An Energy Efficient Routing and Fault Tolerant Data Aggregation (EERFTDA) algorithm for wireless sensor networks

V.K. Subhashree * and C. Tharini

Department of Electronics and Communication Engineering, B.S. Abdur Rahman University, Vandalur, Chennai, Tamilnadu, India
E-mails: subhashree_ece_phd_2013@bsauniv.ac.in, tharini@bsauniv.ac.in

Abstract. Major issues found in large scale wireless sensor networks are energy consumption of sensor nodes, redundant data transmission by neighbouring nodes, outlier generation due to unwanted event sensing and node failure. To address these issues, a novel Energy Efficient Routing and Fault Tolerant Data Aggregation (EERFTDA) algorithm is proposed. The key aspects of the proposed algorithm are reliable data aggregation and transmission by eliminating outlier and redundant data using locality sensitive hashing. Another advantage of the proposed algorithm is the bottleneck problem near sink nodes is overcome by the election of relay nodes based on residual energy. The performance of proposed algorithm is analyzed and compared with Data Routing for In-Network Aggregation (DRINA) algorithm for QoS parameters like delay, packet delivery ratio, energy efficiency and aggregation accuracy. The results indicate that EERFTDA provides improved aggregation accuracy and energy efficiency.

Keywords: Routing, data aggregation, locality sensitive hashing, outliers, fault tolerance, energy efficiency

1. Introduction

A large set of autonomous low-cost, low-power and multifunctional wireless sensor nodes form a self-organized distribution system (without any complex structure) referred as Wireless Sensor Networks (WSNs). These sensor nodes coordinate, monitor, sense and collect environmental conditions automatically. They have a wide range of applications in the areas such as environmental monitoring, domestics, meteorology, agriculture, space, health, and military field surveillance [11,12,24,25]. Despite its wide range of applications, there are many challenges in wireless sensor networks such as high energy consumption, scheduling, target localization, fault diagnosis, security, etc. Energy efficiency is considered to be the most important issue among them. The sensor nodes in WSN spend maximum time in sleep mode in order to save power and increase its lifetime. The limited energy resources to the sensor node make them an ideal target for battery drain attacks. Hence, energy is considered to be an asset and it need to be constantly measured and conserved. Routing, data aggregation, network coding, sleep scheduling, green energy-aware security (GEAS), etc., are prominent techniques for improving energy efficiency in WSN [16]. The hierarchical routing is the most commonly used routing technique. Based on the structure, this routing is classified into tree-based and cluster-based routing. The cluster-based routing uses the data aggregation protocols and is prominently used for energy conservation algorithms.

Data aggregation is the process of gathering and aggregating data from multiple sensors to reduce the number of transmissions. It involves fusion of data from multiple sensors at intermediate nodes and transmission of aggregated data to the sink node. The data aggregation is done at the cluster heads and/or the relay nodes based on the

---

*Corresponding author. E-mail: subhashree_ece_phd_2013@bsauniv.ac.in.

application of the network. The commonly used data aggregation mechanisms are SUM, MAX, MIN, COUNT, AVG, etc [4]. Some of the other data aggregation mechanisms include detection and elimination of outlier and redundant data [19]. The data aggregation protocols are classified based on structure and structure-free networks. In structured-based data aggregation protocols, a tree or cluster-based structure is built during network initialization phase for efficient data gathering. These structures perform well in stable environment. However, there are certain drawbacks in structured aggregation techniques such as: (i) Nodes may move or fail unpredictably in a more practical environment; (ii) The construction and maintenance overhead of structure may offset the data aggregation benefits for dynamic situations; (iii) waiting small period at intermediate nodes, causes smaller number of aggregations and a long period results in higher latency. Comparatively, structure free aggregation is preferred, which rarely spends energy on building any structure. The advantages of structure free solutions are controlled average delay, limited maintenance overhead, and better robustness even if there are node failures. The structure free solutions gain better performance during low aggregation efficiency or high network density condition [6]. The challenges in structure free aggregation are: Routing decisions for efficient packet aggregation should be in progress due to lack of pre-constructed structure; Nodes are unaware of their upstream nodes, hence could not openly wait on data from any node without forwarding its own data [13,22]. The proposed EERFTDA algorithm combines the advantages of both structured and structure free data aggregation.

In addition to the need for energy efficient data aggregation, most of the applications in WSN require real-time data mining. Real time data mining is limited by the presence of Outliers and redundant data. Identifying the outliers and providing fault tolerance in the network are considered to be important issues in wireless sensor networks. Outliers are the measurements that deviate from the normal pattern of data sensed by the nodes in a network [5]. The major sources of outliers are compromised sensor nodes, exhausted batteries, imperfect sensing devices, harsh environments, unwanted event sensing, etc., [3,19]. Some of the prominent outlier detection techniques are based on statistical, clustering, classification, spectral decomposition, near neighbour, etc. The most commonly used outlier detection techniques are Kernel density estimation, Locality sensitive hashing, Rough set theory, Support vector machine, etc. [19]. In the proposed algorithm, Locality sensitive hashing is used for detecting the outlier data and redundant data.

The proposed algorithm EERFTDA works in four phases during which the nodes temporarily form a cluster, on sensing an event. Data aggregation is performed by cluster head by eliminating the outlier and redundant data and the data is transmitted to the sink node. The performance of the algorithm is analyzed and compared with Data Routing for In-Network Aggregation (DRINA) algorithm for QoS parameters like delay, packet delivery ratio, energy efficiency and aggregation accuracy. Section 2 summarizes the related work; Section 3 explains the proposed algorithm; Section 4 describes the performance evaluation; and Section 5 concludes the paper.

2. Related work

In this section, prominent research work focussing on data aggregation is presented. In [21], a novel Data Routing for In-Network Aggregation, (DRINA) which involves limited number of messages to set up a routing tree, higher number of overlapping routes, high aggregation rate, and reliable data aggregation and transmission is proposed. The aggregation quality provided by the DRINA routing tree is best when compared to other algorithms. However, DRINA algorithm requires, on average, 25 percent more control messages than the Shortest Path Tree (SPT) algorithm.

In [9], the authors have proposed techniques for data aggregation without explicit structures. Two corresponding mechanisms namely, Data-Aware Anycast at the MAC layer for spatial convergence and Randomized Waiting at the application layer for temporal convergence are modelled. On combining these mechanisms together, there are more contention losses. In [23], an efficient structure-free algorithm is proposed accompanied with a real-time routing, RAG to save energy by data aggregation. Two mechanisms were utilized for temporal and spatial convergence of packets as follows: A Judiciously Waiting policy and A Real-time Data-aware Anycasting policy at the MAC layer processing the routing decisions for efficient aggregation of data as there is no pre-constructed structure. However, the miss ratio increases on increasing the data rate.

In [22], the authors have designed a novel structure-free Real-time Data Aggregation protocol, RDAG, using a Real-time Data-aware Routing policy and a Judiciously Waiting policy for spatial and temporal convergence of packets. Energy is conserved in this protocol by considering spatial and temporal convergence of packets without incurring the overhead of constructing a structure while meeting the desired level of timeliness. But when traffic increases, the miss ratio and contention increases; resulting in longer queuing delays and congestion at intermediate nodes. In [6], a structure-free and energy-balanced (SFEB) data aggregation protocol is proposed, featuring both efficient data gathering and balanced energy consumption, due to its two-phase aggregation process and the dynamic aggregator selection mechanism. The data aggregation was scheduled among the aggregator neighbours and energy balance mechanism was suggested; though there is delay and additional overhead in SFEB-CL.

In [19], the authors presented a fault-tolerant data aggregation scheme to avoid the false data sent by malfunctioning and/or compromised sensor nodes. A locality sensitive hashing was used for detecting outlier data. In [8], a completely structure free aggregation system is proposed. Fuzzy reasoning was employed to make aggregation decisions relying on a flexible and extensible set of application. In [15], a survey on energy-aware security mechanisms for WSN is presented. The sensor nodes spend maximum time in sleep mode in order to conserve energy. The limited energy nodes become an ideal target for the malicious nodes, as it would be forced to get out of the sleep mode and the lifetime of the node is reduced by the sleep-deprivation attack or denial-of-sleep. Hence the energy is considered to be an asset and is constantly measured. This leads to prevention of attacks aiming at depleting power of nodes and the energetic behaviour of the node acts as an indicator to monitor such malicious activities in the network.

In [1], energy aware routing (ERA) protocol for cluster based WSN is proposed. The data routing is facilitated by constructing a directed virtual backbone of cluster heads (CHs) routed to the sink. It is inferred that the energy consumption of the CHs remains balanced during the routing process. The simulations are performed for both random and grid node deployments and the lifetime of the network are maximized. In [2], the authors have proposed an energy efficient Distributed fault tolerant clustering and routing (DFCR) protocol. The cluster formation is based on the residual energy and the distance between the sink and the cluster head. However, for fault tolerance, only permanent failure of the cluster head is considered.

Most of the algorithms proposed in the literature focus on energy efficient algorithms to increase the lifetime of WSN. However, there is a need to focus on fault tolerant data aggregation along with energy efficient design. Hence the proposed algorithm focuses on energy efficient fault tolerant design by using clustering and data aggregation techniques.

3. Energy Efficient Routing and Fault Tolerant Data Aggregation (EERFTDA)

The key aspect of the proposed algorithm is to perform reliable data aggregation and transmission by eliminating the outlier and redundant data. The EERFTDA algorithm considers the following functions of the nodes during the cluster formation and route establishment as shown in Fig. 1.

- Sensor Node (SN): This node detects the event and reports the collected data to the cluster head.
- Cluster-Head (CH): This node performs data aggregation by eliminating the outlier and redundant data using the locality sensitive hashing technique.
- Secondary Cluster-Head (SCH): This node supports the primary head in case of network congestion or failure.
- Sink: It receives all the data sent from a set of cluster-head and relay nodes.
- Relay: This node helps to forward data towards the sink.

The DRINA algorithm includes three phases namely, building hop tree, cluster formation and route establishment and maintenance to implement data transfer in large scale WSNs. In this paper, the proposed algorithm considers election of secondary cluster head election apart from the cluster head election; data aggregation using outlier and redundant data elimination; election of relay nodes based on residual energy of neighbouring nodes, due to which the proposed EERFTDA algorithm is efficient. The proposed algorithm EERFTDA works in four phases.
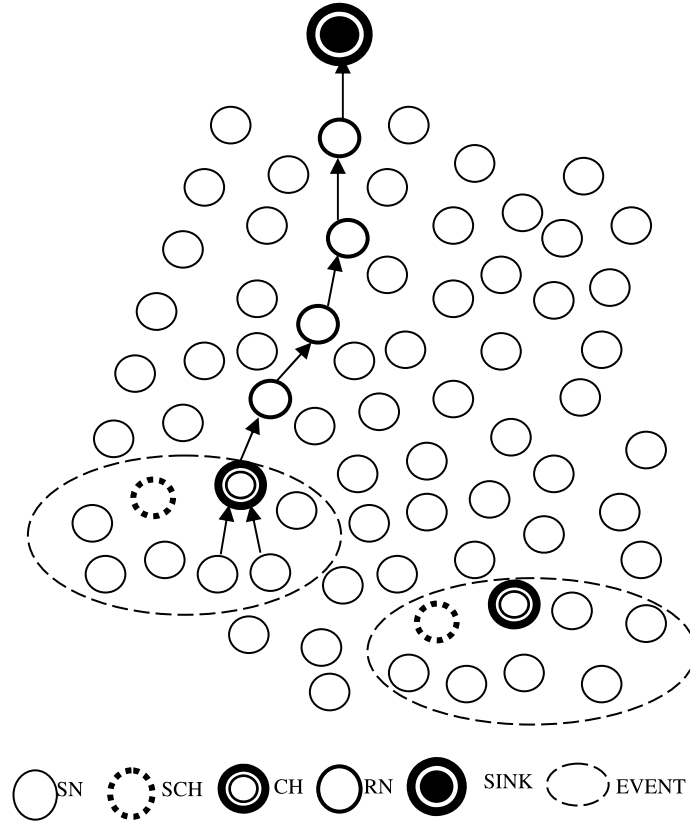
Fig. 1. Cluster formation and route establishment in EERFTDA.

In Phase 1, the hop tree is built. In Phase 2, clusters are formed and the cluster heads are elected. In Phase 3, the Data aggregation is using the Locality Sensitive Hashing technique. In Phase 4, relay nodes are elected based on residual energy and routing is done through the energy efficient paths.

3.1. Phase1: Building hop tree

In this phase, the sink node builds the hop tree for data forwarding purposes. The distance from each node to sink is computed by hops. The sink node uses the flooding technique for route discovery and sends Configuration Message (CM) to the all the nodes in the network. The CM contains two fields namely ID and NH. The node identifier (ID) indicates the node that started or retransmitted the CM and the number of hops (NH) is the distance from the sink in hops.

Initially, all the nodes set the value of their NH field to infinity and sink node sets its NH value as 1. The sink forwards the Configuration message (CM) with its own ID and NH to its neighbouring nodes by the process of flooding. Once the node I receives the CM, from the sink or its neighbouring node, it will check for the NH value of the received packet with its own NH value. If the NH value is found to lesser than its own value, then it will increment the NH value in the received CM with 1 and update this value in its field. Node I would update the CM packet with its own ID and NH value and forwards the CM packet to its neighbours.

If node u receives the CM packet from two of its neighbour nodes 1 and 2, it would first check the NH values of the received packets with its own NH value. If the received NH is lesser, then it would check the NH value between the nodes 1 and 2. If the node 1 NH value is found to be lesser than node 2 NH value, then it would discard the

Algorithm 1: Building hop tree

// Sink node broadcasts CM with NH = 1

// S is the set of nodes that receive the CM from sink.

    If node I ε S,

        If NH (Node I) > NH (CM)

        NH (node I) ← NH (CM) + 1

// node I updates the CM with its ID & NH

        ID (CM) ← ID (node I)

        NH (CM) ← NH (node I)

// node I broadcasts CM with new value

// If node u receives CM from two nodes 1 and 2

        Check NH (node1) > NH (node 2)

          Discard maximum value of NH

Fig. 2. Building hop tree.

CM from node 2 and accepts the CM from node 1. These steps are repeated until the whole network is configured. The algorithm 1 describes the process of building the hop tree as shown in Fig. 2.

All the nodes in the network store the smallest value of NH, which is the smallest distance to the sink. Since the network is not partitioned completely into clusters, before the first event takes place, there is no pre-defined route established between the sink and the nodes. Once the first event takes place, the cluster is formed (discussed in Section 3.2) among those nodes detecting the event and a new route is established to the sink node via the relay nodes.

3.2. Phase 2: Formation of cluster

In EERFTDA, the network is not partitioned initially as done in cluster based approaches. The clusters are formed between nodes that detect the same event. This reduces the energy consumed for partitioning the whole network into clusters. The nodes are considered homogeneous; hence all the nodes will be detecting the same event at different time instances. Let the event refer to sensing of temperature. If a set of nodes detect the event at the same time then, they can be grouped as a cluster. All the nodes in the formed cluster runs for leader election algorithm as described in Fig. 3. A cluster-head (CH) is elected among them on the basis of NH value and secondary cluster-head (SCH) is elected on the basis of residual energy. The role of the SCH is to self – initiate in case of any load congestion or failure of primary CH.

The CH is elected on the basis of number of hops to the sink and if two nodes are having same NH value, then the node having highest residual energy will be elected as CH and the other node will be elected as SCH. If only one node is having minimum NH and all the other nodes are having same NH, the node with the minimum NH

Algorithm 2: Election of CH and SCH

// V be the set of nodes that detect the event

// u be CH and w be SCH

If NH (u) < all other nodes ε V

U ← CH

Re (w)> all other nodes except CH

W ← SCH

Else if NH (u) =NH (w)

If Re (u)> Re (w)

U ← CH

W ← SCH

Fig. 3. Election of CH and SCH.

will be elected as CH and the node having highest residual energy among other nodes will be elected as SCH. The SCH also senses the event and sends the data to the CH. All the nodes in the cluster would send the data to the CH and sets time out to receive an acknowledgement. If the acknowledgment from the CH is received, then the nodes understand that the CH is alive; else the elected SCH realizes that the CH is either in traffic due to congestion or failed due to loss of energy and self-initiates itself as the CH. The algorithm 2 describes the election of primary and secondary cluster-heads as shown in Fig. 3.

3.3. Phase 3: Data aggregation at cluster-head using locality sensitive hashing

Data aggregation in EERFTDA is done at the CH or SCH (in case of CH failure) by eliminating the outliers and redundant data using the Locality Sensitive Hashing (LSH) technique. The nodes send their LSH codes to the cluster head. The CH performs LSH algorithm and eliminates outlier generated due to sensor failure, unwanted event sensing, etc. Thus the elimination of outlier data and redundant data is one form of data aggregation due to which reduced number of transmissions from cluster head to sink node is achieved [14].

LSH belongs to the family of min-hash functions that is used to find the similar entries in large data sets by finding out the approximate near matches efficiently [19]. LSH technique uses the distance metrics to find the similarity between data. In EERFTDA, the cosine distance and hamming distance are used for finding the similarity and detecting the outlier data. The output of the LSH algorithm determines whether there is a K-near neighbour for a data point or not. To make use of LSH technique on sensor node data sets, data sets are converted in the form of vectors [19].

Random Hyper Plane (RHP) projection is one form of LSH algorithm that is used for finding the similarity between the vector data sets. RHP generates a family of hash functions by projecting the collected data in a d-dimensional space to a Hamming cube, whose coordinates are in the form of 0s and 1s bits. Let x and y be the data sets in the vector form. RHP generates a family of hash function by taking dot product with the data vector 'x' and

| Dataset x | 33.32 | 33.65 | 34.03 | 33.98 | 33.95 |

| Dataset y | 37.72 | 37.98 | 38.56 | 38.98 | 38.32 |

| Random vector 'f' | 0.71 | 0.78 | -0.84 | 0.64 | -0.97 |

Fig. 4. Temperature data sets x and y.

random vector 'f'. A common seed is used to produce a spherically symmetric random vector 'f' of unit length from a d-dimensional space. A hash function as h_f is defined in equation (1) as:

$$h_f(x) = \begin{cases} 1, & \text{if } f \times x \geqslant 0 \\ 0, & \text{if } f \times x < 0 \end{cases} \tag{1}$$

Using n random vectors $f_1, f_2, \ldots, f_n$, input data items x and y are mapped into RHP bitmaps LSH(x) and LSH(y) of length 'n' respectively. The probability of existence of similarity between x and y can be found using the cosine distance between the data vectors x and y as shown in equation (2):

$$(x, y) = \arccos\left(\frac{x \cdot y}{\|x\| \cdot \|y\|}\right) \tag{2}$$

$$P_r\left[h_f(x) = h_f(y)\right] = 1 - \frac{\emptyset(x, y)}{\pi} \tag{3}$$

where $\phi(x, y)$ is the angle between the data sets x and y, P_r is the probability for existence of similarity between the two data sets x and y [as in equation (3)]. Moreover, this hash function measures the similarity only in terms of angle between two vectors. But calculation of an angle between two vectors is not an easy task for resource constrained sensor node. Hence further improvement in the technique can be done in equation (3) based on hamming distance:

$$D_h(\text{LSH}_x, \text{LSH}_y) = nx\frac{\emptyset(x, y)}{\pi} \tag{4}$$

$$\emptyset(x, y) = \pi x\frac{D_h(\text{LSH}_x, \text{LSH}_y)}{n} \tag{5}$$

where $\text{LSH}_x, \text{LSH}_y \in [0, 1]^n$ represents LSH codes of vectors x and y respectively. $D_h(\text{LSH}_x, \text{LSH}_y)$ represents the hamming distance between the LSH_x and LSH_y. Here each LSH code is length of n-bit, which has much smaller value than original vectors. The equation (4) helps sensor node to measure the relationship between their data sets with the help of bit comparison.

To explain LSH using RHP, similarity calculation between two temperature data sets x and y are considered. The RHP bitmaps are formed by taking dot product with the random vector f using the equation (1). The temperature data sets x and y along with the random vector is shown in the Fig. 4.

The data sets x and y are mapped into RHP bitmaps by taking dot product with the random vector and the resulting vector is in the form of 0s and 1s. The RHP bitmaps of the data sets x and y are termed as LSH_x and LSH_y respectively. The Fig. 5 illustrates the angle estimation of two data sets x and y using their RHP bitmaps.
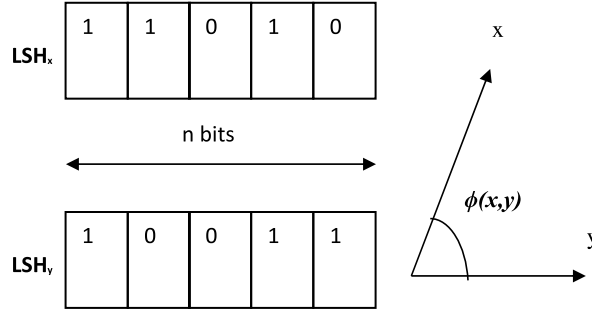
Fig. 5. Angle estimation using RHP bitmaps.

The Hamming distance is the number of bits that differ in the RHP encodings of the data sets x and y (LSH_x and LSH_y) and is given in equation (6):

$$D_h(LSH_x, LSH_y) = 2 \tag{6}$$

Using the equation (5), the angle between the data sets is estimated using the Hamming distance and given by the equation (7):

$$\emptyset(x, y) = \pi x \frac{2}{5} \tag{7}$$

where n is the length of LSH code and the value is 5 in this example. The cosine distance between the data sets x and y is given by the equation (8):

$$\cos \emptyset(x, y) = \cos\left(\frac{2\pi}{5}\right) = 0.3090 \tag{8}$$

A similarity threshold is set by the cluster head and the cosine distance is compared with the threshold. The similarity threshold is set by each cluster head based on the minimum support count assigned by the base station. If the cosine distance is found to be lesser then the support count, then the data from the node is marked as local outlier. The local outliers are those affected by the events in the neighbouring clusters. The local outlier of one cluster would be a normal data of the other cluster. Hence, the CHs of the neighbouring clusters exchange their local outliers. If the local outlier of one cluster does not have minimum support count of its neighbouring cluster, then the local outlier is considered to be global outlier and it is eliminated [19,20].

The data aggregation is done by the cluster head by eliminating the outlier data and the redundant data. The Cluster head eliminates the global outliers based on the threshold value. In order to eliminate redundant data, the CH checks if two or more nodes are having the same LSH code. If it is found to be the same, the CH will ask for data from only one node, thus eliminating the redundant data. In this phase, the data aggregation is performed by merging only the distinct data that lies within the desired threshold set by the CH after eliminating the redundant and outlier data for each occurrence of event. Figure 6 shows the data aggregation using LSH.

3.4. Phase 4: Relay node election and routing

In EERFTDA, the aggregated data from the CH is forwarded via relay nodes to the sink. Since the network is structure-less, there cannot be a predefined path from each node to the sink. Once the event is detected by few nodes and the data aggregation process is completed by the CH/SCH, it starts establishing new route to the sink for event dissemination.

CH sends a route request message requesting node ID and residual energy to its neighbour nodes. The neighbour nodes send route responses with their node ID and residual energy to the CH. The CH compares the residual

---

Algorithm 3: Data aggregation using LSH

// for data aggregation in CH do

//each node collects data set Ds and generates its

own LSH codes

//CH collects LSH codes of all the nodes in its

cluster

//Similarity threshold (ST) is set by CH

    If LSH of nodes lies inside ST

        If two nodes have same LSH codes

            Accept data from one node

            Eliminate data from other node

        If LSH codes of other nodes are distinct

            Accept data from these nodes

    Else

        Eliminate node as an outlier

Fig. 6. Data aggregation using LSH at CH.

energies of the neighbouring nodes and finds the nodes with highest residual energy. The residual energy is the remaining energy present in a node once the network operation is initiated. If there are two or more nodes having equivalent residual energies, then in order to choose the next hop node, the MAC layer of CH uses the anycast mechanism [23]. Anycasting is applied for choosing the best next hop to forward the packets, when multiple next hop choices are available.

The RTS/CTS technique is used by the MAC layer to determine the eligible next hop relay node. The RTS packet of the node contains the type of data sensed and Time-to-Deadline (TTD) of the transmitting packet. The packet's Time-to-Deadline (TTD) is used to indicate the remaining time of the packet before its deadline and is used in routing decisions [23]. To make real-time decisions, before the CH forwards the RTS packets to its neighbour nodes, it computes the required velocity for the packet. The required velocity is the ratio of the distance between the sender and the receiver node to its TTD value. The CH sends the RTS packets to its neighbouring nodes having more residual energy, by setting a specific threshold value for TTD [7].

Once the RTS is received by these neighbouring nodes, they in turn send CTS responses to the CH. The CTS contains the neighbour node's TTD value and type of data sensed. Since the network is homogenous, the type of data sensed would be common for all the nodes. The CTS responses belong to any of the two classes described below:

- Class 1: the minimum TTD of packets in the aggregation queue is lesser than or equal to the TTD value specified in RTS packet of the forwarder node.
- Class 2: the minimum TTD of packets in the aggregation queue is greater than the TTD value specified in RTS packet of the forwarder node.

The nodes that lie in Class 1 will be an efficient relay node since its TTD value is lesser than the forwarder node. This makes the aggregation queue to be less or empty. Hence highest priority is given to the nodes in Class 1. If two

---

Algorithm 4: Election of Relay node

//Send RTS to neighbour nodes set N_n

    If node I ε N_n

      If TTD (I) < TTD (CH)

        Node I $\leftarrow$ class 1

        Select node I as next hop relay node

      // if both node I and J are eligible

      if TTD (J) < TTD (I) < TTD (CH)

        Node I $\leftarrow$ class 1 (high priority)

        Select node I as next hop relay node

    Else

      Reject node I

Fig. 7. Election of relay node.

or more nodes in the class 1 are eligible next hop nodes, the node with a TTD value that is lesser and very much closer to the RTS specified TTD value of the CH is selected as the next hop relay node, as shown in algorithm 4 as shown in Fig. 7.

According to the bottleneck sphere theorem [23], the nodes closer to the sink have higher energy consumption. When the eligible next hop relay nodes near the sink gets failed due to energy depletion, the data will not reach the sink. Using the residual energy of the nodes near the sink, the routing decisions are taken on how to forward the packet to the sink. Once the route is established between the CH and the sink through relay nodes, the data is forwarded to the sink.

The relay nodes can act as an aggregator node if it receives data from two or more CHs in the network. Since the data is of same type, the data aggregation can be done at these nodes. Data aware anycast algorithm is the data aggregation approach used by the relay nodes for performing data aggregation when two or more events use the same relay node for routing [10,17,23] in order to improve the data aggregation efficiency.

4. Results

In this section, the performance of the proposed algorithm EERFTDA is analyzed and compared with the DRINA technique [21]. This analysis and comparison is done using metrics such as transmission delay, packet delivery ratio, energy consumption and aggregation accuracy, using NS2 simulator [18]. The default simulation parameters are presented in Table 1. For each simulation set, the parameters such as transmission rate, number of nodes, number of events and event duration are varied and their impact on EERFTDA and DRINA algorithms are evaluated.

The following metrics were used for the performance evaluation:

- **Delay**: It is the amount of time taken by the nodes to transmit the data packets.

Table 1

Simulation parameters

Parameters	Description
No. of Nodes	100
Area Size	500×500
MAC	IEEE 802.11
Transmission Range	250 m
Simulation Time	50 sec
Traffic Source	CBR
Packet Size	1000
Initial energy	15.1 J
Transmission Power	0.660
Receiving Power	0.395

- **Packet Delivery Ratio**: It is the ratio of the number of packets received to the number of packets sent.
- **Energy Consumption**: It is the amount of energy consumed by the nodes to transmit the data packets to the receiver.
- **Aggregation Accuracy**: It is percentage of true positive data computed by the data aggregator after eliminating the redundant and outlier data [19].

4.1. Impact on transmission rate

The performance of EERFTDA is analysed by varying the transmission rate in kbps and the results are compared with DRINA for the metrics such as delay, packet delivery ratio and energy consumption. The transmission rate is varied as 100, 200, 300, 400 and 500 kbps for 1000 packets and the performance is analysed.

In DRINA, the routing is done through the shortest path irrespective of the number of packets in the aggregation queue and their Time to Deadline (TTD) value. As the transmission rate increases, the contention in the queue increases, as a result the delay increases rapidly. This problem is overcome in EERFTDA by choosing the relay nodes based on the TTD value and number of packets in the aggregation queue. The node having lesser TTD value than the forwarder node is given high priority since the packets would reach the sink faster. The delay is 40% lesser in EERFTDA when compared to DRINA, as shown in Fig. 8(a).

As the transmission rate increases, there is an increase in the contention and queuing of packets in the relay nodes. In DRINA, the relay nodes are chosen based on the distance between the nodes. Hence the node has to forward the packets despite the contention in its aggregation queue leading to more packet drops. In EERFTDA, the relay nodes are chosen based on the number of packets in the queue and TTD value. Data aggregation is performed at the relay nodes using data aware anycast mechanism [23], thus reducing the number of packets to be transmitted. From the Fig. 8(b), the packet delivery ratio of DRINA is 20% lesser than EERFTDA.

The increase in the transmission rate increases the traffic in the network resulting in congestion in intermediate nodes. In DRINA, since shortest path routing is done, there are more chances for the nodes to lose energy in high traffic loads. To avoid this problem, the relays nodes are chosen based on residual energy in EERFTDA. The node having more residual energy and less TTD value is the eligible next hop relay node, thus reducing more energy consumption on transmission of data. From the Fig. 8(c), the energy consumption is 30% less in EERFTDA when compared to DRINA.

4.2. Impact of network size

The performance of EERFTDA is analysed by varying the network size, i.e. by varying the number of nodes in the network as 20, 40, 60, 80 and 100. The performance is compared with DRINA for the metrics such as delay, packet delivery ratio, and energy consumption.

In DRINA, as the number of nodes in the network increases, the network becomes denser and the delay rapidly increases. In EERFTDA, though the number of nodes increases, the delay is 15% less when compared to DRINA

Transmission Rate vs Delay

(a)

Transmission Rate vs Packet Delivery ratio

(b)
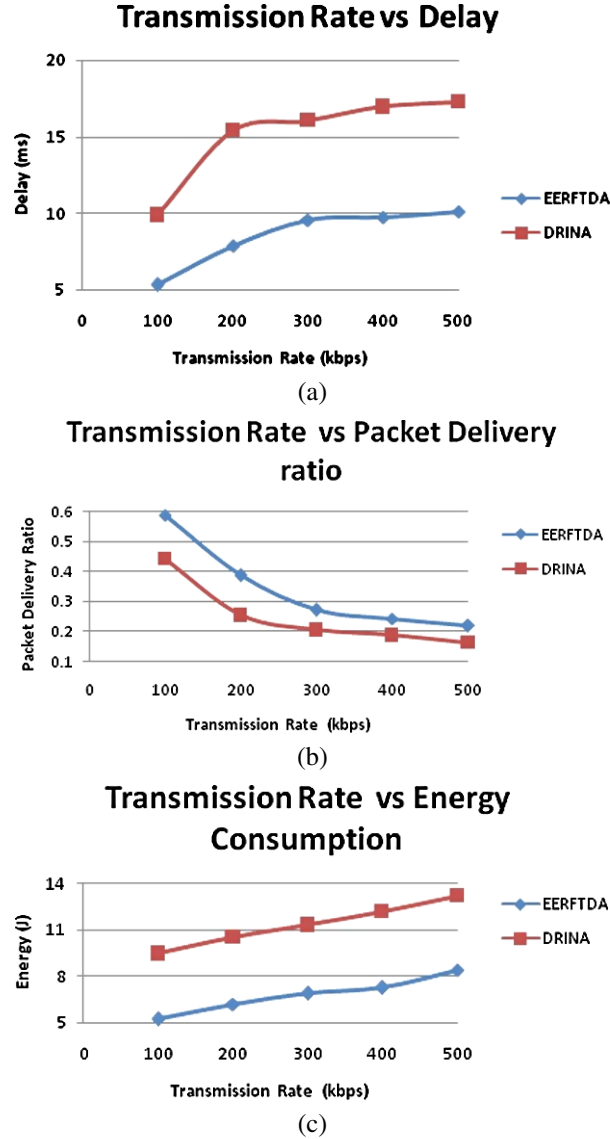
Transmission Rate vs Energy Consumption

(c)

Fig. 8. (a) Transmission rate vs. delay. (b) Transmission rate vs. packet delivery ratio. (c) Transmission rate vs. energy consumption.

as shown in Fig. 9(a). This is due to the election of secondary cluster head that is used to help their primary head during network congestion. Since the load is balanced between them, the data forwarding is done through many relay nodes, so that the delay is not increased to a greater extent when compared to DRINA.

As the nodes in the network increase, there are more chances for simultaneous events to occur, as a result, more clusters will be formed. The cluster heads forwards the data to the sink node via the relay nodes. The data aggregation process takes place in relay nodes based on data aware anycast. Hence one relay node may aggregate data from two or more nodes and as a result, there might be more contention in the network. This leads to decrease in packet delivery ratio. But in DRINA, since this mechanism is not used, as the network size increases, the packets tend to drop more at relay nodes and as a result, the packet delivery ratio is 25% less when compared to EERFTDA, as shown in Fig. 9(b).
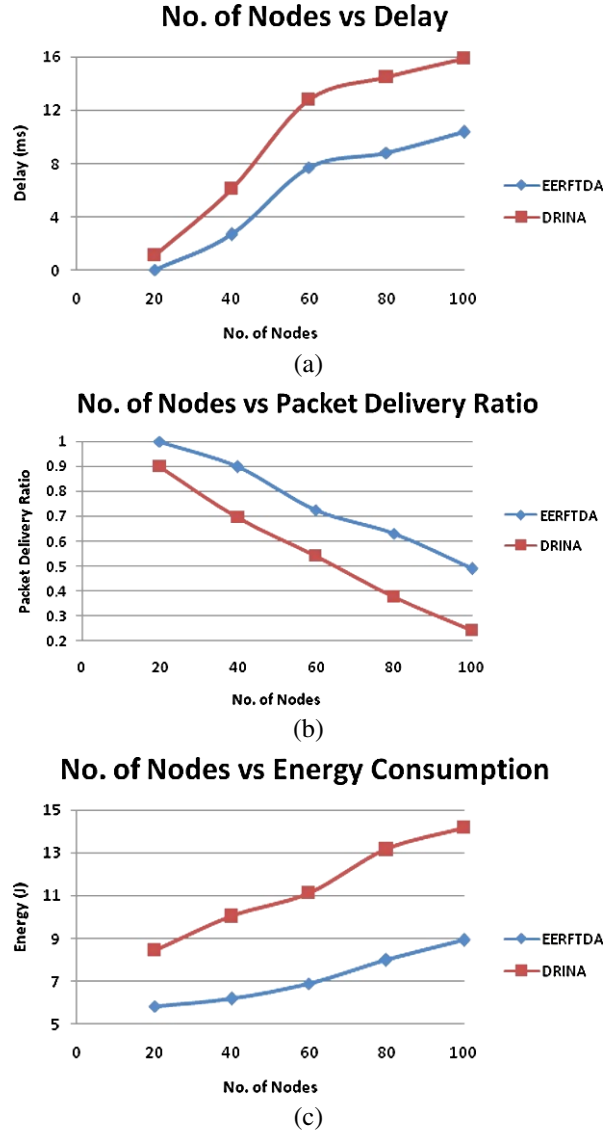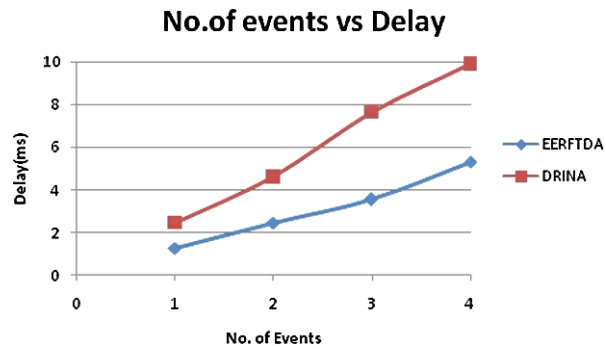
Fig. 9. (a) Network size vs. delay. (b) Network size vs. packet delivery ratio. (c) Network size vs. energy consumption.
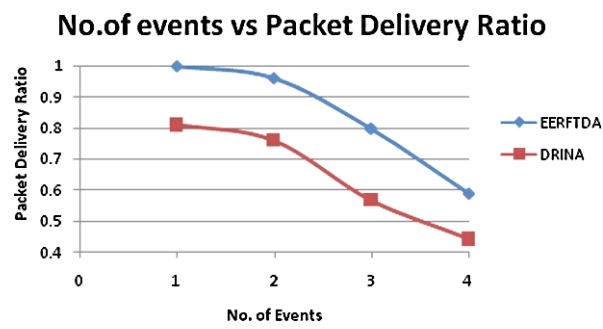
In EERFTDA, the secondary cluster heads and the relay nodes are elected based on residual energy. As the number of nodes increases, the network becomes denser. According to the bottleneck problem [23], the nodes near the sink drain out soon than the other nodes in the network. In DRINA, since shortest path routing is performed, the nodes that is very closer to the sink node drains out of energy and tends to fail. The load gets equally distributed near the sink node and the failure of nodes due to energy depletion is reduced. As shown in Fig. 9(c), the energy consumption is reduced to 40% in EERFTDA when compared to DRINA.
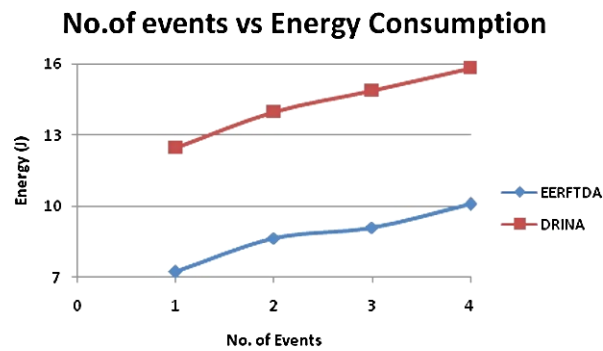
4.3. Impact of number of events

The number of events is varied to evaluate the behaviour of EERFTDA in networks with 1, 2, 3 and 4 events occurring simultaneously and the performance was compared with DRINA for the metrics such as delay, packet delivery ratio, energy consumption and aggregation accuracy.
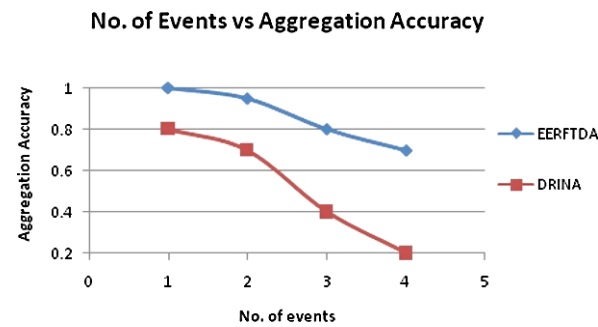
No.of events vs Delay

(a)

No.of events vs Packet Delivery Ratio

(b)

No.of events vs Energy Consumption

(c)

No. of Events vs Aggregation Accuracy

(d)

Fig. 10. (a) Number of events vs. delay. (b) Number of events vs. packet delivery ratio. (c) Number of events vs. energy consumption. (d) Number of events vs. aggregation accuracy.
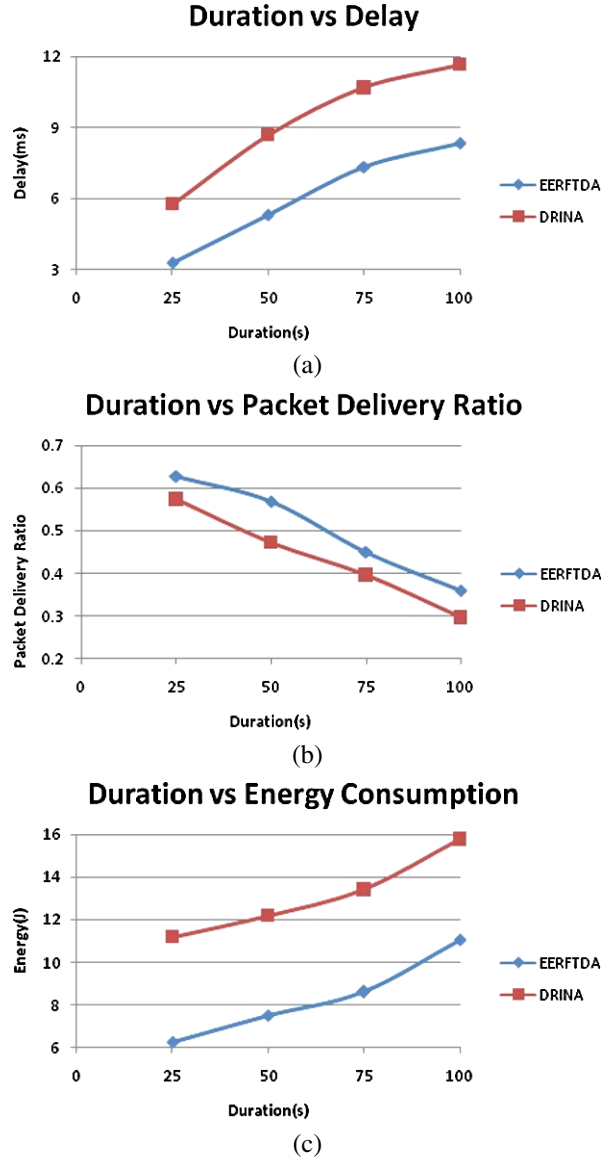
Fig. 11. (a) Duration vs. delay. (b) Duration vs. packet delivery ratio. (c) Duration vs. energy consumption.

Figure 10(a), (b), and (c) show the impact of number of events for the metrics such as delay, packet delivery ratio, and energy consumption. The clusters are formed when a few nodes detect the same event. In DRINA, aggregation is done by receiving few data packets and sending a fixed size of merged packets [21]. If the aggregation queue is having more packets, the delay increases with a decrease in delivery ratio. In EERFTDA, the delay is less due to the presence of extra aggregators as it shares the load of the primary aggregators. As the relay nodes perform data aggregation, the data from various cluster heads are getting aggregated in the relay nodes, as a result, the number of packets transmitted is also reduced. This decreases the packet drops and energy consumption. The energy consumption is 30% lesser when compared to DRINA. As the number of simultaneous events increases, there are more chances for generation of redundant and outlier data. Since there is no mechanism involved to eliminate outlier data, the aggregation accuracy is reduced in DRINA. In EERFTDA, the cluster head uses the locality

sensitive hashing technique to eliminate the outlier and redundant data, the aggregation accuracy is increased to 50% when compared to DRINA as shown in Fig. 10(d).

4.4. Impact of event duration

The event duration is changed as 25, 50, 75 and 100 seconds and the performance of EERFDTA is compared with DRINA for the metrics such as delay, packet delivery ratio, energy consumption.

In DRINA, the packet is forwarded hop by hop and as the event duration increses, the delay increases. In EERFTDA, the packet is forwarded through the relay nodes having lesser TTD value. This reduces the delay in EERFTDA to 30% less compared to DRINA as shown in Fig. 11(a).

The packet delivery ratio decreases as the duration of event increases. In EERFTDA, due to the process of data aggregation both at cluster heads and relay nodes, the number of packets transmitted is reduced, as a result, the contention and queuing is reduced; reducing the packet drop. From the Fig. 11(b), the delivery ratio is 20% more in EERFTDA when compared to DRINA.

Since the events occur for more time, the redundant data are generated and are eliminated in EERFTDA; hence the number of packets transmitted is reduced, thereby the energy consumption is 40% lesser when compared to DRINA, thus maximizing the lifetime of the network.

From the results, it is observed that the proposed algorithm provides fault tolerance in the network (i) by eliminating the outlier data; (ii) by choosing an relay node based on residual energy to avoid bottleneck problem and (iii) by electing secondary CH, which will self initiate and perform the operations of CH in case of any failure.

5. Conclusion

An Energy Efficient Routing and fault tolerant data aggregation (EERFTDA) algorithm for WSN was proposed to provide fault tolerant data aggregation by eliminating the outliers and redundant data using locality sensitive hashing. In order to maximize the network lifetime, relay nodes were elected based on residual energy. Since all the nodes are homogenous, the relay nodes also performed data aggregation when it received data from two or more cluster-heads. In addition to election of one cluster-head per cluster, a secondary head was chosen in order to share the network load during congestion and prevent the failure of network. The simulation results were analysed by varying the transmission rate, network size, number of events and events duration for the performance metrics such as delay, delivery ratio, energy consumption and aggregation accuracy. The proposed algorithm outperformed DRINA approach in all the metrics and for different network characteristics. The aggregation accuracy was found to be 50% greater than DRINA and energy consumption is 40% lesser than DRINA. In future work, the proposed algorithm can be used for heterogeneous network and different outlier detection techniques can be analysed for improving the accuracy of data aggregation.

References

[1] T. Amgoth and P.K. Jana, Energy-aware routing algorithm for wireless sensor networks, *Elsevier Computers and Electrical Engineering* (2015), 357–367. doi:10.1016/j.compeleceng.2014.07.010.

[2] Md. Azharuddin, P. Kuila and P.K. Jana, Energy efficient fault tolerant clustering and routing algorithm for wireless sensor networks, *Elsevier Computers and Electrical Engineering* (2015), 177–190. doi:10.1016/j.compeleceng.2014.07.019.

[3] S.A.-H. Baddar, A. Merlo and M. Migliardi, Anomaly detection in computer networks: A state-of-the-art review, *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA)* **5**(4) (2014), 29–64.

[4] V. Bhoopathy and R.M.S. Parvathi, Secure authentication technique for data aggregation in wireless sensor networks, *Journal of Computer Science* (2012), 232–238.

[5] V. Chandola, A. Banerjee and V. Kumar, Outlier detection: A survey, Technical Report, University of Minnesota, 2007.

[6] C.-M. Chaoy and T.-Y. Hsiao, Design of structure-free and energy-balanced data aggregation in wireless sensor networks, in: *11th IEEE International Conference on High Performance Computing and Communications (HPCC '09)*, Seoul, 2009.

[7] O. Chipara, Z. He, G. Xing, Q. Chen, X. Wang, C. Lu, J. Stankovic and T. Abdelzaher, Real-time power-aware routing in sensor networks, in: *Proceedings of the 14th IEEE International Workshop on Quality of Service (IWQoS'06)*, 2006, pp. 83–92. doi:10.1109/IWQOS.2006.250454.

[8] S. Dietzel, B. Bako, E. Schoch and F. Kargl, A fuzzy logic based approach for structure-free aggregation in vehicular ad-hoc networks, in: *Proceedings of the Sixth ACM International Workshop on VehiculAr InterNETworking, VANET '09*, ACM, New York, NY, USA, 2009, pp. 79–88. doi:10.1145/1614269.1614283.

[9] K.-W. Fan, S. Liu and P. Sinha, On the potential of structure-free data aggregation in sensor networks, in: *Proceedings of 25th IEEE International Conference on Computer Communications INFOCOM*, Barcelona, Spain, 2006, pp. 1–12.

[10] K.-W. Fan, S. Liu and P. Sinha, Structure-free data aggregation in sensor networks, *IEEE Transactions on Mobile Computing* (2007), 929–942. doi:10.1109/TMC.2007.1011.

[11] J. Flathagen and F.T. Johnsen, Integrating wireless sensor networks in the NATO network enabled capability using web services, in: *IEEE Military Communications Conference - Track 2 - Network Protocols and Performance*, 2011, pp. 828–833.

[12] A. Leandro, Villas, D. Guidoni, A. Boukerche, R.B. Araujo and A.A.F. Loureiro, Dynamic and scalable routing to perform efficient data aggregation in WSNs, in: *The Proceedings of IEEE International Conference on Communications*, 2011, pp. 1–5.

[13] C. Lu, B.M. Blum, T.F. Abdelzaher, J.A. Stankovic and T. He, RAP: A Real-Time Communication Architecture for Large-Scale Wireless Sensor Networks. Department of Computer Science, University of Virginia.

[14] B.S. Mathapati, S.R. Patil and V.D. Mytri, Energy efficient reliable data aggregation technique for wireless sensor networks, in: *IEEE International Conference on Computing Sciences*, 2012, pp. 153–158.

[15] A. Merlo, M. Migliardi and L. Caviglione, A survey on energy-aware security mechanisms, *Pervasive and Mobile Computing* **24** (2015), 77–90. doi:10.1016/j.pmcj.2015.05.005.

[16] M. Migliardi, A. Merlo and L. Caviglione, A survey of green, energy-aware security and some of its recent developments in networking and mobile computing, in: *2014 Eighth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS)*, Vol. 24, July, 2014, pp. 241–246.

[17] M.Y. Mohamed Yacoab and V. Sundaram, Secure data aggregation using reliable nodes for wireless sensor networks, *Research Journal of Applied Sciences, Engineering and Technology*, 768–775.

[18] NetworkSimulator, http:///www.isi.edu/nsnam/ns.

[19] S. Ozdemir and Y. Xiao, FTDA: Outlier detection-based fault-tolerant data aggregation for wireless sensor networks, *Security and Communication Networks* **6** (2013), 702–710. doi:10.1002/sec.604.

[20] A. Saeed, A. Stranieri and R. Dazeley, Fault-tolerant data aggregation scheme for monitoring of critical events in grid based healthcare sensor networks, in: *Proceedings of the 19th High Performance Computing Symposia*, 2011, pp. 56–64.

[21] L.A. Villas, A. Boukerche, H.S. Ramos, H.A.B.F. de Oliveira, R.B. de Araujo and A.A.F. Loureiro, DRINA: A lightweight and reliable routing approach for in-network aggregation in wireless sensor networks, *IEEE Transactions On Computers* **62**(4) (2013), 676–689. doi:10.1109/TC.2012.31.

[22] M.H. Yeganeh, H. Yousefi, N. Alinaghipour and A. Movaghar, RDAG: A structure-free real-time data aggregation protocol for wireless sensor networks, in: *17th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications*, 2011, pp. 51–60.

[23] H. Yousefi, M.H. Yeganeh, N. Alinaghipour and A. Movaghar, Structure-free real-time data aggregation in wireless sensor networks, *Computer Communications* (2013), 1132–1140.

[24] X. Zhao, S. Kami Makki and N. Pissinou, Determining the routing path for in-network aggregation, in: *International Conference on Hybrid Information Technology (ICHIT '06)*, Vol. 2, 2006.

[25] J. Zhou, Q. Cao, C. Li and R. Huang, A genetic algorithm based on extended sequence and topology encoding for the multicast protocol in two-tiered WSN, *Expert Systems With Applications* (2010), 1684–1695. doi:10.1016/j.eswa.2009.06.043.

Biographies

V.K. Subhashree received the B.E. degree in Electronics and Communication Engineering from Anna University, Chennai, India, in 2007, and the M.E. degree in Communication Systems from the Anna University, Chennai, India, in 2009. She is currently pursuing her research in the field of wireless sensor networks in the ECE department, B S Abdur Rahman University, Chennai. She has presented few papers in national and international conferences. She is a member of IEEE.

C. Tharini is currently working as a Professor in ECE department, B S Abdur Rahman University Chennai. She has pursued her research in Information and Communication at Anna University, Chennai and completed during the year 2011. She has 15 years of experience in research and teaching. She has presented many papers in International Conference and some of her papers were awarded as best paper. She has published papers in International journals and her papers are cited by many researchers in this field. She is a life member of Computer Society of India. She has reviewed papers for the journals like Journal of Computers and e-scripts, a novel crowd- based submission system introduced by European Alliance for Innovation. Currently she is guiding research scholars in the area of Wireless Communication and Sensor Networks. She has also organized many national workshops and conferences in the field of wireless sensor networks. She has been a resource person for many workshops and conferences organized by several engineering colleges.