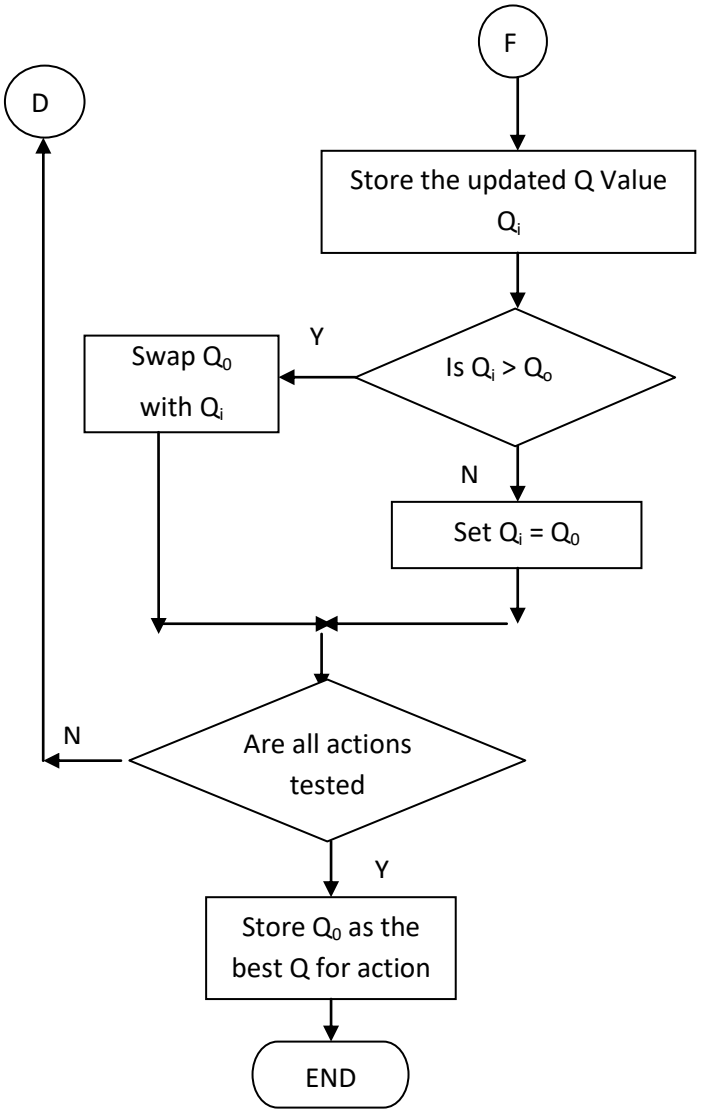


Fig 7 Flowchart for updating Q-values in SRP

V. RESULTS AND ANALYSIS

Experiments were conducted in a simulated environment as well as on hardware test-bed to compare the performance of SRS with RFSN [2], CONFIDANT [13] and [14]. To show the usefulness of alarms SRP, (AS), the results of SRP with, and without AS was compared. It is denoted by SRP and SRP-NAS, respectively. To verify the usefulness of the hierarchical structure, SRP was implemented without any hierarchy, but the method failed to find a reasonable solution (due to the large state/action space), thus not shown in the results. The metrics used for characterizing the WSN security are: the average Packet Delivery Ratio (PDR) i.e, ratio of data packets successfully delivered to the sink and Residual Energy (RE) i.e., average (remaining) energy of each sensor node in the network.

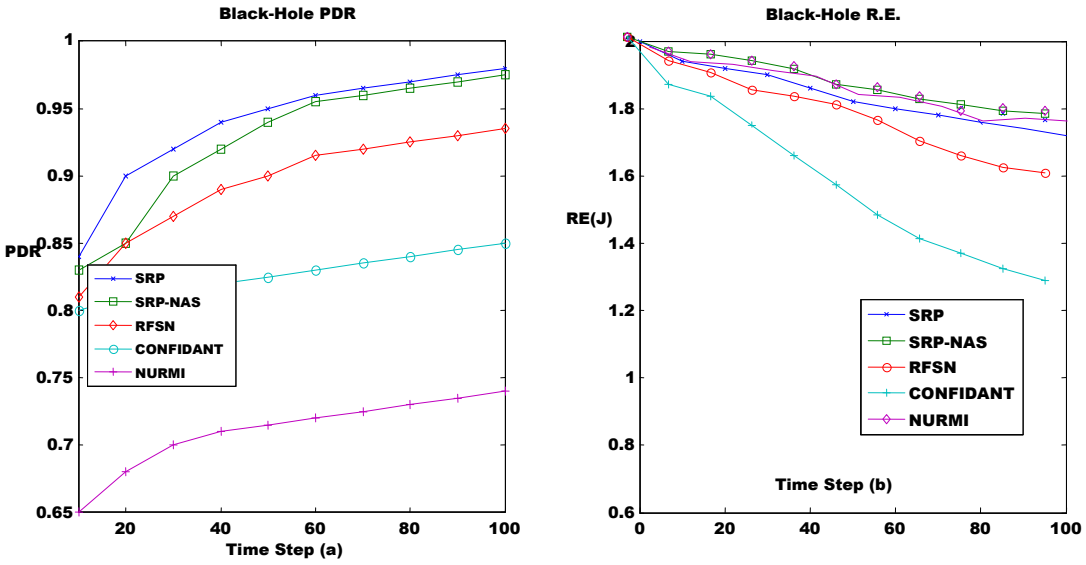


Fig 8(a-b) Graph of (PDR) and Residual-Energy (RE) under Black-hole adversary

For simulation, the MATLAB Simulator was used. The topology includes 100 stationary nodes, uniformly randomly distributed within a 1000m X 1000m square, with the sinks at its right end. The transmission radius is 100m and M=5 (i.e. number of neighbours). Each node generates packets at the rate λ=1 per time step. The size of each data packet is 512 bytes, HELLO packet is 60 bytes, QUERY, ALARM and ACK packet is 125 bytes. The initial energy of each sensor node is 2J. The radio dissipates 50 nJ / bit to run the transmitter/receiver circuitry and 100 pJ / bit for the transmitter amplifier. 20% of the nodes were assumed to be compromised. The experiments are run for 100 time steps, transmitting over 10,000 data packets.

In Fig. 8 (a-b), the following can be deduced from the simulation: (i) The PDR (Packet delivery ratio was highest in the SRP and SRP-NAS (97%), under the black-hole attack, here adversarial nodes drop all data packets routed through them. It can also be seen that SRP shows a slightly better performance than SRP-NAS because it quickly identifies adversarial nodes, while in the case of SRP-NAS, it first routes through the adversarial node before it learns of its integrity.

NURMI as opposed to the other protocols considers routing directly to a neighbour node without seeking integrity factor rating (this is the observation received about the recommendation given by a node about other nodes in the network, also referred as the rating function) from other nodes in the network before routing. This can be seen by its low PDR (73%). Conversely it has the highest remaining energy of all the compared protocols as it doesn't seek integrity factor rating however it pays more proportionally for this by its very low PDR. (ii) From Fig. 8(b), it can be seen that SRP has a lower residual energy compared to SRP-NAS, however the reduction is minimal. This can be attributed to alarm sub-function that is included in the protocol, where the integrity factor rating from other nodes can lead to the execution or non execution on the danger/alarm procedure, this incurs higher communication overhead. The RE values for SRP and SRP-NAS is 1.83J and 1.85J respectively. The residual energy of CONFIDANT is the lowest. This is due to its continuous integrity factor rating recommendation from many nodes in the network before routing data. This causes huge communication overhead with the resultant a lower residual energy. Its RE is 1.2J.

The graph in Fig 8(c-d) gives comparable results to Fig 8(a-b). In this simulation the on-off attackers drop packets every 5 time steps. SRP and SRP-NAS achieve the highest PDR (97% and 95% respectively after 100 time steps). The PDR of SRP is a slight improvement over that of SRP-NAS, because it can quickly identify the on-off adversarial nodes, especially in the beginning.

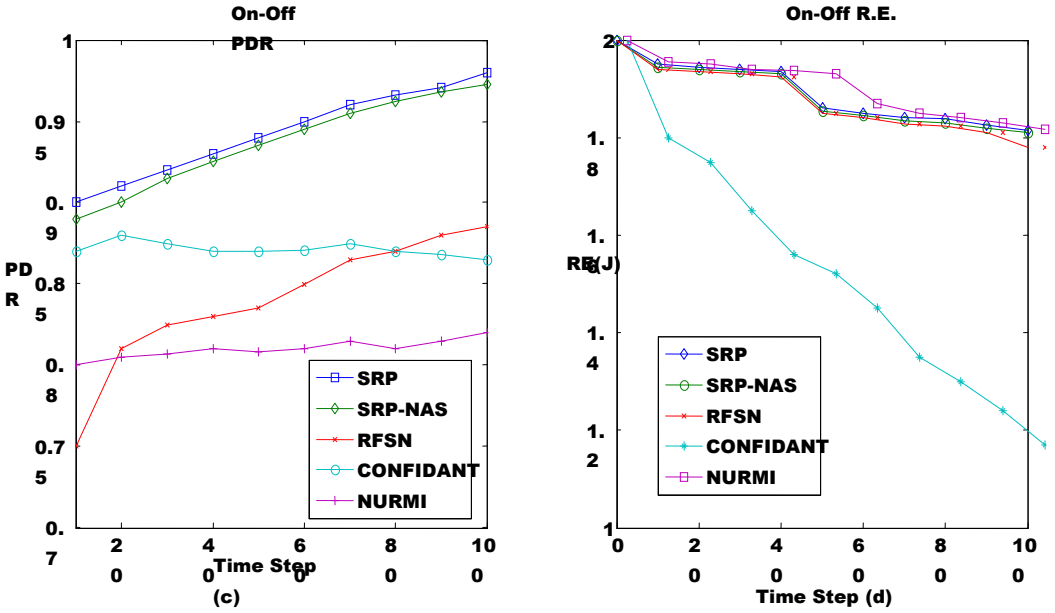


Fig 8(c-d) Graph of Packet Delivery Ratio (PDR) and Residual-Energy (RE) under on-off attack.

However their performance was at par towards the end of the simulation. This is so because after a time interval the SRP-NAS protocol must have identified the on-off adversarial node, thereby excluding them from routing data.

NURMI as opposed to the other protocols considers routing directly to a neighbour node without seeking integrity factor rating from other nodes in the network before routing. This can be seen by its low PDR (73%). Conversely it has the highest remaining energy of all the

compared protocols as it doesn't seek integrity factor rating however it pays more proportionally for this by its very low PDR. Its RE was 1.9J (92%). As opposed to SRP, SRP-NAS, RFSN, CONFIDANT which use both direct evaluation and recommendations, NURMI uses only direct evaluation. This accounts for the low PDR (73%), for NURMI as adversarial nodes that target the trust mechanism were able to give false reputation of a neighbour node. the other being the use of gradient techniques for computing policy. SRP achieves a lower residual-energy than SRP-NAS (90% and 91% respectively), as SRP usually activate the alarms sub-function any time it receives low integrity factor index about a node. This operation usually results in additional communication overhead. RFSN continuously seek integrity factor index from all neighbours before routing while CONFIDANT continuously activates the alarms sub-function any time it receives low integrity factor index from majority of neigbouring nodes, as it is being modeled in its observation. . This is the cause of its lower residual energy of 87% and 75% respectively. As Nurmi does not query other nodes, it achieves a high residual-energy of 1.92 J.(96%). However despite the higher functionality in SRP and SRP-NAS, they still have high remaining energy 1.85J (92.5%) after 100 timestamps. This is due to the use of Q-learning model used in determining the fitness of node and the avoidance of adversarial nodes which could cause quick depletion of the sensor node's energy.. CONFIDANT has the lowest remaining energy (1.2J) due to its relentless sending of alarms about malicious nodes.

In Fig. 8 (e-f), under probabilistic attack, adversarial nodes randomly send and drop data packets routed through them, hence their behaviour is deceptive. Under this scenario SRP, and SRP-NAS achieve high performance 96% and 94% PDR respectively because they can easily identify adversarial nodes with this deceptive behaviour as it is part of the observation probability computation. RSFN and CONFIDANT perform worse 84% and 79% PDR respectively due to the fact that it takes time for the protocol to recognize this type of deceptive adversarial nodes. While NURMI performs worst 72% PDR, as it cannot identify such adversarial nodes, thus it is

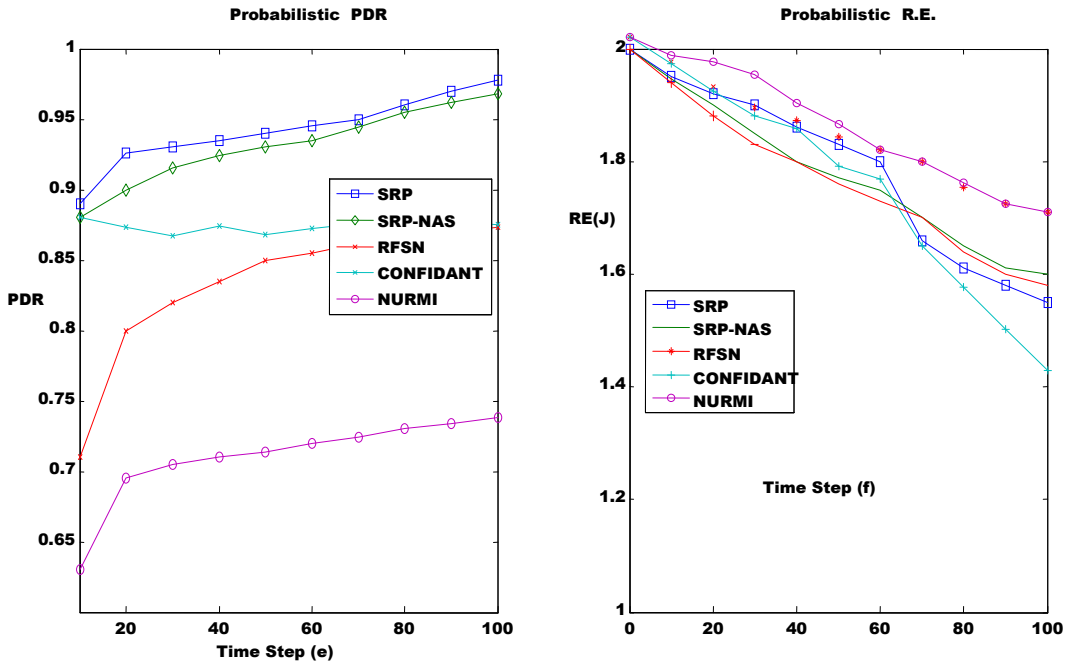


Fig 8(e-f) Graph of Packet Delivery Ratio (PDR) and Residual-Energy (RE) under probabilistic attack

adversely affected by adversarial nodes that sends false reputation about neighbour nodes.. However in terms of remaining energy (RE) NURMI achieves a slightly higher residual energy of 1.63 J than both version of SRP as NURMI does not query other nodes however the deceptive nature of the adversarial nodes cause high energy drain mainly at the beginning of the simulation,. However despite the higher functionality in SFROMS and SFROMS, they still have high remaining energy 1.54J and 1.59 J respectively after 100 timestamps. This is due to the use of Q-learning model used in determining the fitness of node and the avoidance of adversarial nodes which could cause quick depletion of the sensor node's energy. CONFIDANT has the lowest remaining energy (1.2J) due to its continuous activation of the danger sub-function whenever it receives a low integrity factor index about the reputation of a node from other neighbouring nodes..

In Fig. 8(g-h), under the importance attack, the adversarial nodes have other nodes that are sub-nodes to it, so whenever data is routed through them they send it to their child nodes which results in a high number of hop count to the destination. It may even result in a face routing problem where data packet travel in an infinite loop. In the importance attack, the adversarial nodes are increased to 60%, making them the majority due to the number of sub-node that depend on them for routing.. Here SRP with PDR 96% performs better than SRP-NAS, as it is able to isolate such adversarial nodes by activating the danger sub-function, while SFROMS-NAS with PDR 92% initially obtains mis-routed information from the adversarial nodes. This causes its initial routing through such nodes, until they are recognized after routing. NURMI has the lowest PDR of 72%. In terms of remaining energy (RE) SRP-NAS has a higher remaining energy to SFROMS, 80% and 76% respectively. NURMI has the highest RE of 85% while CONFIDANT has the lowest RE of 65%

In general from Fig 8(a-h) it was shown that AS improves the performance of SRP i.e.(PDR of SRP is always greater than SRP-NAS, although AS involves additional energy drain, in some cases). Also SRP outperforms the compared secured routing protocols by between 8% - 25%.

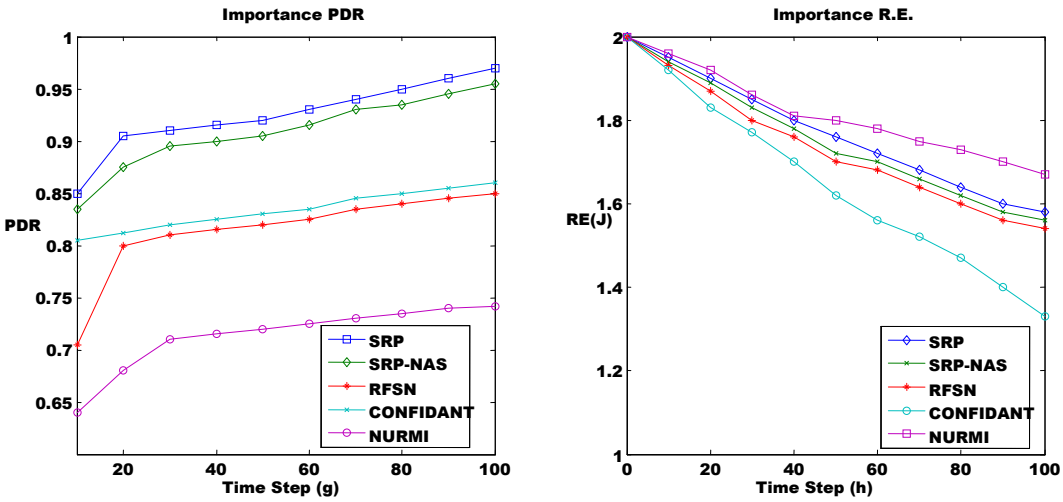


Fig 8 (g-h) Graph of Packet Delivery Ratio (PDR) and Residual-Energy (RE) under importance attack

In order to validate the secured routing protocol, in a real-world test-bed, the performance of SRP was compared with RFSN, CONFIDANT and NURMI. The experimental setup consists of arduino-uno (microcontroller), programmable xbees (radio transceiver) and LM 35 temperature sensor (sensing device) the combination of arduino uno xbee and LM 35 temperature sensor acts as the end device while the combination of the arduino uno and xbees acts as the router and co-ordinator nodes. The results of SRP were compared with and without AS denoted by SRP and SRP-NAS, respectively. Two performance metrics were used for comparison: The average Packet Delivery Rate and Residual Energy (RE). The neighborhood radius is 75m, size of packets is 62 bytes, initial node energy is 2J. For this purpose the individual protocols uploaded separately onto the arduino board, while the PDR and RE results were taken by connecting the laptop connected to the co-ordnator node with MATLAB Support package for Arduino. In Fig 9(a) under importance attack, the results show that SRP with PDR 91:5% outperforms other compared protocols. by between 15 to 25% , while in fig 9(b) SRP with the exception of NURNI was able to conserve the energy in

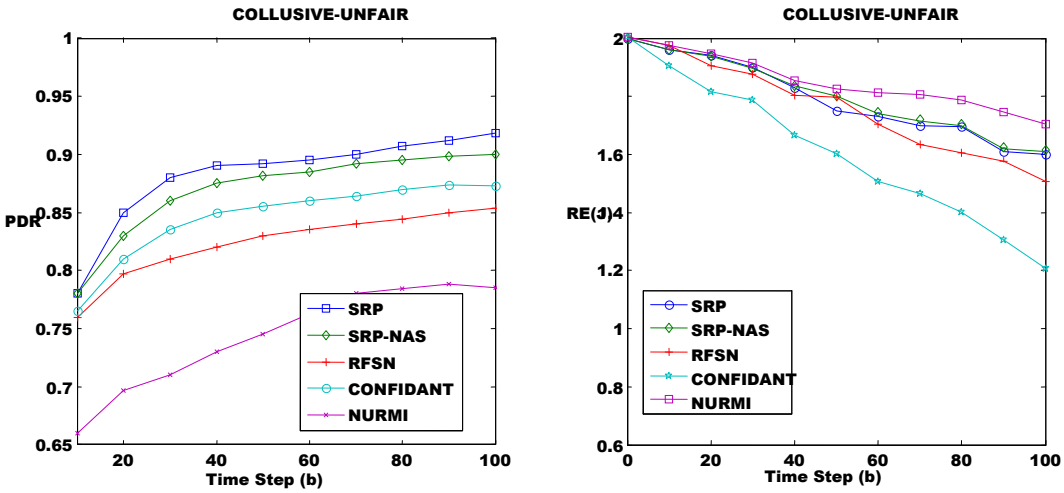


Figure 9 : Graph on the Collusive unfair adversarial nodes (Hardware Test-bed)

comparison to the other protocols by between 0.2 – 0.5J within the simulation period. This is due to its conservative propagation of alarms only where appropriate unlike CONFIDANT AND RSFN that repeatedly propagates alarm. The higher residual energy in NURMI is due to the fact that it doesn't propagate alarms, which is the reason for its low PDR of 75%.

VI. CONCLUSION

The Secure Routing POMDP (SRP) approach is presented in this paper, to select suitable next-hop neighbours and successfully route packets to the sink. It is a subset of the protocol to route to multiple mobile sinks. SRP can deal with black-hole, on-off attacks, etc., and other attacks targeting the trust system. It balances the exploration/exploitation tradeoff in gaining/exploiting information about sensor nodes, thereby effectively addressing their energy constraints. Experiments both in simulation and on hardware test-bed show that SRP consistently achieve higher packet delivery rate by coping with various categories of adversarial nodes, while still maintaining high residual energy.. Hence it guarantees secure and energy-efficient routing in WSNs. This paper has established that SRP is robust against various categories of adversarial threats that can compromise the trust mechanism employed in current secured routing protocols in WSN.

REFERENCES

[1]. Mac Ruair´ı and Keane Mark, (2007). An energy-efficient, multi-agent sensor network for detecting diffuse events. In International Journal of computational intelligence , pp 23-40
[2]. Ganeriwal Saurabh, Laura Balzano, and Mani Srivastava, (2008) Reputation-based framework for high integrity sensor networks. ACM Transactions on Sensor Networks (TOSN), Vol 4 pp 3 -15.
[3]. Rezgui Abdelmounaam and Eltoweissy Mohammed, (2007). Tarp: A trust-aware routing protocol for sensor-actuator networks. In MASS, 2007.pp 125-136
[4]. Walkins C. (1995) Learning from Delayed Rewards. PhD thesis, Kings College, Cambridge, England
[5]. Bertsekas D. P and J. N. Tsitsiklis. J. N (2001) Neuro-Dynamic Programming. Athena Scientific, Belmont, Massachusetts, .
[6]. Jaakaala T, Jiang Siwei , Jie Zhang, and Yew-Soon Ong (1999) An evolutionary model for constructing robust trust networks. In Journal of Artificial Intelligence pp 180 - 194.
[7]. Even E and Mansour Y . (2006) Learning rates for Q-learning. Journal of Machine Learning Research, Vol 5: pp 1–25,
[8]. Szepzsv C .S and Szita I (2010) Model-based reinforcement learning with nearly tight exploration complexity bounds. In Proceedings of the 27th International Conference on Machine Learning, pp 1031–1038
[9]. Irissappane Athirai, Frans Oliehoek, and Jie Zhang (2014). A POMDP based approach to optimally select sellers in electronic marketplaces. In AAMAS,2014 pp 67 - 79.
[10]. Poupart Pascal (2005). Exploiting structure to efficiently solve large scale Partially Observable Markov Decision Processes. PhD thesis, University of Toronto
[11]. Ross Stephane, Joelle Pineau, Sebastien Paquet, and Brahim Chaib-draa (2007). Online planning algorithms for POMDPs. Journal of Artificial Intelligence Research, Vol 32(1) pp 663–704.
[12]. Marti Sergio,Thomas J. G, Kevin L, and Mary B (2000). Mitigating routing misbehavior in mobile ad hoc networks. In MobiCom 2000 Vol 6 pp 67 - 78
[13]. Buchegger and Le Boudec, (2002) Performance analysis of the CONFIDANT protocol (Cooperation of nodes: Fairness in dynamic ad-hoc networks). In MobiHoc, 2002, pp 102 - 116
[14]. Nurmi Petteri (2007). Reinforcement learning for routing in ad hoc networks. In WiOpt, pp 124 - 138.
[15]. Zhang Shiqi and Sridharan Mohan, (2012). Active visual sensing and collaboration on mobile robots using hierarchical POMDPs. In AAMAS, pp 45 – 54.
[16]. Pineau Joelle and Thrun Sebastian (2002). An integrated approach to hierarchy and abstraction for POMDPs. Carnegie Mellon University Technical Report CMU-RI-TR pp 2-21.
[17]. Theocharous Georgios (2002). Hierarchical learning and planning in partially observable Markov decision processes. PhD thesis, Michigan State University.
[18]. Foka Amalia and Trahanias Panos, (2007). Real-time hierarchical POMDPs for autonomous robot navigation. Robotics and Autonomous Systems, vol 55(7) pp 561–571