Egyptian Informatics Journal (2016) 17, 273–284

Cairo University

Egyptian Informatics Journal

www.elsevier.com/locate/eij
www.sciencedirect.com

FULL-LENGTH ARTICLE

Energy efficient structure-free data aggregation and delivery in WSN

Prabhudutta Mohanty *, Manas Ranjan Kabat

Dept. of Comp. Sc. & Engg., Veer Surendra Sai University of Technology, Burla, Sambalpur, Orissa, India

Received 7 May 2015; revised 27 December 2015; accepted 13 January 2016
Available online 4 March 2016

KEYWORDS

Energy efficient;
Structure-free;
Data aggregation;
Redundant data;
WSN

Abstract In Wireless Sensor Networks (WSNs), the energy consumption due to the sensed data transmission is more than processing data locally within the sensor node. The data aggregation is one of the techniques to conserve energy by eliminating the redundant data transmission in dense WSNs. In this paper, we propose an energy efficient structure-free data aggregation and delivery (ESDAD) protocol, which aggregates the redundant data in the intermediate nodes. In the proposed protocol, waiting time for packets at each intermediate node is calculated very sensibly so that data can be aggregated efficiently in the routing path. The sensed data packets are transmitted judicially to the aggregation point for data aggregation. The ESDAD protocol computes a cost function for structure-free, next-hop node selection and performs near source data aggregation. The buffer of each node is partitioned to maintain different types of flows for fair and efficient data delivery. The transmission rates of the sources and intermediate nodes are adjusted during congestion. The performance of the proposed protocol is evaluated through extensive simulations. The simulation results reveal that it outperforms the existing structure-free protocols in terms of energy efficiency, reliability and on-time delivery ratio.

1. Introduction

The WSN is a collection of low-cost, small, energy constrained and unreliable multifunctional microsensor nodes, which cooperatively transmit the sensed data to base station (BS) [1,2]. These sensor nodes are randomly deployed in a remote hostile environment to measure the temperature, light intensity, humidity, noise level or any other physical conditions in its locality. These sensor nodes collect information from their surroundings and transmit them to one or more BS. The data may either be accessed from a remote location through Internet or the actuators directly perform actions in response to the event. WSNs are being widely deployed for different applications such as battlefield surveillance [3], healthcare applications [4], environment and habitat monitoring [5], homeland security [6], biomedical research [7], human imaging [8] and agricultural monitoring [9].

* Corresponding author. Tel.: +91 9937650035.
E-mail addresses: prabhudutta.mohanty@gmail.com (P. Mohanty), manas_kabat@yahoo.com (M.R. Kabat).

Peer review under responsibility of Faculty of Computers and Information, Cairo University.

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.eij.2016.01.002
1110-8665 © 2016 Production and hosting by Elsevier B.V. on behalf of Faculty of Computers and Information, Cairo University.

Each sensor in a WSN has limited energy, memory and computational capacity [10]. Nodes are densely deployed in a WSN, usual sense highly co-related and redundant data. The transmission of all these redundant data to BS increases energy consumption and congestion in the network. The energy consumption due to transmission of all the redundant data can be minimized by collecting these data locally and converting it into a valid valuable data packet before forwarding to the BS. The node that collects the similar data packets and converts it into a single packet is called the aggregator node. The process of converting multiple similar packets into one packet is called data aggregation [11]. The aggregator eliminates redundant data using various methods such as statistical approaches [12,13], probabilistic approaches [14,15] and artificial intelligence [16–18].

Several data aggregation protocols have been developed for data transmission and avoid overwhelming amounts of traffic in the network. These data aggregation protocols are broadly classified into a structure based [19–27] and structure-free [2,28,29] data aggregation. In structured data aggregation, different structures of the sensor nodes are formed to collect data, aggregate the collected data and transmit the aggregated data to the BS. Data aggregation structures are chain-based [19], tree based [20], cluster based [21–25], tree-cluster based [26] or hierarchical cluster based [27]. The intermediate nodes are designated as leader node [19], root [20] and cluster heads [21–27] in the chain, tree and cluster respectively. These intermediate nodes collect data from the sensor nodes and aggregate them. The structured data aggregation incurs low maintenance overhead for static and unchanged traffic pattern. However, in case of event-based applications, the benefits of data aggregation may outweigh the overhead of construction and maintenance of the structure [28]. In distributed event, the absence of an explicit center or any evident point for optimal aggregation makes structured aggregation approaches inapplicable. The structured approach that centrally computes the aggregation tree [30] incurs excessive communication overhead for which it is impractical in dynamic scenarios. Furthermore, the performance of structured data aggregation depends on the waiting period of data in the intermediate nodes after receiving from all downstream nodes. A small waiting period may lead to poor aggregation and a long waiting period may lead to higher latency. Therefore, it is important to compute the optimal waiting period from the relative position of the node with respect to the entire sub-tree for dynamic scenarios.

In structure-free data aggregation approaches [2,28,29], the multiple sources having similar data, select the same downstream node so that the redundant data can be aggregated at that node and the energy spent to build a structure can be saved. However, the energy spent due to the data transmitted by the sensor nodes and data aggregation at the aggregation point cannot be avoided. Thus, the main issue in structure-free aggregation is the routing decisions for efficient aggregation of packets due to the unavailability of a pre-constructed structure. The waiting periods of the packets are required to be computed dynamically at the intermediate nodes for delay efficient data aggregation. In structure-free data aggregation, the number of downstream nodes for receiving data packets is not fixed. Therefore, an efficient buffer management and scheduling scheme are also required to avoid network congestion and increase throughput.

In this paper, we propose an Energy efficient Structure-free Data Aggregation and Delivery (ESDAD) protocol to assure efficient data aggregation and delivery without explicit maintenance of a structure. The ESDAD protocol addresses different levels of sensing reliability required in the sensing field. The data packets are judicially transmitted to the next-hop node on the basis of required reliability for aggregation. The waiting time of packets at each intermediate node is calculated very sensibly so that data can be aggregated efficiently in the path. The proposed protocol performs near source data aggregation and computes a cost function for structure-free next-hop node selection. The buffer of each intermediate node is partitioned to support fair and efficient data delivery with buffer management.

The rest of this paper is organized as follows. Section 2 presents the related work on data aggregation techniques. The proposed structure-free data aggregation, data forwarding, the loss recovery and congestion control mechanism are presented in Section 3. The simulation results of the proposed protocol are presented in Section 4. Finally, the paper is concluded in Section 5.

2. Related work

In this section, we present a thorough study on the existing protocols that prolong the network lifetime by structure based [19–27] and structure-free [2,28,29] data aggregation. In structure based data aggregation, the data are transmitted to the base station by creating chain [19], tree [20], cluster [21–25], tree-cluster [26] or hierarchical clustering [27]. The Power-Efficient Gathering in Sensor Information Systems (PEGASIS) [19] is a chain-based protocol that forms a chain by using greedy algorithms. Each node performs data fusion where it fuses its own data with neighbor's data to generate a single packet of the same length and then transmits it to its next neighbor. A leader node is selected in the chain for transmitting data to the sink. The TRee based Energy Efficient Protocol for Sensor Information (TREEPSI) [20] is a tree-based protocol that selects root node randomly among all the sensor nodes. Then it starts building hierarchical paths to form a tree structure. The path is computed either centrally by a sink or broadcasting the path information over the network or locally by using a common binary tree construction algorithm in each node.

The protocols such as Low Energy Adaptive Clustering Hierarchy (LEACH) [21], Threshold-sensitive Energy Efficient protocol (TEEN) [22], Adaptive Periodic Threshold-sensitive Energy Efficient protocol (APTEEN) [23], and HEED [25] form the clusters of sensors. Data from different sensors are aggregated at the cluster head (CH) and the CH sends these data to downstream CH or to the base station. The LEACH [21] selects the CHs randomly on the basis of remaining energy to distribute energy load uniformly among sensor nodes in a network. The cluster members send their data directly to the CH for data fusion and the fused data are forwarded to the BS. The Centralized LEACH (LEACH-C) [24], a variant of LEACH, uses simulated annealing method to select the CHs for specific time slots so that the average transmission power between sensors and their CHs is minimized. This is a centralized approach that cannot be scaled to very large numbers of sensors.

TEEN [22] is a 2-tier clustering topology where the CH sends two threshold values to its members. The first one is a hard threshold value of the sensed attribute and the next is a soft threshold value which is a small change in the value of the sensed attribute. The hard threshold tries to reduce data communications by allowing the nodes to transmit only when the sensed attribute is in the range of interest. The soft threshold further reduces data communications when there is little or no change in the sensed attributes. The APTEEN [23] is an extension of TEEN that aims at both transmitting periodic data and reacting to time critical events. It allows user to set the count-time interval (CT) and the threshold values for the better energy utilization during data communication. The count time is the maximum time period between two successive reports sent by a node. In TEEN and APTEEN protocols, the CH is overburdened with parameter calculations and settings.

The Hybrid Energy-Efficient Distributed clustering (HEED) [25] is a multi-hop WSN clustering algorithm where CHs are periodically selected based on the residual energy of the nodes and the cost required by that node during intra-cluster communication. It is an energy-efficient clustered routing technique.

The Tree-Clustered Data Gathering Protocol (TCDGP) [26] is a hybridization of tree and cluster based approach. The TCDGP forms cluster on the basis of location and energy information about sensor nodes. This builds a path in the minimum spanning tree using the Prim algorithm. The sink computes the distance between the CHs to form tree instead of CHs that decrease computational overhead of the CHs. These structure based data aggregation protocols consume a significant amount of energy during the construction of network structure.

The Hierarchical Energy Efficient Reliable Transport Protocol (HEERTP) [27] is hierarchical cluster based transport protocol that minimizes energy consumption by minimizing sensed redundant data transmission with the co-ordination of the BS. The BS detects the redundant data even without receiving the redundant data from the sensor node. The proposed method identifies the redundant data at the receiver side when timeout occurs. If the receiver receives non-redundant data, then it updates the data table of the BS. These structure based data aggregation protocols consume a significant amount of energy during the construction of network structure. Therefore, structure free data aggregation protocols are proposed to handle redundant data without consuming the energy required for construction of structures.

The structure-free data aggregation protocol proposed by Fan et al. [28] does not use any explicit structures for data aggregation. This protocol achieves spatial convergence (packets to meet at the same node) through a MAC layer anycast based approach called Data-Aware Anycast (DAA) and temporal convergence (at the same time) through Randomized Waiting (RW) at the application layer of the source node. The combination of DAA with RW improves the normalized load (in terms of the number of transmissions) compared to opportunistic aggregation, and it performs better than the structured approach when the aggregation function is not perfect. DAA is based on anycasting at the MAC layer to determine the next-hop for each transmission. Anycasting requires the use of Right To Send (RTS) packets to elicit Clear To Send (CTS) responses from the neighbors before transmission of the packet.

Another structure-free data aggregation protocol (RAG) [2] is proposed to handle redundant data without consuming the energy required for construction of structures. The RAG uses judicial waiting policy and real-time data aware of anycasting for handling both temporal and spatial redundancy in real-time WSN. The judicial waiting policy calculates a waiting time out for each forwarding packet at the intermediate nodes in such a way that it can be delivered to the BS within the stipulated time bound. The real-time data aware of anycasting policy help a node to decide the next hop that can achieve better aggregation performance. However, this protocol increases energy consumption due to the broadcast of control message to all the neighboring nodes by a node. The packets may experience more delay at near sink nodes and the RAG increases the speed of the packet transmission to meet the deadline constraint which increases more energy consumption as well as may incur congestion.

Lastly, we studied structure-free and energy-balanced data aggregation protocol (SFEB) [29] that operates in a multi-hop network. It assumes that the packet with same event identification (EID) can be aggregated. The protocol works in two phases. In phase one, primary aggregators (PA) and secondary aggregators (SA) are selected by partitioning the network into virtual parallelograms. The virtual parallelograms are constructed by using the communication range of sensor nodes. The aggregator pair collects data and finds a node to forward its data toward the sink. In phase two, the aggregator as well as orphan nodes sends their collected information to the sink. The data aggregation effect is improved by selecting a waiting time as a summation of the aggregator selection time, data collection time and ACK transmission time.

3. Energy efficient structure-free data aggregation and delivery (ESDAD) protocol

In this section, we present our proposed framework for energy efficient data aggregation and delivery in structure free WSNs. The first subsection presents the procedure to construct the logical topology. In the next subsection, we present the approach to select the sensors that are eligible to transmit the sensed data depending on the required reliability of the occurred event. The judicial waiting policy for efficient data aggregation and data forwarding is presented in the next subsection. In the last subsection, we present an efficient congestion control mechanism to reduce the packet loss and the local recovery of the lost packets.

3.1. Logical topology construction

The topology control minimizes the issues evolved from a redundant number of nodes and their dense deployment i.e. interference, maximum number of possible routes, use of maximum power to communicate to distant nodes directly. The topology control preserves connectivity with the use of minimal power. A sensor node must know about its own position, the position of the neighboring nodes and base station during logical topology construction phase.

The logical topology construction phase is initiated by the BS after the deployment of sensors in an area. The BS initiates topology construction by broadcasting a "*HELLO*" message. The nodes that receive this "*HELLO*" message further transmit "*HELLO*" message to carry the logical topology construction process. Each node transmits the "*HELLO*" message after a random waiting period to avoid collision among peer nodes. The "*HELLO*" message contains node ID, location information, energy level, buffer status and its *hop-count* (*hc*) to reach to BS. The *HELLO* message broadcast by the BS contains $hc = 0$ (Zero). The sensor node that hears the *HELLO* message transmitted from BS sets its *hc* as one and transmits *HELLO* message. The node that receives this HELLO message sets its *hc* as two and so on. A node selects its neighbor with minimum *hc* value from the received multiple *HELLO* message and sets its own *hc* as received minimum $hc + 1$. The nodes having *hc* one form the first logical level. The nodes having *hc* two form the logical level two and so on. This process continues till all the nodes are included in the hierarchy or the time for logical topology construction is over. The orphan nodes that are not included in the logical topology, broadcast *HELLO* messages to know their position after time out. The orphan node sets their position and *hc* after receiving a *HELLO* message from their neighborhood.

We don't construct a static structure rather we use a structure free topology where the hierarchy of sensor nodes decides the data forwarding to a level not to a specific node. After the topology construction phase, each node knows its own logical level, available energy, position and buffer occupancy of all the neighboring nodes in its radio range. The logical topology is constructed only once at the beginning and is not required to be repeated like other structure based topology constructions. The initial topology construction phase helps to identify the neighborhood of a node. The WSN is dynamic in nature and the topology changes when the node dies. The structure based topology control protocols initiate topology construction phase when the network energy goes below a threshold energy level or after a significant number of nodes die in the network. This increases a significant amount of energy consumption in WSN. The proposed protocol saves the energy wasted in topology construction by adopting a structure-free data delivery approach. In case of node death in the neighborhood of a node, the new nodes are selected for data transmission on the basis of the cost function. Thus, the proposed protocol can handle topology changes without reconstructing the topology. Fig. 1 shows the logical topology construction in a sensing field. Fig. 1(a) shows the HELLO packet transmission initiated by the BS. Fig. 1(b) shows that the nodes N1, N2, N3 and N4 transmit the HELLO packet for further neighbor discovery and constructing the logical level one. Fig. 1(c) shows the hierarchical HELLO packet transmission for the rest of the level identification and neighbor discovery. Fig. 1(d) shows the formation of all logical levels in the sensing field.

3.2. Judicial data transmission

The dense deployment of sensor nodes in the sensing field produces highly co-related and redundant data. The energy consumption of the WSN due to data transmission can be minimized by selectively forwarding the sensed data to the aggregation point. Most of the WSN applications require different levels of sensing reliability in the sensing field. For example in the forest fire monitoring application, the entire forest can be divided into subregions. The different subregions of the forest need to be monitored with different levels of interest. The subregion containing precious trees needs to be monitored with higher interest than a subregion containing un-precious trees. Similarly, in WSN based healthcare applications the patients in ICU should be monitored with a better interest than the patients in indoors. Therefore, the subregions may be assigned with different reliability weights ($w_j, 1 \leqslant j \leqslant n_s$), where n_s is the number of subregion. The weight factor assigned to the subregion decides the QoS requirements such as delay and data delivery ratio of that region. Fig. 2 shows the sensing field with different levels of sensing reliability requirements.

The more we collect data about the event produces a more accurate picture of the event. Thus, we can say that the amount of data collected about the event is directly proportional to the event occurrence reliability [31,32]. The proposed method divides the entire sensing zone into subregions based on the required level of reliability requirements. The number of sensors allowed to transmit the sensed data is decided on the basis of the reliability requirement of that region. The event occurred in higher reliability required subregion needs to collect more data from the sensor nodes than the lower reliability required subregion. Thus, the proposed protocol judicially transmits data to achieve the required level of reliability in a subregion to conserve energy.

A sensor node can sense the event if the event occurs within a sensing radius (d_{sense}). After sensing the event, each sensor node assigns a reliability factor (r_f) to its sensed data. The r_f of sensory data is computed by using Eq. (1).

$$r_f = \frac{d_{sense} - d_{Event\_Sensor}}{100} \tag{1}$$

where d_{Event_Sensor} is the distance between the event and the sensor node that senses it. The d_{sense} is the sensing radius of the sensor node. The node closes to the event, producing more accurate value than the nodes far away from it [31,32]. For example, the heat and smoke produced due to the fire broke out in the forest spread gradually. Thus, the sensor node near to the event detects more accurate value as compared to far away node. On the other hand the far away nodes play a vital role in finding radius of the affected region.

A sensor node decides whether to transmit the data on the basis of its r_f or forbids. A sensor node decides to transmit the data if $r_f \geqslant tr_f$. The tr_f for a node is the threshold reliability factor of an event. The tr_f for a node is computed by using Eq. (2).

$$tr_f = \frac{100 - w_j}{100} \tag{2}$$

3.3. Next-hop node selection

A sensor node selects the next-hop node to transmit its own packet, or aggregated data packet toward BS. The sensor node that senses unique data (data different from its neighbors) selects the next-hop node sensibly to transmit the sensed data. The next-hop node selection is an important issue in the structure-free aggregation and routing. Fig. 3 shows the pseudo code for next-hop selection.

In the proposed protocol, the next-hop node is selected on the basis of a cost function. The cost function is computed from the residual energy and available buffer space of the next-hop node and strength of the link between these current node and next-hop. Each sensor node maintains a neighbor information table that helps in computing cost function and next-hop node selection during data forwarding. A neighbor

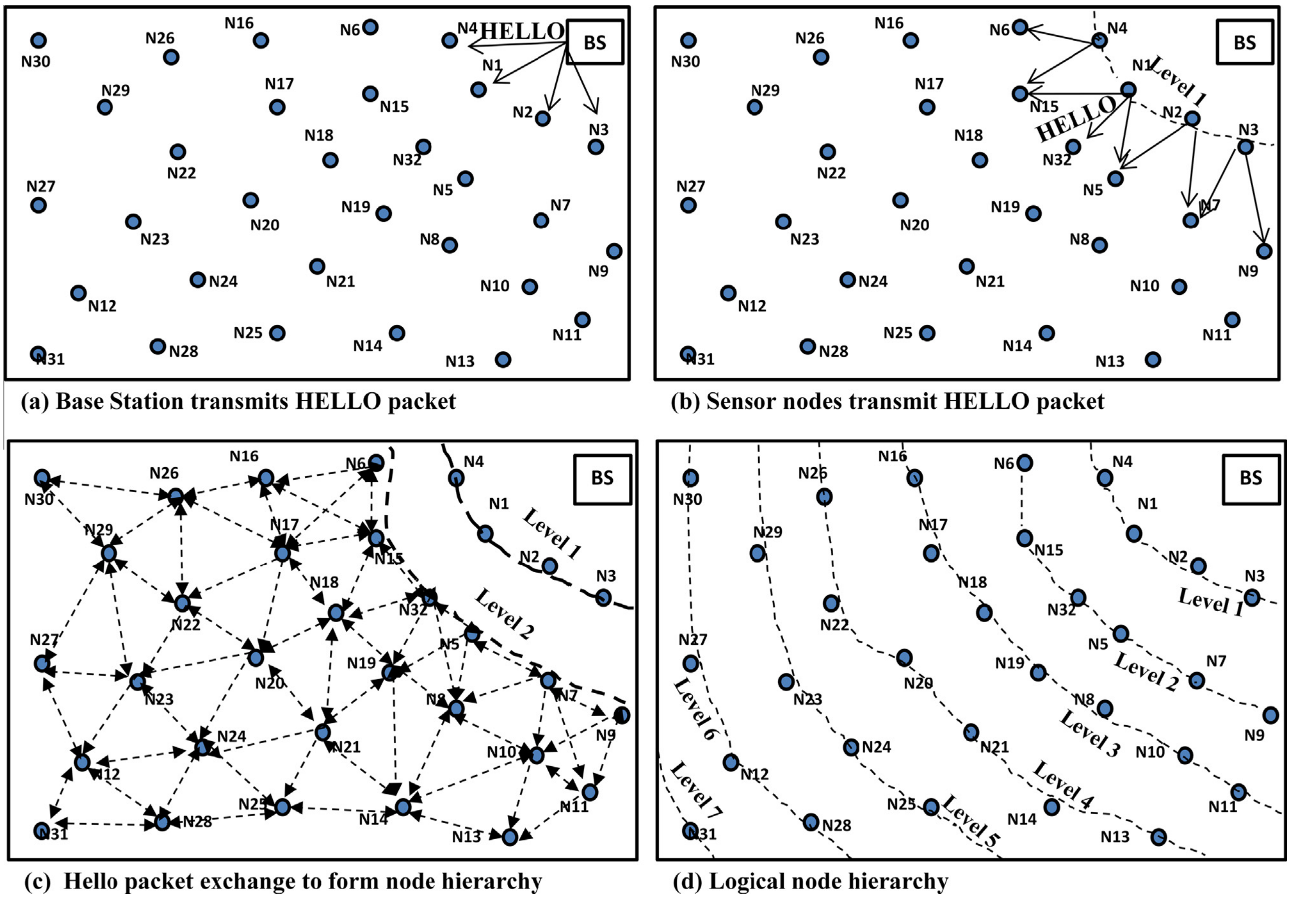

Figure 1 Logical topology constructions in a sensing field.

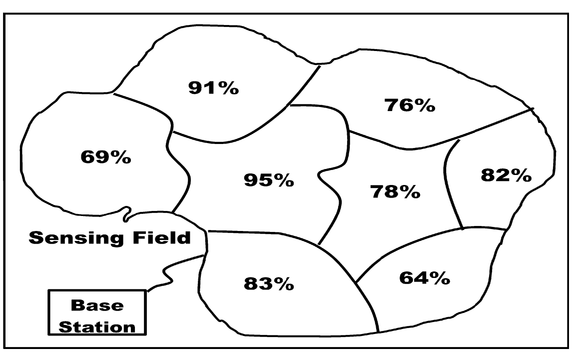

Figure 2 Sensing field with required different levels of sensing reliability.

information table contains the information of next-hop neighboring nodes that are node id (N_{ID}), coordinate position ($N_{(x,y)}$), available buffer ($Buff_{st}$), link strength (l_s) and residual energy (E_{resd}). When a node senses data or receives a data packet from the nodes of lower level for forwarding it to BS, the node first computes the cost function for all its next higher level nodes from the neighbor information table. A node j (N_j) selects a next hop node i (N_i) with maximum cost function value (cf_{max}). The cf_{max} is computed as given below.

$$cf_{max} = \max_{i \in N}\{\propto (E_{resd.i} + Buff_{aval.i} + l_{s.i})\} \tag{3}$$

where N represents a set of neighbors of N_j and α is the weight factor computed as the inverse of distance between N_j and N_i.

$$\propto = 1 \Big/ \sqrt{(N_{j.x} - N_{i.x})^2 + (N_{j.y} - N_{i.y})^2} \tag{4}$$

The residual energy of node i is computed as given below.

$$E_{resd.i} = E_{level.i} - \left\{ (E_{TX}(k, d_{tran}) + E_{RX}(k)) + \sum_{i=1}^{N} E_{RX}(k) + E_{agg} \right\} \tag{5}$$

The $E_{TX}(k,d)$ is the energy required to transmit k number of bits to a distance d_{tran}, where d_{tran} = maximum transmission range of a sensor node, $E_{RX}(k)$ is energy spent for receiving a packet which is computed as $E_{RX}(k) = E_{elec} \times k$ and E_{agg} is energy spent to aggregate n_p number of packets.

The available buffer space is computed from the current buffer status and the expected number of packets to be transmitted from neighborhood of N_j. The $Buff_{st}$ and E_{level} are generally piggybacked in acknowledgment packets sent for the data packet and updated in the neighbor information table. At the initial stage, when the first round of communication begins the $Buff_{st}$ and E_{level} of a neighbor node are set as the value received during topology construction phase. The E_{level} of neighbor node is sent with hello packet during topology construction phase and the B_{uffst} status is assigned a value

```
next-hop_node()
Define: DID destination node ID
Define: Cf [i] cost function value of node i
begin:
   sensor node has DS to transmit;
   if ( receive (control_packet)== true ) then
      set next-hop node_ID = control_packet DID;
      set transmission_schedule for the node;
   else
      The sensor node(s) begin sense data DS of Event E;
      For i = 1 to en do
         Cf [i] = α X(Eresd.i + Baval.i + ls.i);
      end for;
      set Cfmax = maximum { Cf [1,2,…,en] };
      set next-hop node_ID = Cfmax node_ID;
      set transmission_schedule for the node;
   end if;
end;
```

Figure 3 Pseudo code for next-hop node selection.

equal to the total length of buffer available for a node. The buffer available to a node i is computed as given below.

$$Buff_{aval.i} = Buff_{st.i} - \sum_{i=1}^{N} k \tag{6}$$

The link strength for a neighbor node is computed and updated in the neighbor information table when the node receives an acknowledgment packet from the neighbor node by using Eq. (7). The link strength is the signal interference noise ratio (SINR) for the link between N_j and N_i. The $\mathrm{Rec}_{\text{Signal Power}}$ is computed from Eq. (8) and $\mathrm{Rec}_{\text{no.of bits}}$ is the number of bits present in an acknowledgment packet from the neighbor node N_i.

$$l_{s.i} = \frac{\mathrm{Rec}_{\text{Signal Power}}}{\mathrm{Rec}_{\text{no.of bits}}} \tag{7}$$

The $\mathrm{Rec}_{\text{Signal Power}}$ is computed as given by Sergiou et al. [33]

$$\mathrm{Rec}_{\text{Signal Power}} = \left[\frac{P_r(d)}{P_r(d_0)}\right]_{db} = -10\beta \log\left(\frac{d}{d_0}\right) + X_{db} \tag{8}$$

where $P_r(d)$ is the mean received power at distance d, which is computed relative to a reference power $P_r(d_0)$ at distance d_0. β is the path loss exponent and X_{db} is a Gaussian random variable with zero mean and standard deviation δ_{db}.

3.4. Structure-free data aggregation

A near source data aggregation can reduce energy consumption, delay and traffic load than a near sink aggregation. In structure based data aggregation, data are collected periodically from all its children and then aggregated to form single packet. The structure based data aggregation aggregates fixed number of packets in the routing path. The number of packets needed to be aggregated depends on the routing structure. In structure-free aggregation, the number of packets aggregated varies from node to node may be much more than structure-based data aggregation. With the increase in the number of packets to be aggregated, the traffic load, buffer requirement and chance of network congestion increase. Thus, we aggregate selected packets for an event as early as possible.

The nodes that have the data to transmit set their timer as $(100 - r_f)$ at the beginning of the control period. The nodes, then start decrementing their timers. When the timer of a node becomes zero, it sends the control packet to its neighbor selected as described in Section 3.3. The other nodes freeze their timer and listen to the control packets of that sender. The control packet contains the source id, next-hop id, packet type and waiting period. The control packet is overheard by the sensor nodes that are in the low power listening (LPL) area of that node. The neighbor nodes set their transmission schedule and data forwarding to the next-hop node. The selected next-hop node aggregates the received and forwards the data to the next-hop in the direction of BS.

3.5. Waiting period for data aggregation and data forwarding

The received data packets are buffered at the aggregating node for data aggregation before transmitting it to the next-hop. The data received from the nodes are aggregated after expected waiting time duration (tw_{expt}). The aggregating node waits for tw_{expt} to collect a significant number of data packets before data aggregation. The spatial and temporal convergence of data packets toward the aggregator node can be improved through tw_{expt}. The tw_{expt} of a node for data aggregation is computed by using Eq. (9). For real-time data delivery in WSN, the sensed data packet must be reported to the BS with certain delay bound. The delay components depend on propagation delay (τ_{pro_delay}), transmission delay (τ_{tran_delay}), channel access delay (τ_{chan_delay}) and buffering delay (τ_{buff_delay}).

$$tw_{exp} = \{(\mathrm{TTD} - \tau_{pro\_delay}) - (hc[\tau_{tran\_delay} + \tau_{chan\_delay} + \tau_{buff\_delay}])\} - \beta \tag{9}$$

The Time-To-Deadline (TTD) is computed as remaining time to meet the deadline. The τ_{pro_delay} is computed as d_{n_BS}/p_s where d_{nagg_BS} is the distance between aggregator node (nagg) and BS. p_s is the wave propagation speed. The τ_{tran_delay} is computed as k/r_t where k bits of data are transmitted at the data rate of r_t. For the simplicity of computation, we assume that, per hop buffering delay and channel access delay do not vary. The β is the random slack time margin that provides safety for a timely data delivery of aggregated data packet to the BS. The tw_{expt} is reset if a new data packet of the same event is received with a lowest delay time bound.

Structure based routing protocols in sensor networks are designed to discover fixed path between the source destination

```
data_aggregation()
Define: Dagg aggregated data packet of an event E
Define: S sensor node perform data aggregation
Define: Di sense data of node i
begin:
   for all receive (Di) until  twexpt of S expire do
   if Di ϵ E and S is the aggregator node then
      if (buffer_available == true)then
          store Di;
      else
          find low rf data_packet of E from the buffer;
          if (low rf < Di rf)then
              replace low rf data_packet with Di;
          else
              drop Di;
                end if;
          end if;
    end for;
    find median (MD)among all stored Di
    set Dagg = MD;
    select next-hop_node();
    transmit(Dagg);
end;
```

Figure 4 Pseudo code for data aggregation.

pair for data transmission. A new path is established when the energy of the next hop node is exhausted or route fails due to network faults. Therefore, the route maintenance cost is high and proper load balancing is also a challenging task. On the other hand, the structure-free routing protocol dynamically selects the next hop nodes for data forwarding. Thus, in each round of communication a different routing path may be selected dynamically. This structure-free routing can provide load balancing and fault tolerance and reduce the chance of congestion.

We assume that the WSN consists of sensor nodes that are initially deployed randomly, but uniform in space and the BS is located at a specific point in the network topology. Each sensor node has a unique identity (N_{ID}). The number of source nodes varies since an event is possible to be random and be captured by more than one sensor nodes. The base station is connected to a data collection center via an external network. We also consider that each node knows its position and the position of the BS. The BS has no resource limitation and the sensors are battery-operated with limited energy and have the same physical capabilities. The sensor works no more if its energy exhausts. Moreover, the source nodes may transmit different types of data packets and the intermediate nodes are responsible for performing in-network aggregation of same-type packets. The nodes that are not adjacent communicate with each other through hop-by-hop. A node transmits a packet with size k bits to distance d. The data aggregation procedure compresses n packets with size k to produce one packet of size k. If the packet size is bigger than k, it will be sliced into pieces with constant size k. If the size is smaller than k, it will be enlarged into size k [34]. Fig. 4 shows the pseudo code for data aggregation.

3.6. Congestion control and buffer management

The buffer of each node is divided into three parts. The first part of buffer stores aggregated data packets from its neighborhood. The second part of the buffer is utilized by the orphan flow and the control packets, exchanged between the nodes. The third part stores the data packets needed to be aggregated by the node. Multiple events may occur in a sensing field and they may need to be reported through a node. Thus, the third part of the buffer needs to be shared among the multiple event flows that need to be aggregated. We name the third part of the buffer as event flow. Fig. 5 shows the buffer partition of a sensor node. Both data aggregation and orphan flow have en (expected number of neighbors) buffer space for each. The event flow ($Buff_{event\_flow}$) occupies $Buff_{event\_flow} = Buff_{total} - 2 \times en$ buffer space. The proposed procedure shares the queue in each node for each flow passing through the node. However, the boundaries between queues are not fixed means if one of the active flow has free space in its queue then other flow facing a lack of space can use this free space on demand. Let us consider that in a sensor network, each sensor node always has traffic to send. The traffic originated by the event j is denoted as the jth flow i.e. f_j. Each event flow f_j is assigned a buffer space ($Buff_{event,j}$) from buffer of $Buff_{event\_flow}$ which is computed from Eq. (10).

$$Buff_{event,j} = \frac{Buff_{event\_flow}}{en} \tag{10}$$

The threshold (Q_{th}) for queue occupancy of an event flow f_j is set as $\frac{3}{4} \times Buff_{event,j}$. When a new packet arrives, the node computes the hit frequency (h_t) by examining whether the packet is from the same flow as one in the buffer, if yes then h_t is increased by one. For a new flow first buffer is allocated if available, and then the h_t is computed from the beginning. The node with higher hit frequency and low buffer available space has more chance of congestion. The expected congestion source node is informed to adjust its transmission rate so that the congestion can be avoided. The f_j flow is assigned fair and efficient transmission rate flow r_j. The r_j is reduced by $\frac{1}{ht} \times r_j$ during chance of congestion. The proposed protocol tries to avoid congestion by predicting a chance of congestion from

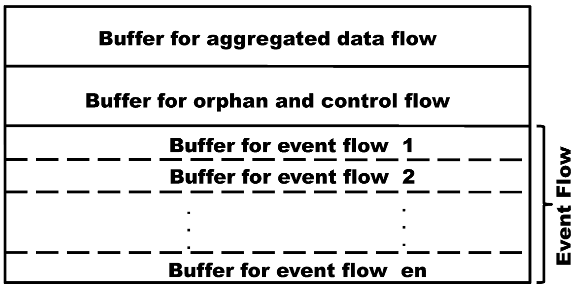

Figure 5 Buffer partition of a sensor node.

the available buffer space and adjusting the transmission rate of the source nodes of the event flow. If congestion occurs even after congestion avoidance steps, then the received packets are dropped to avoid congestion. When the buffers of a flow reach to $Buff_{event.j}$ and no further free space is available, then the dropping probability (p_d) of the arriving packet is computed using Eq. (11).

$$p_{d.j} = \frac{3}{4} \times \frac{h_{t.j}}{Q_{th.j}} \tag{11}$$

The fairness of data transmission is achieved by transmitting the data packet according to their remaining earliest deadline. Furthermore, the transmission of the data packets is scheduled in a ratio of 2:1 for aggregated buffer flow and orphan flow. In case of unavailability of data packets in one flow, the other flow takes the chance of data forwarding.

Fig. 6 shows the flowchart that summarizes the basic steps of the proposed ESDAD protocol.

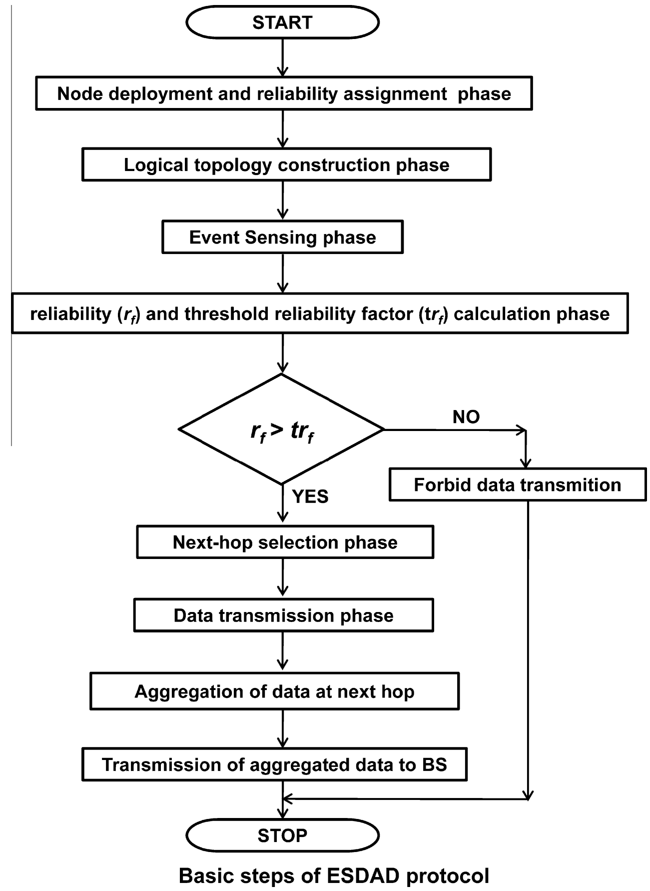

Figure 6 Basic steps of ESDAD protocol.

4. Performance evaluation

In this section, we evaluate the performance of our proposed protocol through NS-2.30 simulation. The goal of the simulation is to compare the performance of our proposed protocol, ESDAD, with the existing structure-free data aggregation protocols such as RAG [2], and (SFEB) [26] that operates in a multi-hop network. Table 1 summarizes the simulation parameters of the proposed protocol. We consider 400 sensor nodes deployed randomly in an area of 500 × 500 m^2. The transmission range (d_{tran}) of a sensor node is set to be 50 m with a data rate of 200 kbps. The sensing range of a sensor node is 50 m. The packet length is 60 bytes long. Each sensor node has an initial energy of 0.6 J (joule). The energy consumption for transmitting and receiving a bit is 50 nJ/bit. The energy spent in sensing, aggregation and radio amplification is 0.083 J/s, 5 nJ/bit/signal and 10 pJ/bit/m^2 respectively. Events are generated randomly in every 3 s in the sensing field. The packets with same event ID (EID) can be aggregated. The buffer length of each node is set as 65 packets. The sensing area is divided randomly into subregions and each subregion randomly assigned sensing reliability in each simulation. We run the simulation 25 times each for 35 s. The average value of the runs is considered for best possible results.

We study and compare the performance of our proposed protocol with other existing protocols in terms of average energy consumption, miss ratio and end-to-end delay. The average energy consumption is the most important parameter to evaluate the performance of the WSN. It shows the energy consumption during data transmission and helps to predict the life span of the entire sensor network. The miss ratio is computed as the percentage of packets that are not delivered to the BS on time and discarded in delay sensitive application. We set Time-To-Deadline (TTD) as in RAG [2] for each generated packet to compute the miss ratio. The miss ratio plays a vital role in the performance evaluation during reliable event reporting in the congested real-time sensor network application. The end-to-end delay is the time from the packet generated by the source till it is delivered to the destination. The end-to-end delay plays a vital role during the performance evaluation in a time bound data communication system.

Table 1 Simulation parameters.

Area of sensor field	500 × 500 m^2
Number of sensor nodes	400
Packet length	60 bytes
Buffer length	65 packets
Initial node energy	70 J
Bandwidth	200 Kb/s
Sensing length	50 m
Radio range	40 m
Propagation model	Two ray
Eelec	50 nJ/bit
Esense	0.083 J/s
Eagg	5 nJ/bit/signal
Eamp	10 pJ/bit/m^2

Figs. 7–10 show the comparison of average energy consumption of our proposed ESDAD protocol with SFEB and RAG protocols for varying data rate, sensing reliability, event generation time and the number of nodes respectively. It is observed that ESDAD conserves more energy than RAG and SFEB with the increase in data rate, sensing reliability, event generation time and the number of nodes respectively. RAG adopts a Judiciously Waiting policy for efficient aggregation of data packets to conserve more energy and tries to eliminate the inherent redundancy of raw data. In SFEB, all nodes transmit data for aggregation result of the whole network involved in the data transmission process. The ESDAD out-performs all these different proposed protocols due to selective data forwarding and aggregation in time. Thus, it saves the burden of broadcasting the data packet to its entire neighbor. It performs near source data aggregation, which reduces traffic load in multi-hop transmission. In Fig. 8, the sensing reliability of the subregions increased from 10 to 100 respectively. We found that the energy consumption remains constant for RAG and SFEB. The energy consumption of ESDAD increases proportionally with respect to sensing reliability. With the increase in sensing reliability the data transmission node, i.e. source nodes increases; thus, average energy consumption of the WSN increases.

Figs. 11–13 show the performance in terms of miss ratio with respect to the increase in data rate, event generation time and the number of nodes respectively. When an event is generated in a sensing field the RAG and SFEB protocols transmit the sensed data of all the source nodes that sense the data. The ESDAD protocol selectively transmits packet for data aggregation that satisfies the transmission reliability criteria. The packets are dropped in the buffer during aggregation due to lack of buffer space.

In Fig. 11, it is observed that the packet drop rate increases with the increase in data rate proportionally. However, the proposed ESDAD protocol performs better than the existing protocols due to less number of source nodes than RAG and SFEB. Fig. 12 shows the miss ratio of ESDAD, RAG and SFEB protocols with respect to the increase in event generation time. The traffic load decreases with the increase in event generation time leading to the decrease in packet miss ratio. Furthermore, the traffic load in our ESDAD is less as compared to the other structure free data aggregation protocols. Thus the miss ratio of our ESDAD protocol is less than RAG and SFEB.

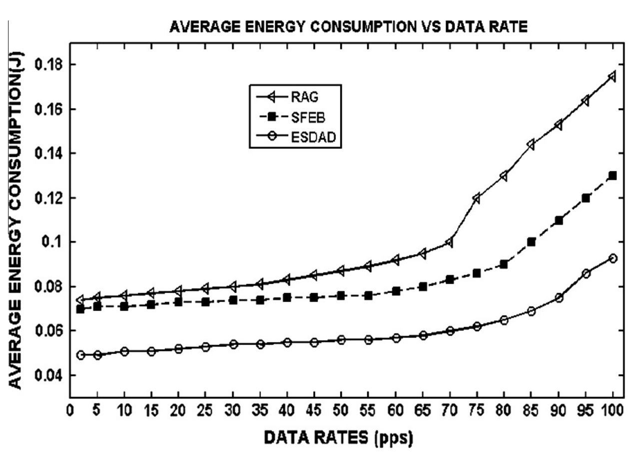

Figure 7 Average energy consumption in different data rates.

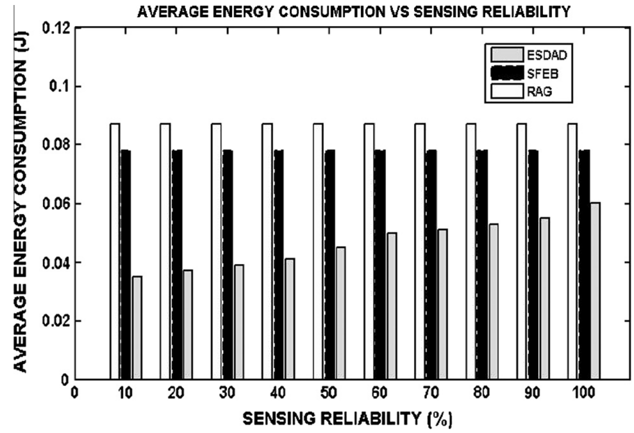

Figure 8 Average energy consumption with respect to different sensing reliability.

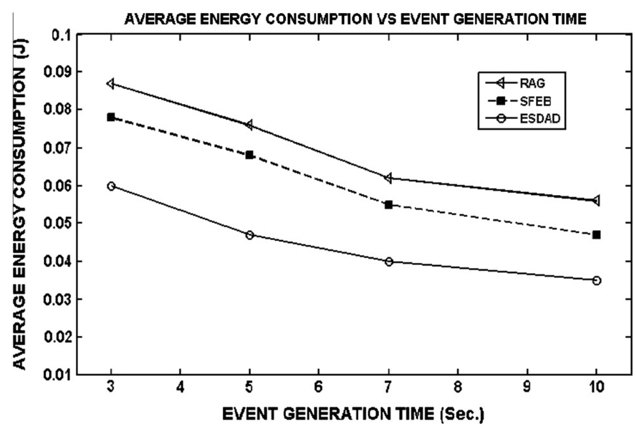

Figure 9 Average energy consumption with respect to event generation time.

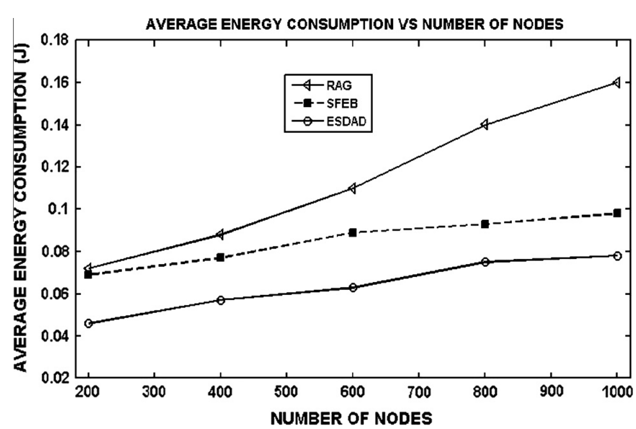

Figure 10 Average energy consumption with respect to the number of nodes in the sensing field.

Fig. 13 shows the miss ratio with respect to the number of nodes. There may be a large number of sensor nodes in a region due to the dense deployment of sensors in a WSN to sense the events occurred in that region. Therefore, an event may be sensed and transmitted by a number of sensors which

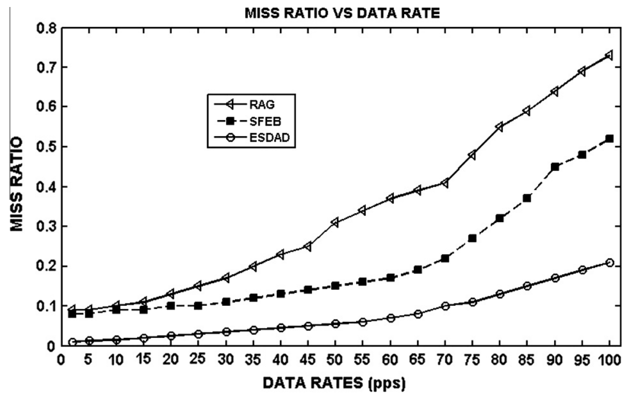

Figure 11 Miss ratio in different data rates.

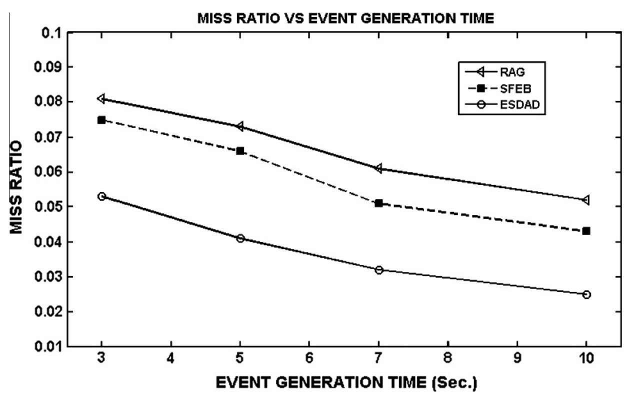

Figure 12 Miss ratio with respect to event generation time.

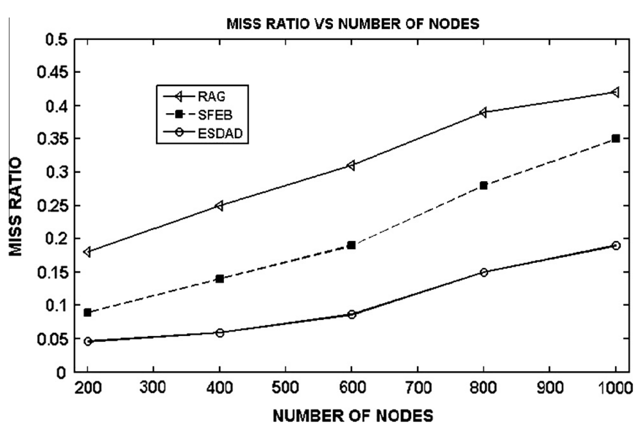

Figure 13 Miss ratio with respect to the number of nodes in the sensing field.

increases the traffic load and buffer overflow at the intermediate sensors. The buffer overflow leads to packet drop at the intermediate sensor nodes. The packet drop rate increases with the increase in traffic load. The traffic load of a sensing region increases, if the node density of that region is more. Thus, we can say that packet drop rate is directly proportional to the node density of that region. The packet drop rate due to increased traffic load can be minimized by decreasing the traffic load. The traffic load decreases, if the data aggregation is performed as early as possible. The data aggregation near

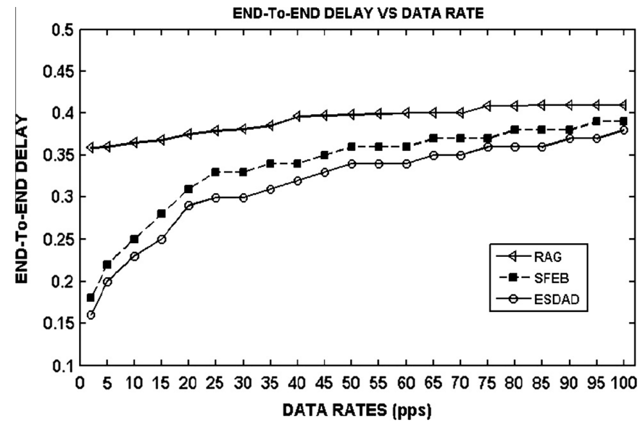

Figure 14 End-to-end delay in different data rates.

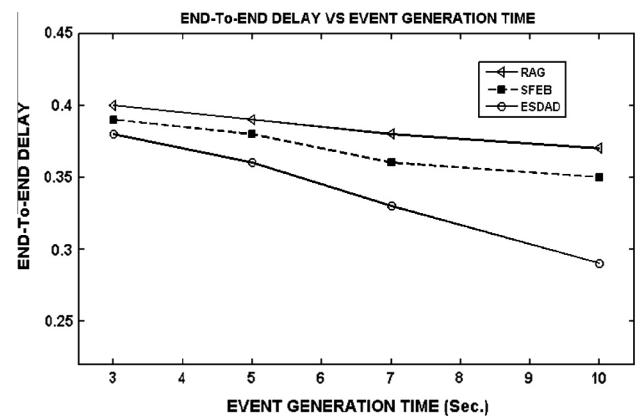

Figure 15 End-to-end delay with respect to event generation time.

the source is the earlier data aggregation technique than data aggregation near the sink and it reduces the traffic significantly. Thus, the ESDAD outperforms in the high density due to near source aggregation whereas SFEB performs near sink aggregation by selecting primary aggregator near to sink.

Figs. 14–16 show the end-to-end delay with respect to increase in data rate, event generation time and the number of nodes respectively. Our waiting policy dynamically adjusts the waiting time limit in the aggregator node on the basis of deadline of the packet. Fig. 14 shows that end-to-end delay of RAG, SFEB and ESDAD increases with the increase in data rate. In RAG, due to increase in data rate the end-to-end delay increases slowly. This is only due to the increase in congestion. However, the end-to-end delay of SFEB and ESDAD proportionally increases with an increase in data rate due to the congestion, waiting time and loss recovery. Our proposed protocol has less traffic load than SFEB and RAG because of the selected number of senders. Therefore, the performance of proposed ESDAD is better than the existing protocols in terms of end-to-end delay.

Fig. 15 shows end-to-end delay with respect to increasing event generation time. The traffic load as well as event reporting frequency decreases with increase in event generation time. Thus the better aggregation and loss recovery are possible in ESDAD. The end-to-end delay is reduced in case of increasing

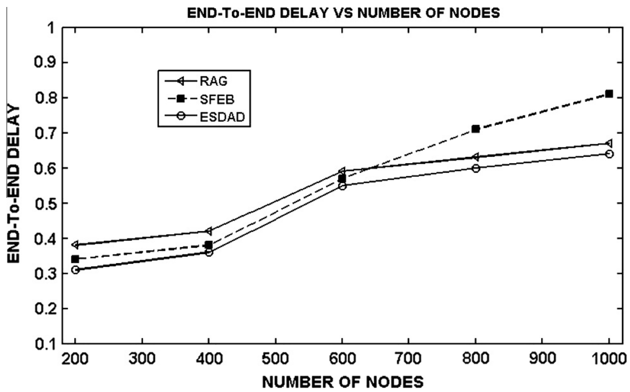

Figure 16 End-to-end delay with respect to the number of nodes in the sensing field.

at the event generation time. The end-to-end delay of RAG SFEB and ESDAD versus the number of nodes is shown in Fig. 16. In SFEB, the data aggregation time depends on the downstream nodes of the aggregation tree. The aggregation tree grows with the node density and thus the end-to-end delay increases in SFEB. The waiting time of RAG is computed judicially in the intermediate nodes during the aggregation process. It aggregates the packet on the basis of their sensitive packet delivery time, i.e. Time-To-Deadline (TTD). In RAG and ESDAD, the prioritized transmission schedule of data packets helps to meet the deadline of the delay sensitive data packets. Thus, in high node density RAG and ESDAD perform better packet delivery than SFEB. However, our proposed protocol outperforms RAG and SFEB in terms of delay due to the prioritized data forwards to real-time packets and less traffic load and loss recovery.

5. Conclusion

In this paper, we propose a reliable energy efficient and structure free data aggregation protocol for WSNs. In the ESDAD, we consider the sensing region divided into different subregions that are sensed with different reliability requirements. Our proposed protocol allows a selected number of senders to transmit the sensed data depending on the reliability requirement. This not only saves the energy consumption of the sensors, but also decreases the traffic load in the network. This less traffic load in case of our proposed protocol in comparison with the existing protocols decreases the end-to-end delay due to congestion and loss recovery. Furthermore, the structure-free data aggregation approach used in our protocol also saves energy consumed due to the computation of a structure and also increases the performance of WSN. We also considered the problems of near sink data aggregation in WSNs and proposed a near source data aggregation. The efficient next node selection method used in our protocol improves spatial and temporal convergence for data aggregation. The sensed data are aggregated selectively to improve energy consumption and decrease miss ratio as well as end-to-end delay. The miss ratio is also minimized through an efficient buffer partition and management. In our future study, the proposed protocol needs to be modified and tested to adopt the real-time dynamic environment. The protocol needs to be tested for real-time WSN applications that require diverse reliability required in the sensing field.

References

[1] Mohanty P, Kabat MR. Transport protocols in wireless sensor networks. In: EI Emay Ibrahiem MM, Ramkrishan, editors. Wireless sensor networks: from theory to applications. CRC Press, Taylor and Francis Group; 2013. p. 265–305 [chapter 10].

[2] Yousefi H, Yeganeh MH, Alinaghipour N, Movaghar A. Structure-free real-time data aggregation in wireless sensor networks. Comput Commun 2012;35:1132–40.

[3] Qian H, Sun P, Rong Y. Design proposal of self-powered WSN node for battle field surveillance. Energy Proc 2012;16:753–7.

[4] Rezaee AA, Yaghmaee MH, Rahmani AM, Mohajerzadeh AH. HOCA: healthcare aware optimized congestion avoidance and control protocol for wireless sensor networks. J Network Comput Appl 2013;37:216–28.

[5] Hadjidj, Souil M, Bouabdallah A, Challal Y, Owen H. Wireless sensor networks for rehabilitation applications: challenges and opportunities. J Network Comput Appl 2013;36:1–15.

[6] Jiang SF, Zhang CM, Zhang S. Two-stage structural damage detection using fuzzy neural networks and data fusion techniques. Expert Syst Appl 2011;38(1):511–9.

[7] Abreu, Mendes P. Wireless sensor networks for biomedical applications. In: 2013 IEEE 3rd Portuguese meeting in bioengineering (ENBENG); 2013. p. 1–4.

[8] Roopa, Kumar S, Manvi S. Image fusion techniques for wireless sensor networks: survey. ITSI Trans Electr Electron Eng 2014:13–9.

[9] Xinqing X, Lin Q, Lei Z, Xiaoshuan Z. Wireless real-time monitoring system for table grape cold-chain logistics. Appl Electron Tech 2013;8:77–83.

[10] Hadim S, Mohamed N. Middleware: middleware challenges and approaches for wireless sensor networks. IEEE Distr Syst 2006;7(3).

[11] Vuran MC, Akan OB, Akyildiz IF. Spatio-temporal correlation: theory and applications for wireless sensor networks. Comput Networks J 2004;45(3):245–61.

[12] Zhang W, Liu Y, Das SK, De P. Secure data aggregation in wireless sensor networks: a watermark based authentication supportive approach. Pervasive Mobile Comput 2008;4:658–80.

[13] Jayram TS, Mcgregor A, Muthkrishnan S, Vee E. Estimating statistical aggregates on probabilistic data streams. ACM Trans Database Syst 2008;33.

[14] Chen J-Y, Pandurangan G, Xu D. Robust computation of aggregates in wireless sensor networks: distributed randomized algorithms and analysis. IEEE Trans Parallel Distrib Syst 2006;17(9):987–1000.

[15] Huang, Leung H. An expectation maximization based interactive multiple model approach for collaborative driving. IEEE Trans Intell Transp Syst 2005:206–28.

[16] Dhasian HR, Balasubramanian P. Survey of data aggregation techniques using soft computing in wireless sensor networks. IET Inf Secur 2013;7:336–42.

[17] Croce1 S, Marcelloni1 F, Vecchio M. Reducing power consumption in wireless sensor networks using a novel approach to data aggregation. Comput J Math Stat 2007;51:227–39.

[18] Liao WH, Kao Yucheng, Fan CM. Data aggregation in wireless sensor networks using ant colony algorithm. J Network Comput Appl 2008;31:387–401.

[19] Lindsey S, Raghavendra C. PEGASIS: power-efficient gathering in sensor information systems. In: IEEE aerospace conference proceedings, vol. 3; 2002. p. 1125–30.

[20] Satapathy SS, Sarma N. TREEPSI: tree based energy efficient protocol for sensor information. In: 2006 IFIP international

conference on wireless and optical communications networks; 2006. p. 1–4.
[21] Heinzelman W, Chandrakasan A, Balakrishnan H. Energy-efficient communication protocol for wireless microsensor networks. In: Proceedings of the 33rd Hawaii international conference on system sciences (HICSS '00); 2000. p. 3005–14.
[22] Manjeshwar E, Agrawal DP. TEEN: a routing protocol for enhanced efficiency in wireless sensor networks. In: Proceedings of the 15th international parallel and distributed processing symposium (IPDPS), San Francisco, CA, USA; 2001. p. 2009–15.
[23] Manjeshwar A, Agarwal DP. APTEEN: a hybrid protocol for efficient routing and comprehensive information retrieval in wireless sensor networks. In: Parallel and Distributed Processing Symposium. Proceedings international, IPDPS; 2002. p. 195–202.
[24] Heinzelman WB, Chandrakasan AP, Balakrishnan H. An application-specific protocol architecture for wireless microsensor networks. IEEE Trans Wirel Commun 2002;1:660–70.
[25] Younis O, Fahmy S. HEED: a hybrid energy-efficient distributed clustering approach for ad hoc sensor networks. IEEE Trans Mobile Comput 2004;3:366–79.
[26] Huang KC, Yen YS, Chao HC. Tree-clustered data gathering protocol (TCDGP) for wireless sensor networks. In: Proceedings of the future generation communication and networking (FGCN 2007), vol. 02; 2007. p. 31–6.
[27] Mohanty P, Kabat MR. A hierarchical energy efficient reliable transport protocol for wireless sensor networks. Ain Shams Eng J 2014;5:1141–55.
[28] Fan Kai-Wei, Liu Sha, Sinha Prasun. Structure-free data aggregation in sensor networks. IEEE Trans Mob Comput 2007;6(8):929–42.
[29] Chao CM, Hsiao TY. Design of structure-free and energy-balanced data aggregation in wireless sensor networks. J Network Comput Appl 2014;37:229–39.
[30] Wong J, Jafari R, Potkonjak M. Gateway placement for latency and energy efficient data aggregation. In: Proc 29th ann IEEE int'l conf. local computer networks; Nov. 2004. p. 490–7.
[31] Su L, Hu S, Li S, Liang F, Gao J, Abdelzaher TF, et al. Quality of information based data selection and transmission in wireless sensor networks. In: RTSS; 2012. p. 327–38.
[32] Abdelgawad A, Bayoumi M. Resource-aware data fusion algorithms for wireless sensor networks. Lect Notes Electr Eng 2012;118:17–35.
[33] Sergiou C, Vassiliou V, Paphitis A. Hierarchical tree alternative path (HTAP) algorithm for congestion control in wireless sensor networks. Ad Hoc Netw 2013;11:257–72.
[34] Mao G, Fidan B, Anderson B. Wireless sensor network localization techniques. Comput Network 2007;51:2529–53.