Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 6, July, 2021: 8643- 8654

Research Article

Secure Data Aggregation Technique using Audit based scheme for Wireless Sensor Network

N.Nithya.MCA. Mphil[1], Dr.N.Rajendran.PhD[2]

Abstract

Wireless Sensor Network is resource-constraint nature of a sensor network makes it an ideal medium for malicious attackers to intrude the system. Thus, providing security is extremely important for the WSN due to limited computational power and energy resources. Aggregation of data from multiple sensor nodes done at the aggregating node is usually accomplished by simple methods such as averaging. However such aggregation is known to be highly vulnerable to node compromising attacks.WSN is usually unattended and without tamper resistant hardware, they are highly susceptible to such attacks. Thus, ascertaining trustworthiness of data and reputation of sensor nodes is crucial for WSN. As the performance of very low power processors dramatically improves, future aggregator nodes will be capable of performing more sophisticated data aggregation algorithms, thus making WSN less vulnerable. Iterative filtering algorithms hold great promise for such a purpose. In this paper, the data generated from the sensor nodes is aggregated from multiple sources and provide trust assessment of these sources, usually in a form of corresponding weight factors assigned to data provided by each source. It implicitly provides the data traffic balancing and energy consumption by reducing the overhead of the network with reference to cache of data on audit based strategies. Analytical and simulation results on the proposed model proves that Secure Audit scheme based on data aggregation protocol provides better performance terms of packet delivery ratio , Network overhead, transmission delay, computation time and packet loss

Keywords: Wireless Sensor Network, Dempster-Shafer Theory, Trust Model, Data Aggregation, Secure Model

1. Introduction

Wireless Sensor Networks composed of multiple sensor nodes which deployed to monitor the environment and track the objects or application[1]. Each data from multiple sensors is aggregated at an aggregator node which then forwards to the base station only the aggregate values. Data aggregation can significantly reduce the amount of data transmitted to the base station, therefore improve the energy efficiency and prolong the wireless network lifetime. A routing protocol is used to decide the best route suitable for sending data to the sink from a source node[2]. Due to dynamic environment in which a WSN is deployed, it is not sufficient to rely on a single path between the source and destination. In order to deal with such problems,

[1] Assistant Professor, Vivekananda Arts& Science College for women
[2]Principal, Vivekananda College for Women

multiple paths between source and destination need to be discovered, so that if one path fails, data can be transmitted to another discovered path[3]. So, there is a need to develop a routing protocol which is efficient in terms of resource consumption (energy resource is the major concern) and perform data aggregation by discovering multiple paths between source and destination [4].

Data aggregation technique is energy efficient if it provides the maximum functionality with minimum energy consumption in WSNs. Energy efficiency is a ratio of amount of data successfully transferred in a sensor network to total energy consumed to transfer those data. Data from multiple sensors is aggregated at an aggregator node which then forwards to the base station only the aggregate values. At present, due to limitations of the computing power and energy resource of sensor nodes, data is aggregated by extremely simple algorithms such as averaging. In paper, secure trust audit scheme for data aggregation has been deployed using D-S belief theory to evaluate the trustworthiness of mobile nodes propagating the networks.

The rest of paper is organized as follows. Section 2 provides the related works on trust modelling in Secure Data Aggregation applications. Section 3 defines and designs the proposed Secure Audit Scheme for data aggregation applications towards data routing. In Section 4, Simulation environment, simulation results and performance of the proposed model has been detailed. Finally Section 5 concludes the paper

2. Related works

In this section, various existing model enumerating Data Aggregation system with desirable properties to trust model has analysed on basis of data routing and aggregation aspects towards secure communication between nodes in detail as follows

2.1.Dynamic Multiobjective Heuristics defined Signature based Scheme

The signature mechanism combined with key based mechanism is used to select the efficient and appropriate sensor nodes for effective data transmission on data aggregated information by the sensor nodes in the cluster based network. Due to high node density in sensor networks, same data is sensed by many nodes, which results in redundancy. This redundancy can be eliminated by using data aggregation approach while routing packets from source nodes to base station.

2.2.Merkle hash tree based mechanism

In most of the secure data aggregation protocols, Merkle Hash tree is used as a very important tool for data verification. Merkle Hash tree based scheme not only can keep data integrity, but also can reduce bandwidth and storage cost for WSNs. In energy and data correlation based data aggregation, fuzzy logic is used to select the CH and distribute the equal amount of load among clusters to transfer data in optimal way with minimum energy consumption.

3. Proposed model

In this section , holistic approach named as Secure Audit Scheme based data aggregation model based on the Dempster–Shafer theory has been developed to evaluate the trustworthiness of interacting nodes on data routing has been designed using various trust establishing step as follows

3.1.Network Model

The sensor nodes are divided into disjoint clusters, and each cluster has a cluster head which acts as an aggregator. Data are periodically collected and aggregated by the aggregator. We assume that the aggregator itself is not compromised and concentrate on algorithms which make aggregation secure when the individual sensor nodes might be compromised and might be sending false data to the aggregator.

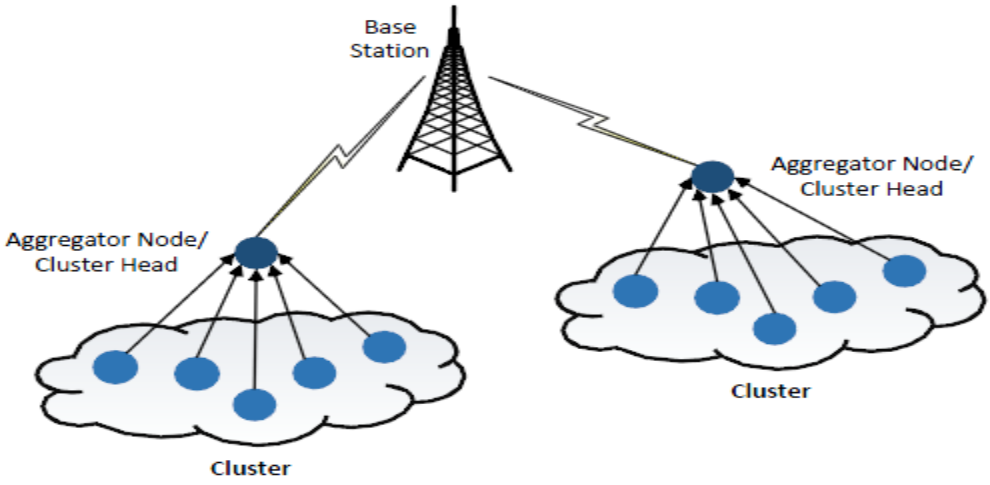

Figure 1 Network Model of Data Aggregation process

In figure 1, Wireless Sensor Network is considered with n sensor nodes which is represented as S_i,i=1,....n. In this model, assumption has been made such that node information is aggregated on the reading at a time of m consecutive instants. The aggregation information is represented as matrix X = {x1,x2,x3...x_n} represents m dimensional reading of the node S_i.

Let r={r1,r2...r_m}^{T} denote the aggregation value of instant t=1,2,3...m, which is represented as reputation vector computed iteratively and simultaneously with a sequence of weights w=[w1,w2,w3....w_n]T reflecting the trustworthiness of sensors[7]. The iterative procedure starts with giving equal credibility to all sensors with initial value w(0)=1. The value of the reputation vector r(l+1)in round of iteration l+1 is obtained from the weights of the sensors obtained in the round of iteration l as

$$R^{(l+1)} = \frac{X.w(l)}{\sum w(l)}$$

The reputation vector is just the sequence of simple averages of the readings of all sensors at each particular instant. The new weight vector w(l+1) to be used in round of iteration l+1 is then computed as a function g(d) of the normalized belief divergence d which is the distance between the sensor readings and the reputation vector $r^{(l)}$.

$d= [d1,d2,d3...d_n]^T$

$$d_i = \frac{1}{m} || x_i - r^{(l+1)} ||^2$$

In this g(x) is called the discriminant function and it provides an inverse relationship of weights to distances d. Discriminant function was a reciprocal of the distance between sensor readings and the current computed reputation.

Reciprocal: $g(d) = d^{-k}$

Exponential : $g(d) = e^{-d}$;

Affine: $g(d) =1- k_1 d$ where $k_l>0$ is chosen so that

$g(\max_i\{d_i^{(l)}\})=0$

A robust variance estimation method in the case of skewed sample mean is an essential part of secure aggregation of the sensed data towards planning the routing. Further the stochastic components of sensor errors are independent random variables with a Gaussian distribution [6]. Moreover, if error distribution of sensors is either known or estimated, it can be adapted to other distributions to achieve an optimal performance.

3.2.Secure Data Aggregation

The secure aggregation method operates with three stages. In the first stage, initial estimate of two noise parameters for sensor nodes, bias and variance has been provided. On the details of the computations for estimating bias and variance of sensors are presented. Based on such an estimation of the bias and variance of each sensor, the bias estimate is subtracted from sensors readings and in the second phase of the proposed model, an initial estimate of the reputation vector calculated using the MLE has been provided has been provided as notation in the table 1

Table 1: Notation Employed for trust computation

Notation	Description
m	Number of reading for each sensor
n	Sensor reading
r^t	True Node Density of node at time t
x^t	Data from sensor S
e^t	Noise error at time t

The mean of the bias of all sensors is not zero and then there would be no way to account for it on the basis of sensor readings. On the other hand, bias of sensors, under normal circumstances, comes from imperfections in manufacture and calibration of sensors as well as from the fact that they might be deployed in places with different environmental circumstances where the sensed scalar might in fact have a slightly different value. Table 2represents the trace information of the node at attack scenario.

Table 2 Trace Information of the node at attack scenario

Sensor Reading					
Instant	S1	S2	S3	S4	S5
T=1	19.76	19.61	19.77	20.20	20.41
Sensor weights					
1st iteration	1	1	1	1	1
2nd iteration	6.68	8.17	6.27	3.48	2.74
3rd iteration	64.21	130.29	53.13	13.26	8.56

Gaussian distribution random variable with a sensor Bias b_s and sensor variance S_v. Let r_t denotes the true value of the signal at time t. Sensor reading x^t can be written as

$$x_s^t = r_t + e^t_s$$

The sensor reading is stored in matrix form. Matrix is an estimator for mutual difference of sensor bias. The mean of the bias of all sensors is not zero and then there would be no way to account for it on the basis of sensor readings. On the other hand, bias of sensors, under normal circumstances, comes from imperfections in manufacture and calibration of sensors as well as

from the fact that they might be deployed in places with different environmental circumstances where the sensed scalar might in fact have a slightly different value.

In the third stage of the proposed model, the initial reputation vector provided in the second stage is used to estimate the trustworthiness of each sensor based on the distance of sensor readings to such initial reputation vector.

3.3.Trust Computation

Trust Computation between Neighbor Nodes, obtains the trust value of node j by itself directly but also through k1, k2 and k3 indirectly. And then node i integrates the different kinds of trust values to get an integrated one (for node j). Then, the trust of nodes changes gradually depending on behavior, history and time. Trust in the network has the feature 'hard to acquire and easy to lose'. There are two trust initialization strategies: the pessimistic and the optimistic strategy.

- The pessimistic strategy can eliminate the possibility that the malicious node creates a new identity and impersonates a new node to rejoin in the network with the purpose of throwing away its bad trust value [14].
- Optimistic strategy is one has the foundation of completely trust at the beginning of network deployment, and is propitious to the quick network expansion. Therefore, in the initial period, trust value between neighbor nodes is depended on the factual application.

3.3.1. D-S Theory for Auditing of the Data

A malicious node does not execute protocols to identify misbehaviour, vote honestly, and revoke malicious attackers. In particular, it is able to falsely accuse a legitimate node to revoke its permission successfully. The so-called attacker node is defined as a special malicious node which can launch attacks on its neighbors to disrupt secure communications in the network.

In D-S scheme, these nodes can be further classified into three categories based on their reliability: normal node, warned node, and revoked node. When a node joins the network and does not launch attacks, it is regarded as abnormal node with high reliability that has the ability to accuse other nodes. Moreover, System should note that normal nodes consist of legitimate nodes and potential malicious nodes.

Nodes that are listed in the warning list are deemed as warned nodes with low reliability. Warned nodes are considered suspicious because the warning list contains a mixture of legitimate nodes and a few malicious nodes. Warned nodes are permitted to communicate with their neighbours with some restrictions, e.g., they are unable to accuse neighbors any more, in order to avoid further abuse of accusation by malicious nodes[8].

D-S theory achieves per-packet behavior evaluation without incurring a per-packet per-hop cost. AMD is a comprehensive solution that integrates identification of misbehaving nodes, reputation management, and trustworthy route discovery in a distributed and resource-efficient manner. AMD enables the per-packet evaluation of a node's behavior without incurring a per-packet overhead. The Dempster shafer theory is established on basis of theory of evidence and degree of belief on the nodes and it is based on principle of indifference. It is given as

$$P(H)+P(H^{|})=1$$

Where P=1/N.

D-S rule of Combination

$$M1+M2= \sum_{xny=z}^{n} m1(X) + m2(Y)$$

In D-S Theory Belief function is modelled on elements of the node which is mutually exclusive to each other. All possible elements in the node are considered as evidence on the basis probability assignment. Probability theory is considered as ignorance.

- AMD enables the concurrent first-hand evaluation of the behavior of several nodes that are not necessarily one-hop neighbors. Overhearing techniques are limited to one hop.
- AMD can operate in multi-channel networks and in networks with directional antennas. Current packet overhearing techniques are only applicable when transmissions can be overhead by peers operating on the same frequency band.
- AMD detects selective dropping behaviors by allowing the source to perform matching against any desired selective dropping patterns. This is particularly important when end to- end traffic is encrypted. In the latter scenario, only the source and destination have access to the contents of the packets and can detect selective dropping.

3.3.2. Reputation Model

AMD isolates misbehaving nodes by implementing a reputation based system. Nodes with low reputation values are excluded from routing paths, thus being unable to drop transit traffic[9]. The reputation module is responsible for computing and managing the reputation of nodes. We adopt a decentralized approach in which each node maintains its own view of the reputation of other nodes.

A reputation evaluation is considered to be first-hand, if it originates from the audit module running. This is because the audit module can make direct observations of the behavior of nodes in a path PSD based on behavioural audits. Due to the multiplicative factor α, the reputation of a misbehaving node rapidly declines with repeated misbehaviour. The difference between the reputation decrease and increase mechanism prevents a selectively misbehaving node from oscillating between periods of misbehaviour and good behavior for the purpose of dropping packets while remaining in active paths.

Parameters α, β can be empirically tuned according to network characteristics such as number of alternative paths (network connectivity degree) and expected congestion conditions. The reputation distribution operation implements a tradeoff between the communication cost of updating reputation values and obtaining a current view of the reputation of other nodes.

Such implementation alleviates the communication overhead for transmitting information to a centralized location, and readily translates to the distributed nature of ad hoc networks. Moreover, it allows nodes to hold individualized reputation metrics for their peers depending on their direct and indirect interactions[10]. The malicious nodes could distribute low reputation values for other nodes, in order to exclude them from the network (badmouthing).

3.4.Secure Route Prediction – Routing

The route discovery module is responsible for the discovery of trustworthy paths from a source to a destination. This module is invoked by the source whenever there is no cached path to the destination. Hence, a malicious node with low reputation value cannot increase the path reputation to a value higher than its own reputation. A malicious node can, however, lower the reputation value of a path by lying about the reputation values of other nodes. This strategy decreases the path reputation, leading to the exclusion of the lying node from routing paths[11].

Through Network delay and computing overhead of the network path, zone portioning is carried out. Partition the network fields into the Zones by employ hierarchical zone partitioning.

Hierarchical zone Partitioning carried out with vertical and horizontal portioning until source and destination not lies in same node

Z_d= Destination Zone

TD = Temporary Destination is randomly chosen in another zone to the Source in different zone

Destination node is calculated based on the node density and No. of nodes" H"

RF =Random Forwarder is Calculated using GPRS routing technique

Heuristic strategies of route planning against Node propagation

1. Data location service through Distance of the nodes

2. Zone partitioning based on the random Forwarder and Temporary destination

Zone Partitioning ∂ Temporary Destination with Node Density strength

3. All or Nothing transformation for Routing Control mechanism to minimize the Hop count between the source and destination, it also utilizes the dissemination time for data transmission.

The source node adaptively resizes each packet in its packet stream for each neighbor node according to the neighbour's mobility in order to increase the scheduling feasibility of the packets from the source node. While throughput guarantee is also important, it is automatically guaranteed by bounding the transmission delay for a certain amount of packets against various attacks propagating on various aspects of the channel and data propagating in the particular path.

4. Experimental Results

In this Section, we simulate Secure Audit Scheme for Data Aggregation Scheme in the Wireless Sensor Network using NS2 Simulator. Through extensive experiment, we demonstrate the properties and measure the network performance in terms of Throughput, Packet delivery ratio, Network Overhead and Packet Loss. The Proposed protocol through inclusion of Spatial and Temporal Constrainst mechanism on various attacks. In the Simulation, the set up of the network is described in the following table 1

Table 1: Simulation Parameters used to build a protocol

Simulation Parameter	**Value**
Simulator	NS2
Topology Size	1000m *1000m
Number of Nodes	200
Bandwidth of the Network	2Mbps
Traffic Type	CBR
Pause Time	10s,20s
Data Packet size	512 bytes

Buffer Size	30 packets
Simulation Time	30 minutes

The whole duration of mobile sink's moving trajectory on the path is divided into several consecutive time slots. The data collection hop count m = 3 is defined and use the nearest neighbor algorithm to find the shortest path [15]. Node energy consumption at different rates during transmission, reception, idle waiting and sleeping is calculated for evaluation of the proposed model

Trust computation is composed of a time domain anomaly detector and a trust model based on the Dempster–Shafer theory or Bayesian function has been evaluated on increase of the nodes. A trust analysis is then conducted based on the anomaly detection results against the various attacks towards data aggregation. The Dempster–Shafer theory or Bayesian function is used to model Mobile Sink behaviour patterns towards data collection and routing which is reliable.

The proposed secure audit mechanism decrease the overhead even then more data packet communication in the particular network on minimizing the hop distance and hop count of the particular data transmission period on multiple mobile sink. The Overhead is Reduced as the malicious node is dropped too much extent by data transmission strategies .The System also achieves high detection rate. Packets are exchanged over an end-to-end path to maintain channel synchronization between consecutive hops via dynamic multiple mobile sink. The performance in depicted in the figure 2

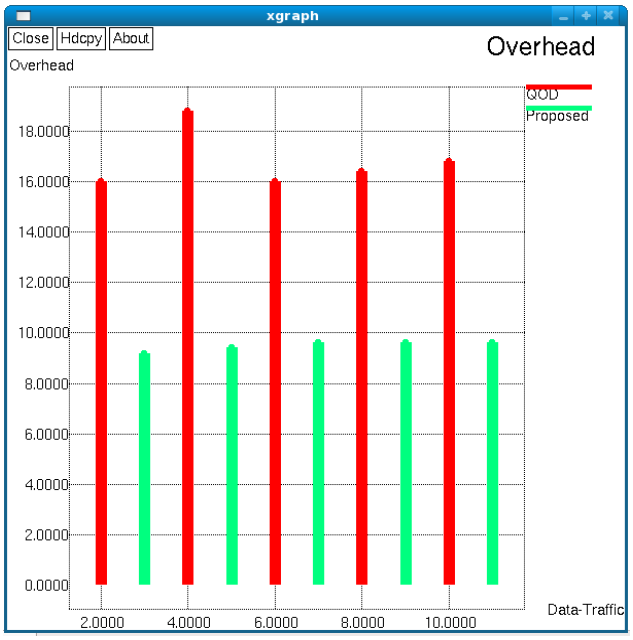

Figure 2 Overhead Analysis of the proposed Protocol against existing protocol via the Data Traffic

The packet loss analysis computation identifies the good transmission rate of the dynamic mobile sink towards data collection on various distance placement of the mobile sink. Security of audit and reputation scheme of the model increases the Network lifetime. Three

aspects decide the networks functionality and there by the network lifetime: the number of awake nodes, connectivity, and coverage on sleep state. Figure 3 depicts the packet loss of the network.

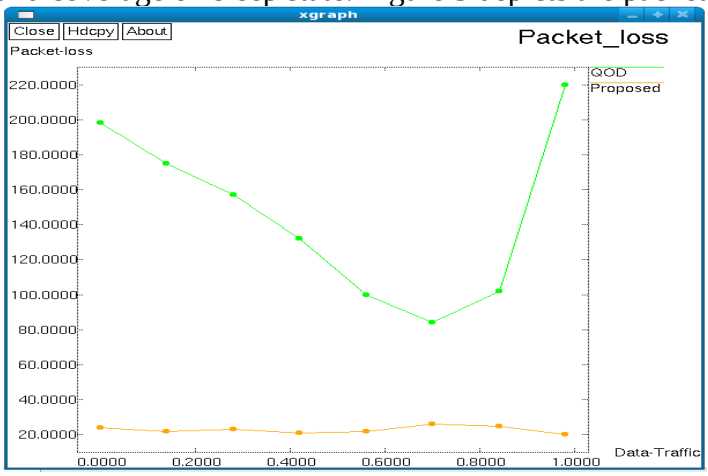

Figure 3. Packet Loss Analysis Proposed Protocol against the Existing protocol via Data Traffic

Energy Consumption is recorded by varying the node density of both sensor node and sink nodes on the simulation cycle for 50 seconds which is determine the data losses. Data losses are determined on basis of security establishment of the network. From the simulations it is found around 5 to 10 Joules of energy is preserved and this conservation increases as the node density of the neighboring node employed for transmission.

The lower packet delivery ratio is not just due to the node density but due to lower prediction accuracy during the start of the simulations which is depicted in the figure 4. A drastic reduction in number of data packets dropped at the end of simulation because of the reliability in selecting next hop links for the data packet [12]. This time delay therefore consists of the transmission times between the two points of a node on data collection. In Proposed Routing model, the increase in node density will decrease the average End to End Delay of the data collection through dynamic multiple sink nodes.

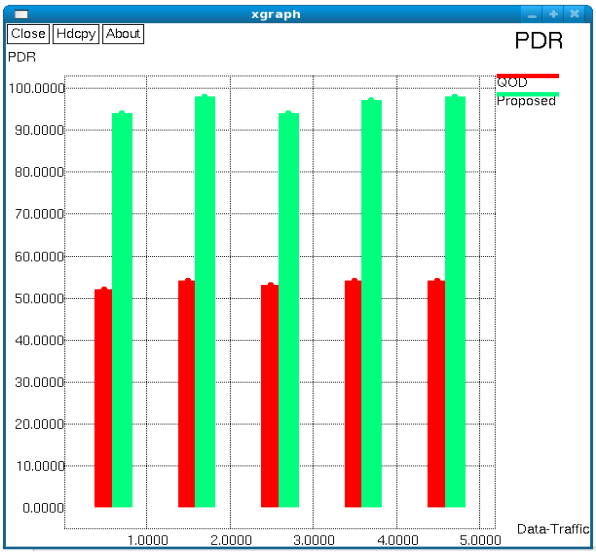

Figure 4. Packet Delivery Ratio Analysis Proposed Protocol against Existing protocol via the Data Traffic

The packet delivery ratio of transmitting packet on the specified network depends the transmission speed of dynamic mobile sink in the specified flooding topology. The impact is much reduced in the proposed model as collaboration of sink node and sensor node at short span of time towards effective data collection on basis of the D-S theory. Finally, the effect of the number of effective nodes M depends on the number of nodes with respects to successful data transmission evaluated. The network performance with respect to packet delivery ratio as described in the figure 5

The proposed work estimates data transmission rate on flooding constraints using throughput analysis on the efficient data cluster. Hence this evaluation is considered during routing process which optimizes route selection there by increasing the throughput of the data collection through dynamic multiple mobile sink on scheduling with hop measures as described in the figure 5.

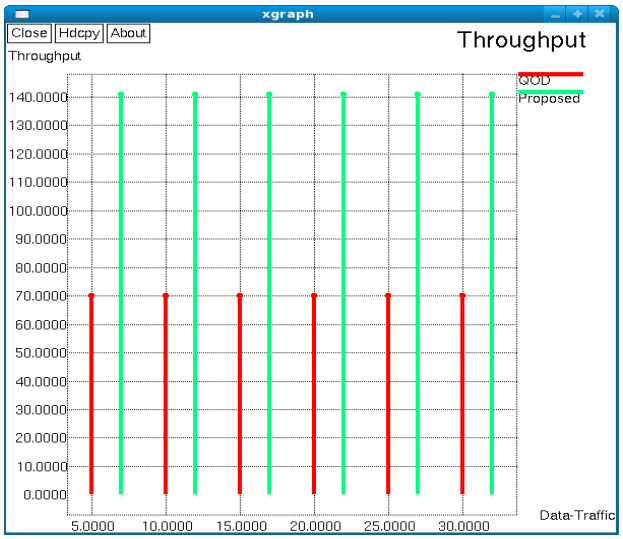

Figure 5: Throughput Analysis of the Proposed Protocol against Existing protocol through the Data Traffic

Also results shows the high performance is reached against most standard technique on varying working cycles of the audit scheme. As mentioned above, the number of neighbors depends nodes depends on speed of the mobile sink and node density of the node during flooding process to achieve effective energy utilization. Normalized configuration utilizes these parameters: number of nodes, deployment area, strategy, and radio link range of the reputation based model.

Table 4 – Performance Evaluation of the Dynamic Multiobjective Heuristics and secure Audit scheme against data traffic of the network

Technique	Throughput in mbps	Overhead in mbps	Packet Delivery Ratio	Packet loss in Percentage of data lost

Dynamic Multiobjective Heuristics Protocol – Existing	67.93	13.29	98.97	0.27
Secure Audit Protocol – Proposed	70.25	12.29	99.97	0.23

It is dynamic to changes such new sensing node and transmission of the dynamic multiple mobile sink in the network. Reduction in the queuing delay also manages the network efficiently. The performance comparison of the state of art approaches displayed as table with obtained values for each technique during security enhancement of the network on aspect of the secure framework.

As the number of source nodes in the system increases, the percentage of the delayed packets increases. This is because as more packets are generated, every packet in the scheduling queue needs to wait for more time to be forwarded out, which leads to higher delay and hence more delayed packets.

Conclusion

We designed and implemented trust computation model has been established to identify various attacking nodes especially selfish and malicious nodes efficiently on behavioural strategies and solve the security problems for node failure or capture in WSNs effectively using D-S Belief Theory

References

[1] Y. Yang, X. Wang, S. Zhu, and G. Cao, "SDAP: A secure hop-by-hop data aggregation protocol for sensor networks," ACM Transactions on Iriformation and System Security (TISSEC), vol. II, no. 4, pp. 18:1-18:43, 2008.

[2] A. Mahimkar and T. S. Rappaport, "SecureDAV: A secure data aggregation and verification protocol for sensor networks," in Proceedings of the 47th IEEE Global Telecommunications Conference (Globecom), 2004, pp. 2175-2179

[3] J.-W. Ho, M. Wright, and S. Das, “ZoneTrust: Fast zone-based node compromise detection and revocation in wireless sensor networks using sequential hypothesis testing,” Dependable and Secure Computing, IEEE Transactions on, vol. 9, no. 4, pp. 494–511, july-aug. 2012.

[4] S. Ozdemir and H. C¸ am, “Integration of false data detection with data aggregation and confidential transmission in wireless sensor networks,” IEEE/ACM Trans. Network., vol. 18, no. 3, pp. 736–749, Jun. 2010.

[5] H. Chan, A. Perrig, and D. Song, “Secure hierarchical innerwork aggregation in sensor networks,” in Proceedings of the 13th ACM Conference on Computer and Communications Security, ser. CCS ’06. New York, NY, USA: ACM, 2006, pp. 278–287.

[6] Y. Yang, X. Wang, S. Zhu, and G. Cao, "SDAP: a secure hopby- hop data aggregation protocol for sensor networks," in MobiHoc, 2006, pp. 356–367.

[7] S. Roy, M. Conti, S. Setia, , and S. Jajodia, "Secure data aggregation in wireless sensor networks," Information Forensics and Security, IEEE Transactions on, vol. 7, no. 3, pp. 1040–1052, 2012.

[8] E. Ayday, H. Lee, and F. Fekri, "An iterative algorithm for trust and reputation management," in Proceedings of the 2009 IEEE international conference on Symposium on Information Theory - Volume 3, ser. ISIT'09, 2009, pp. 2051–2055.

[9] H. Liao, G. Cimini, and M. Medo, "Measuring quality, reputation and trust in online communities," ArXiv e-prints, Aug. 2012.

[10] B.-C. Chen, J. Guo, B. Tseng, and J. Yang, "User reputation in a comment rating environment," in Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, ser. KDD '11, 2011, pp. 159–167.

[11] C. T. Chou, A. Ignatovic, and W. Hu, "Efficient computation of robust average of compressive sensing data in wireless sensor networks in the presence of sensor faults," Parallel and Distributed Systems, IEEE Transactions on, vol. 24, no. 8, pp. 1525–1534, Aug 2013.

[12] D. Wagner, "Resilient aggregation in sensor networks," in Proceedings of the 2nd ACM workshop on Security of ad hoc and sensor networks, ser. SASN '04, 2004, pp. 78–87.

[13] M. Rezvani, A. Ignjatovic, E. Bertino, and S. Jha, "Secure data aggregation technique for wireless sensor networks in the presence of collusion attacks," School of Computer Science and Engineering, UNSW, Tech. Rep. UNSW-CSE-TR-201319, July 2013.

[14] B. Awerbuch, R. Curtmola, D. Holmer, C. Nita-rotaru, and H. Rubens, "Mitigating byzantine attacks in ad hoc wireless networks," Department of Computer Science, Johns Hopkins University, Tech, Tech. Rep., 2004.

[15] M. C. Vuran and I. F. Akyildiz, "Spatial correlation-based collaborative medium access control in wireless sensor networks," IEEE/ACM Trans. Netw., vol. 14, no. 2, pp. 316–329, Apr. 2006.