Energy based-Genetically Derived Secure Cluster-based Data Aggregation in Wireless Sensor Networks

N.C. Nethravathi
M.Tech, Department of Computer Science and Engineering (PG),
NMIT, Yelahanka,

Prathibha A. Ballal
Associate Professor, Department of Computer Science and Engineering,
NMIT, Yelahanka

ABSTRACT

Wireless sensor networks (WSNs), are spatially distributed autonomous sensors used to monitor physical and environmental conditions, such as temperature, sound, pressure, etc. Each device can sense, process, and talk to its peers. All the sensor nodes are considered as little objects and therefore the data communication between sensor nodes is maximum. Extending of network period and information measure utilization becomes critical. This is achieved through the existing technique, routing protocol technique called destination sequenced distance-vector routing (DSDV). In the existing technique, initially it performs clustering and cluster head election process. Finally, the aggregated data from the cluster heads is transmitted to the sink. In the existing technique, the energy utilization is not efficient and it does not concentrate on network security. Therefore, in order to overcome these issues, a technique called, Energy-Based Genetically Derived Secure Cluster Based Data Aggregation (EB-GDSDA) clustering technique has been proposed for WSN. This technique highly minimizes the energy utilization and increases the life span of the network. In the proposed technique, the selection of CH is static throughout the simulation. To make cluster head selection dynamic, a Dynamic Selection-CH technique has also been implemented, where the user can have a choice of selecting the CH in each cluster based on the threshold residual energy. By simulation results, the performance of the Dynamic Selection-CH is more efficient than the existing and proposed static cluster head techniques, based on the simulation parameters.

General Terms

EB-GDSDA, RSA encryption-decryption, node connectivity, fitness function.

Keywords

WSN, data aggregation, clustering process, data security, packet delivery, energy consumption, packet drop, transmission overhead.

1. INTRODUCTION

WSN consists of low cost, little devices called sensor nodes and are in nature self-organizing ad hoc systems. The sensor nodes are considered as either routers (address) or computer-host (port), which has three components i.e., a CPU, a radio transceiver and a sensor array. It is a collection of sensing devices that can communicate wirelessly. Each device can sense, process, and talk to its peers. The job of the sensor network is to monitor the physical environment, gather and transmit the information to the sink node through other sensor nodes. Generally, the data has to be transmitted from one node to another node based on the radio range (specified in the parameter set up). Hence, the transmission of data is done in hop-by-hop to the sink in a multi-hop network. This process helps to reduce unwanted transmission or unwanted data (redundant data). Networks can also reduce the consumption of energy by aggregating similar data in the sensor nodes. Thereby increasing the network lifetime [3]. The main applications of WSN are wild habitat monitoring, forest fire detection, building safety monitoring, military surveillance and so on.

The main disadvantages of WSNs are: all the sensor nodes have been considered as small objects and the data communication between every sensor node is maximum. This leads to maximum failures of a network and causes severe bandwidth problem, measured in bits per second (bit/s) i.e., the rate of data transmission, bit rate and throughput, are different from each transmission to transmission. Sensor nodes generate more transmissions and thereby consume more energy to transmit the desired data. Hence, the amount of data transmission needs to be minimized to consume limited energy and make best use of the available resources [3]. This plays an important role in extending of network lifetime and bandwidth utilization. This can be achieved through data aggregation.

Data aggregation is nothing but collection of large amount of data from various sensor nodes within the network and aggregated into a data aggregator [8]. Finally collected data has to be sent to the sink. This leads to minimizing the number of transmission between nodes and avoids similar data to be transmitted (no redundant data). Hence the data aggregation is used to minimize the energy consumption and reduce the transmission overhead. This also enables enhancing the network lifetime [2]. However, in hostile environments, the aggregated data need to be protected from various attacks for attaining the data confidentiality, integrity and authentication. Hence security plays a major role in data aggregation.

As every sensing element has a finite battery supply, a vital feature of sensing element network is energy efficiency, to increase the network's lifespan. Each node in a sensor network is typically equipped with one or more sensors, a radio transceiver or other wireless communications device, a small microcontroller, and an energy source. Since in most Wireless sensor networks the energy source is a battery, energy consumption plays an important role, and reducing the consumed energy of each node is an important goal that must be considered when developing a routing protocol for wireless sensor networks. To do so first we need to identify a protocol which fits for a data aggregation application. In this paper we have proposed contributions of EB-GDSDA protocol for the data aggregation and to enhance the lifetime of the network.

In wireless sensor network, serious security threat is originated by passive attacks in cluster-based data aggregation [7]. It creates a high risk for data confidentiality and data authentication. Hence data confidentiality and authentication is an important issue in wireless sensor networks. To avoid these issues, we have demonstrated the proposed technique called, energy-based genetically derived secure cluster-based data aggregation (EB-GDSDA) in WSN. Finally EB-GDSDA has been compared with the DSDV protocol and showing that EB-GDSDA is having better performance than the DSDV protocol .

2. LITERATURE SURVEY

Lathies *et al.*[1] have projected GDSDA in WSN to produce less energy consumption, higher network economy and security. Initially, the CHs are selected based on the node connectivity, that acts as an information soul i.e., data aggregator (DAG). Then, the clustering process is executed using the genetic algorithm. This system extremely minimizes energy consumption and thereby enhancing the network period. When the cluster member wants to transmit the data to the aggregator, a data ncryption technique are utilized. The crypto module (CyM) utilized offers confidentiality to the information packet (DP), so guaranteeing the credibility and integrity of the perceived knowledge. The downside is that they have failed to consider time synchronization in capital punishment rule.

Mehrjoo et al. [2] have proposed a hybrid genetic algorithm (GA) and artificial bee colony (ABC)-based clustering algorithm. GA was used to choose the CHs and their number and ABC to select the clusters' members. The drawback of this approach is that the use of intelligent algorithms like ABC increases energy consumption.

Arijit et al.[3] have planned a theme for privacy preservation for secure knowledge aggregation in WSN. Within the planned theme (SMC), security of the information is ensured by a secured and strong key institution policy. Privacy of the information of the supply nodes is preserved by the thought of organization through standard arithmetic. However, in this work the communication overhead isn't mentioned within the overall theme. The most limitation of privacy-preserving knowledge aggregation in WSN is malicious attacks.

HevinRajesh and Paramasivan [4] have planned a method, consisting of 3 phases. Within the initial part, the network is split into clusters. The sensing element nodes with the upper signal strength are elected as CH. Within the second part, 3 parameters: distance, power consumed and also the trust price of the sensing element nodes were verified for each members. Within the third part, symbolic logic is employed to classify the sensing element nodes into best node, traditional node and worst node that support the chosen parameters. Finally the aggregated information is transferred by every cluster head to the sink. Since the values of malicious and faulty sensors don't seem to be aggregated, secure information aggregation is ensured within the wireless sensing element network. The downside of FBSDA approach is, not targeting network security potency and a lot of energy consumption.

Thakkar et al. [5] have planned a power-efficient cluster-based information aggregation protocol for WSNs. They performed CH choice and cluster formation supported residual energy criterion. Also, their multi-level information aggregation data among CHs minimizes the packet size, thereby minimizing transmission and reception energy.

3. PROBLEM STATEMENT

3.1 Overview

Examining the existing methods related to protected data aggregation, the following issues are noted: High communication overhead, high complexity, higher overhead whenever cryptographic technique is used, consumption of more bandwidth and no discussion about minimizing the energy consumed.

In this proposal, we propose to design a energy-based GDSDA secure data aggregation algorithm. This algorithm consists of 3 steps. The first step is to create initial population of sensor nodes. In the Second step, formation of cluster infrastructure and a sink node, the CH's are chosen from each cluster, based on their distance between the sensor nodes (node-connectivity formula) and in the third step, fitness function is used for the best CHs selection by using EB-GDSDA protocol. When a cluster member wants to transmit the data to aggregator, an energy-efficient RSA encryption-decryption technique is utilized, that offers authenticity, confidentiality and integrity.

3.2 Proposed system

In this paper, first we have created the initial population of sensor nodes moving from 0th position to their respective position (x-axis and y-axis). Second step is to form four clusters(Fig 1) and a sink node, the CH's are chosen from each cluster based on their distance between each of the sensor nodes (connectivity formula) and in the third step, we have used performance metrics formula for the best CHs selection by using EB-GDSDA protocol [1]. This technique highly minimizes the energy utilization. Thereby EB-GDSDA can increase the life span of the network. When the cluster member wants to send the data/messages to the CHs, a data/message encryption/decryption technique called RSA algorithm is employed which deals with the accuracy, reliability, validity, correctness, privacy, confidence, incorruptibility and reduces the broadcasting expenditure/consumption.

3.3 Energy based-Genetically derived secure cluster-based data aggregation algorithm.

The genetic algorithm is shown in the following.

Begin

- Generation of sensor nodes
- Deploy nodes in cluster infrastructure and a sink node
- Computing connectivity or distance formula for CHs selection
- Computing fitness of each individual for the best CHs selection
- Final CH selection is done based on fitness function.
- After CH selected in each cluster transmission starts in each cluster, where CH collects data from its respective cluster members(CM's)
- Finally data is transmitted from CH to sink.

End

3.4 Estimation of metrics

3.4.1 Fitness function: the fitness function of the sensor node is calculated based on the node distance, cluster distance and transmission energy [1].

$$f_i = 1/E_T + (D - D_c) + (n - n_c)$$

Where E_T is the energy required for data transmission from the cluster to sink. D is the sum of the distance of all nodes to sink. D_c is the sum of the distance of normal nodes to clusters and the sum of distances of all the clusters to sink. n is the number of nodes. n_c is the number of clusters.

In the above computation, N and D are kept as constants. Although the value of E_T, D_c and n_c keep varying.

3.4.2 Distribution factor: it gives the average number of nodes per cluster [1].

$$ɑ = N/C$$

Where N is the total sensor node counts. C is the total CH counts.

3.4.3Cost factor: it is estimated based on the distance among the sensor nodes[1].

$$ß = z + 10\ \rho \log 10(D/D_{ref}) + g$$

where, z is the path loss at D_{ref}, D_{ref} is the reference distance. D is the distance among the sensor nodes.

ρ is the path loss exponent in the range (2,4). g is the zero-mean Gaussian random variable. It gives the deviation in the path loss from its average value.

3.4.4 *Energy factor* : It is defined as the sum of the transmission energy (E_{tx}), reception energy (E_{rx}), energy in idle state (E_{idle}) and energy during sensing state (E_{sen}) of each CH.

$$Y = E_{tx} + E_{rx} + E_{idle} + E_{sen}$$

3.4.5 Residual virtual energy: The residual energy (E_r) of each node (N_i) is estimated using the following formula [1].

$$E_r = E_i - (Etx + E_{rx} + E_{en} + E_{dec} + E_a)$$

Where E_i is the initial energy of the node. E_{tx} is the energy utilized at the time of transmission of data. E_{rx} is the energy utilized at the time of reception of data. E_{enc} is the energy utilized for encoding the data. E_{dec} is the energy utilized for decoding the data. E_a is the energy required to keep the node active.

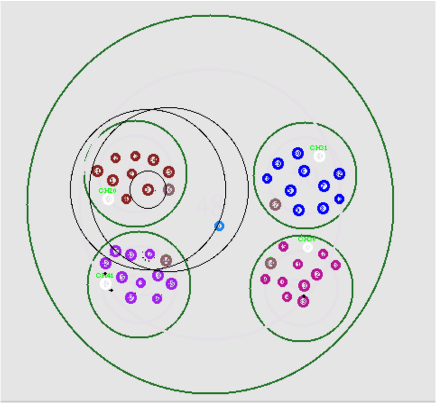

Figure 1: cluster formation

3.5 An Energy efficient RSA encryption-decryption technique

This technique provides the secure communication framework between each node. The key to encryption scheme varies as a function of residual energy of the nodes, thus preventing the rekeying requirements. All normal nodes will broadcast a unique value which contains their dynamic keys, to the selected set of nodes in first round of data aggregation. Let E_{ri} be the initial residual energy of Node N_i.

Let IV be the initialization vector pre-distributed to all the cluster members.

In RSA encryption-decryption technique, initially nodes are deployed in cluster infrastructure. Source node transmits the data along with its dynamic key to CH. Initially enter a node number which contains data to be transmitted. A node number is nothing but a dynamic key. Now entered node data will be encrypted which is in unknown binary format and encrypted data will be saved in CH. Then CH sends its encrypted data to the sink. The encrypted data will be decrypted by the sink node which is in readable (known binary) format.

3.6 EB-GDSDA Dynamic Selection-CH Technique.

In this system the user will have a chance to select the cluster head (CH) based on the inputs available for the user where, user can select the cluster head in each of the clusters. Once the CH is selected then transmission starts, if the selected CH is less than the threshold value than it dynamically selects the default CH which is having higher threshold or energy in each cluster.

The EB-GDSDA Dynamic Selection-CH technique is shown in the following.

Begin

- Nodes are deployed in cluster infrastructure.
- Four clusters with 1 cluster head (totally 4CH) and a sink are created
- Electing CH can be done by user itself i.e. dynamic CH selection.
- Inter communication between CH and CM's.
- Based on threshold value default CH will be elected i.e. if energy of particular node is less than threshold then dynamically default node will be selected.
- When the selection of CH is complete , all respective CM's send packets to its respective CH.
- Finally packets moves from CH to sink.

End

4. SIMULATION RESULTS

4.1 Simulation parameters

We evaluate our EB-GDSDA through ns-2.33 (NS-2),on Ubuntu operating system, front end as Tool Command Language (TCL) and protocol developed in C++. We use a bounded region 1000 X 1000 sq. in which we place nodes using a random distribution. The number of nodes is varied as 11, 12, 12 and 12. We assign the power levels of the nodes such that transmission range 250m. In our simulation, the channel capacity of mobile hosts is set to the same value: 11Mbps. We use the Distribution Coordinate Function (DCF) of IEEE 802.11 for wireless Local Area Networks (LANs) as the Media Access Control (MAC) layer protocol. The simulated traffic is Constant Bit Rate (CBR).

4.2 Performance metrics

The performance of EB-GDSDA Dynamic Selection-CH technique is compared with the existing DSDV and proposed EB-GDSDA. The performance is evaluated mainly, according to the following metrics. Table 1 summarizes the simulation parameters used for EB-GDSDA (similar simulation parameters for DSDV).

1) Average Packet Delivery Ratio**:** it is the ratio of the number of packets received successfully and the total number of packets transmitted.

2) Control overhead**:** the time taken to transmit data on a packet-switched network. The process of controlling the overall transmission speed of the raw data.

3) Bit error rate: it is the total number of packets dropped during the data transmission process.

4) Average Energy Consumption**:** the average energy consumed by the nodes in receiving and sending the packets.

5) Throughput : the number of packets received by the sink successfully.

The following table summarizes the simulation parameters used.

Table 1: simulation parameters

Number of nodes	11, 12, 12, and 12
Area size	1000 X 1000
MAC	802.11
Simulation time	25 s
Traffic source	CBR
Packet size	256
Transmit power	5000.0316w
Initial energy	10 J
Transmission rate	250m
Routing protocol	EB-GDSDA
Rate	20, 30, 40 and 50 kb

4.3 Results and analysis

In our experiment, four performance metrics are used to compare the performance of our protocols.

From Figure. 2, we can see that the packet delivery ratio of dynamic selection of CH (blue) is more than the EB-GDSDA (red) and the DSDV (green) with time v/s no. of packets (Kbit/s).

From Figure. 3, we can see that the control overhead of dynamic selection of CH (blue) is less than the EB-GDSDA (red) and the DSDV (green) with time v/s no. of packets (Kbit/s).

From Figure. 4, we can see that the bit error rate i.e., packet drop of dynamic selection of CH (i.e. blue) is less than the EB-GDSDA (i.e. red) and the DSDV (i.e. green) with time v/s no. of packets (Kbit/s).

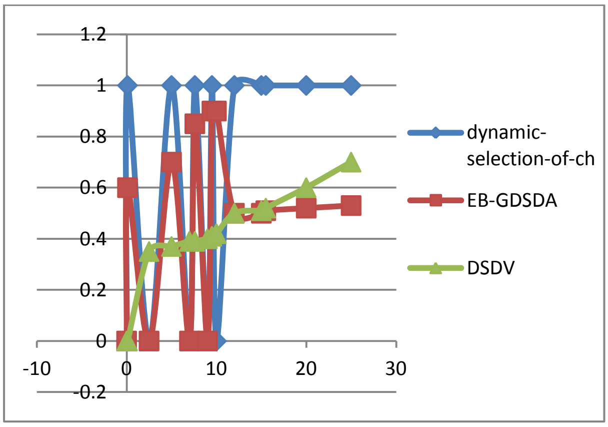

Figure 2. Time Vs no. of packets (Kbit/s)

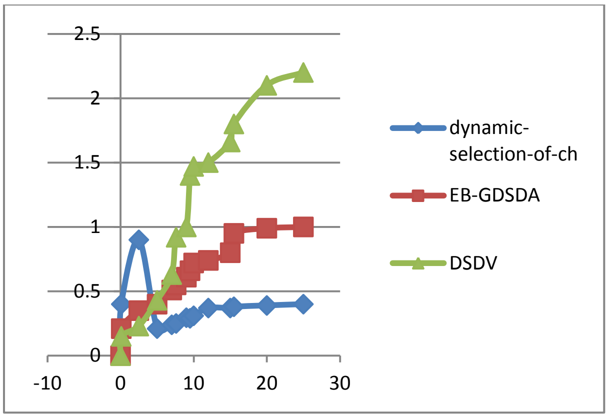

Figure 3. Time Vs no. of packets (Kbit/s)

From Figure. 5, it can be seen that the throughput of the dynamic selection of CH (blue) is more than EB-GDSDA (red) and DSDV (green). Hence less energy consumption in Dynamic Selection-CH than the EB-GDSDA and DSDV method with time v/s no. of packets (Kbit/s).

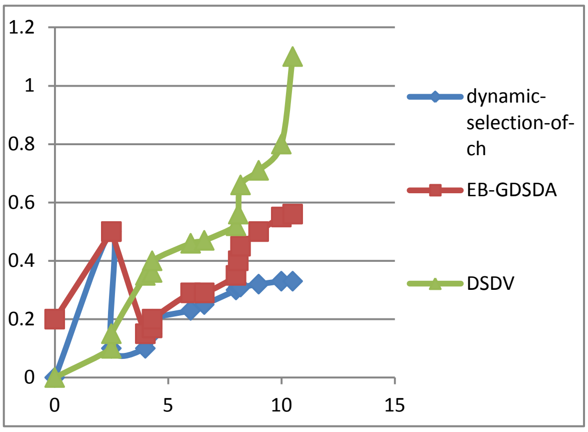

Figure 4 Time Vs no. of packets (Kbit/s)

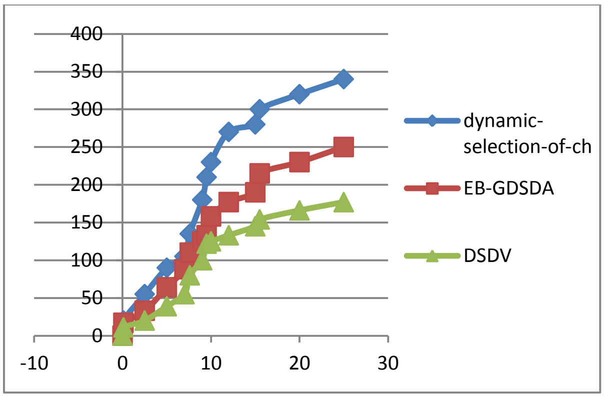

Figure 5 Time Vs no. of packets (Kbit/s)

Table 2 gives the comparison of the existing protocol DSDV, the proposed protocol EB-GDSDA and the EB-GDSDA dynamic selection of CH.

Table 2: Comparison between the protocols

Parameters	**Existing (DSDV)**	**Proposed (EB-GDSDA)**	**EB-GDSDA,Dynamic Selection -CH**
Packet Delivery Ratio	Low	Average	High
Control overhead	High	Average	Less
Bit error rate	Average	Average	Low
Throughput	Less	Average	High
Energy consumption	High	Less	Less

5. CONCLUSION

In WSN, all the sensor nodes are considered as little objects and therefore the data communication between each of the sensor nodes is maximum and utilization of energy is more. In order to overcome these issues, an Energy-Based Genetically Derived Secure cluster based Data Aggregation (EB-GDSDA) clustering technique is used to increase the life span of the network and an RSA encryption-decryption security technique has been proposed in order to provide efficient and secure data transmission. As the selection of CH is static throughout the simulation process, a Dynamic Selection-CH technique has been implemented, to make cluster head selection as dynamic. Hence, simulation results shows that the Dynamic Selection-CH has less energy consumption and high throughput in comparison with current and proposed techniques based on the performance metrics such as energy consumption, packet delivery ratio, bit error rate and control over head.

6. ACKNOWLEDGMENTS

Nothing is envisaged without the help and guidance of an experienced person respected in the field of concerned subject. Though the benefits achieved from them can never be adequately valued, I would like to express hearty gratitude towards them. I am grateful to thank my guide Ms.Prathibha Ballal Associate professor Dept. of CSE (UG) for guiding me throughout the work. I sincerely acknowledge her timely help, hearty guidance and constant support.

7. REFERENCES

[1] Lathies Bhasker : ‘Genetically derived secure cluster-based data aggregation in wireless sensor networks ’Department of Computer Science and Engineering, Manonmaniam Sundaranar University, Tirunelveli, India, 2014.E-mail: lathiesbhasker@gmail.com

[2] Mehrjoo, S., Aghaee, H., Karimi, H.: ‘A novel hybrid GA–ABC based energy efficient clustering in wireless sensor network’, Can. J. Multimedia Wirel. Netw., 2011, 2, (2), pp. 41–45

[3] He, W., Liu, X., Nguyen, H., Nahrstedt, K., Abdelzaher, T.: ‘PDA: privacy-preserving data aggregation in wireless sensor networks’.IEEE ICWMC, 2010.

[4] HevinRajesh, D., Paramasivan, B.: ‘Fuzzy based secure data aggregation technique in wireless sensor networks’, J. Comput. Sci., 2012, 8, (6), pp. 899–907.

[5] Thakkar, H., Mishra, S., Chakrabarty, A.: ‘A power efficient cluster-based data aggregation protocol for WSN (MHML)’,Int. J. Eng. Innov. Technol. (IJEIT), 2012, 1, (4), pp. 241–246

[6] Zahmatkesh, A., Yaghmaee, M.H.: ‘A genetic algorithm-based approach for energy- efficient clustering of wireless sensor networks’. Int. Conf. Network Communication and Computer (ICNCC), 2011.

[7] Patil, N.S., Patil, P.R.: ‘Data aggregation in wireless sensor network’. IEEE Int. Conf. Computational Intelligence and Computing Research, 2010.

[8] Roy, S., Conti, M., Setia, S., Jajodia, S.: ‘Secure data aggregation in wireless sensor networks’, IEEE Trans. Inf. Forensics Sec., 2012, 7, (3)

[9] NS is available at http://www.isi.edu/nsnam/ns

[10] Spector, A. Z. 1989. Achieving application requirements. In Distributed Systems, S. Mullender