A Hybrid Algorithm for Reliable and Energy-Efficient Data Gathering in Wireless Sensor Networks

Mohammed Falah Abood Salman [1] and Leili Farzinvash [1]

[1] Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran

Abstract: Reliability and energy efficiency are two important requirements of the data gathering process in wireless sensor networks. Accordingly, we propose a novel data gathering algorithm which meets these requirements. The proposed scheme categorizes the sensed data into valuable and regular data and handles each type of data based on its demands. The main requirement of valuable data is reliability. Thus, the adopted strategy to gather this type of data is to send several copies of data packets toward the sink. The rise of energy exhaustion in this scheme is tolerable. This is due to that, the valuable data is generated at a low rate. On the other hand, our main concern in gathering regular data is energy efficiency. As most of the sensed data is regular, an energy-efficient approach to gather regular data results in considerable energy conserving. Thus, we exploit clustering technique for regular data gathering. We also propose a lightweight intrusion detection system to detect malicious nodes. Simulation results and theoretical analysis confirm that our proposed algorithm provides reliability and energy efficiency at an acceptable level.

Keywords: wireless sensor network, data gathering, reliability, energy efficiency, clustering, intrusion detection system.

1. Introduction

Wireless sensor networks (WSNs) have been broadly employed in various applications such as wild environmental monitoring and military reconnaissance. Consisting of hundreds of tiny sensors, these networks offer an efficient and cost-effective solution for system monitoring. Each sensor node measures the environmental parameters and sends the collected data toward the sink in a multi-hop manner. The gathered data is processed for decision making and system control [1].

The proposed data aggregation algorithms for WSNs should fulfill the requirements of these networks especially reliability and energy efficiency. In these networks, packet lost may occur due to different reasons, such as collision, link error, battery exhaustion of sensors, permanent or temporal node failure, and the existence of malicious nodes. To ensure reliable monitoring, the amount of lost data should be diminished as much as possible. The existing strategies to increase reliability are redundancy injection and packet retransmission [2]. In the redundant-based solutions, some copies of the same packet are sent to the sink over distinct paths [3-10]. On the other hand, in the retransmission-based schemes, the sender node confirms packet delivery through receiving the acknowledgment [11-12].

The other important issue concerning data gathering in WSNs is energy efficiency. The sensors are low-cost nodes with limited energy. The energy of the sensor nodes exhausts quickly if a good energy management plan is not employed. This problem particularly arises for the nearby sensors to the sink because these nodes forward much more data in comparison to further ones. Considering the importance of energy-efficient data gathering, it has been extensively investigated in recent years [13-19].

The challenge in designing reliable and energy-efficient data aggregation algorithms is that these requirements are contradictory. In other words, increasing reliability necessitates more energy depletion. References [3-4] considered both of the mentioned criteria and balanced reliability against energy consumption. Felemban et al. [3] employed multipath routing and forwarded duplicated copies of the same packets to ensure reliability. To avoid overutilizing of energy resources, the number of paths are regulated based on the required level of reliability. The proposed algorithm in [4] guaranteed that the expected percentage of delivered data to the sink exceeds a pre-defined threshold. Using statistical reliability metric avoids excessive retransmissions, which conserves energy considerably.

The drawback of the abovementioned algorithms is that they did not adopt the collection strategy based on the type of the sensed data. The collected data by the sensor nodes can be categorized into valuable and regular data. In this context, regular data refers to the data that is collected in the normal condition. In this case, the environmental parameters change at a low rate. Losing some regular data does not impact on monitoring procedure severely. Therefore, it is acceptable to conserve more energy at the cost of reducing reliability. On the other hand, the sensed data in emergency circumstances, which is named as valuable data in this paper, has to be delivered to the sink for quick decision making and reaction. In such scenarios, the aim is to achieve high reliability, and energy efficiency is not a concern.

According to the above discussion, we propose a hybrid algorithm which follows different approaches to handle valuable and regular data. To transmit valuable data, the source node selects some trustworthy next-hop forwarders from its neighbors that are closer to the sink. Each forwarder receives a copy of the data packet from the source and transmits it toward the sink using a random path. In this way, several copies of the same packet are sent and the data is delivered to the sink with a high probability. As the rate of valuable data generation is low, the increase of exhausted energy due to multi-copy data transmission is not significant. In addition, the clustering technique is applied to collect regular data in an energy-efficient manner. The employed proactive strategy for reliability enhancement (i.e., transmitting several copies of valuable data) does not provide enough reliability. Therefore, we also include a lightweight intrusion detection system (IDS) in our design to detect the malicious nodes and increase reliability. The contributions of our algorithm can be summarized as the following:

- Balancing reliability against energy efficiency.
- Combining reactive and proactive strategies to enhance reliability.

The rest of this paper is organized as follows. Section 2 summarizes the related studies to our research. Section 3 describes the employed network model. In Section 4, we explain the proposed algorithm for reliable and energy-efficient data collection. Simulation results are presented in Section 5. Finally, we conclude the paper in Section 6.

2. Related Work

WSNs are mainly deployed to perform monitoring functions. The primary task in these networks is to collect and forward the sensed data toward the sink. As a result, efficient data gathering has been extensively investigated in the literature [3-20]. These studies considered different requirements of WSNs, such as reliability [3-12], energy efficiency [13-19], and security [20]. In this paper, reliability and energy efficiency requirements are covered. Therefore, in the following we consider the proposed schemes to meet these requirements.

Reliable data gathering refers to the procedure of recovering lost data packets. The well-known techniques for this purpose are to send redundant data [3-10] or retransmit the lost packets [11-12]. In the first approach, several copies of the packets are sent over different paths. The multipath construction problem was investigated in [3, 5-9]. Marina et al. [5] proposed on-demand multipath link disjoint distance vector algorithm based on AODV routing protocol. To fully separate the paths, the proposed scheme in [6] constructed node-disjoint paths. Using link and node-disjoint paths provides high-level reliability because a link or node failure at most impacts on a single path. This solution is not practical for sparse scenarios. Therefore, the braided multipath routing was introduced in which paths are partially disjoint [7-8]. Sun et al. [8] combined braided multipath routing with opportunistic routing. In this algorithm, each node selects its parents from the neighboring nodes that are closer to the sink. The number of parents is regulated based on link quality to achieve the requested level of reliability. Reference [10] constructed a fault-tolerant spanning tree over the sensors. To increase the reliability, each sensor chooses a primary parent and a number of backup parents. In the case of parent failure, the sensor adopts the best backup parent as the primary one.

Transmitting several copies of the packets causes significant overhead. To reduce this overhead, different coding techniques such as erasure coding and network coding have been integrated into multipath routing schemes. The main idea of these coding approaches is to divide the packet into m fragments and encode them to generate $m+k$ coded fragments. The packet can be recovered at the sink if at-least m coded fragments are received. References [21-22] utilized erasure and network coding, respectively. In these works, the fragments of a single packet are sent over different paths to increase the possibility of receiving adequate fragments to the sink. Ding et al. [23] applied multipath routing to the cluster-based WSNs. In this study, the clusters are considered as the basic units and data is transmitted from one cluster to the next-hop. To forward fragments in each cluster, the required number of sensors are activated per time slot in a probabilistic manner.

Redundancy injection is a proactive approach and may result in transmitting duplicated data to the sink. The retransmission strategy avoids unnecessary data transmission and energy depletion at the cost of high transmission delay. In this scheme, the source node waits for the acknowledgment of the sent packet to confirm its reception. The packet is retransmitted if its acknowledgment does not receive before the specified timeout. The retransmission procedure can be performed in a hop-by-hop or end-to-end basis [11-12]. The proposed algorithm in [11] ensured hop-by-hop reliability. To conserve energy, it considered event reliability where each node transmits one packet per event. The focus of [12] was to provide end-to-end reliability for WSNs similar to TCP protocol. The algorithm assumed that each flow can tolerate a specific lost rate. Accordingly, a lightweight congestion control mechanism is proposed to regulate data transmission rate of each flow such that its reliability requirement is satisfied.

To further improve reliability, some studies identified and removed malicious nodes from the WSN [24-29]. The main idea of these proposals was to determine one or more watchdogs per node to monitor its activity. The misbehavior of the node is determined according to the reports of its corresponding watchdogs. In the proposed algorithm in [24], some neighbors of each node are selected as its watchdogs. The watchdogs report the activities of misbehaving nodes to the sink. The network administrator identifies the malicious nodes according to the delivered reports and removes them from the network. A distributed malicious node detection system was proposed in [25]. In this work, the neighbors of each node served as its watchdogs. Each node monitors its neighbors and removes malicious ones from its neighboring list. Later studies used more sophisticated methods for malicious node detection. For example, the given algorithm in [26] applied the Dempster-Shafer theory to aggregate the evidence of different watchdogs. In this scheme, the nodes are organized into clusters. In addition, a number of watchdogs are allocated to the CHs and the relay nodes among the clusters. A fault-tolerant algorithm for WSN was proposed in [30]. In this work, the faulty nodes are replaced with sleeping ones to preserve connectivity and full coverage of the network.

The clustering technique, which is employed for regular data collection, has a great impact on diminishing energy exhaustion. This scheme has been deeply investigated in the literature [13-19]. In the proposed algorithms, each CH collects and aggregates the sensed data in its corresponding cluster. The aggregated data is sent to the sink directly [13-15] or in a multi-hop manner [16-19]. Some clustering algorithms in the latter category used only CHs for data forwarding [16]. The advantage of this approach is that the cluster members can be turned into sleep mode after sending data to the CH. On the other hand, the algorithms that selected forwarders from all available nodes have more options for choosing relay nodes [17-19].

3. Network Model

The assumed network model is presented in this section. The employed parameters in the proposed algorithm are described in Section 3.1. In addition, the model of energy consumption to send/receive data is explained in Section 3.2.

3.1 WSN Model

We consider a homogenous WSN with node set V and link set E. All nodes have the same transmission and interference range. Let e_i shows the remaining energy of node v_i. In addition, set N_i contains the neighbors of sensor node v_i. In the proposed algorithm, it is assumed that some nodes act maliciously. The percentage of the malicious nodes is shown by α. To identify the malicious nodes, each node computes the trust level of its neighbors. Let t_{ij} presents the trust of node v_i to node v_j. The overall trust of sink to v_i is shown by ts_i.

In the proposed algorithm, the data is categorized into regular and valuable classes. A sample data is assumed to be valuable if it differs considerably from the sensed data in the regular condition. In other words, a sample data is considered as valuable if it exceeds a predefined threshold or considerably differs from the previous sample. The amount of threshold depends on the application and is preloaded into the sensors. It is assumed that γ percentage of the sensed data is valuable on average. In the proposed scheme for valuable data collection, sensor v_i sends m copies of the valuable packets toward the sink. These copies are sent to m distinct nodes which are selected from NH_i. In this context, set NH_i consists of the neighboring nodes of v_i which are closer to the sink. In addition, the clustering scheme is employed to collect the regular data in an energy-efficient manner. The utilized notations in the proposed cluster-based scheme are as follows. The set of available clusters is presented by CL. The cluster number j and its CH are denoted by C_j and CH_j, respectively. Let CN_j presents the set of neighboring CHs of CH_j that are closer to the sink.

3.2 Energy Model

The amount of required energy for data transmission/reception is computed using the proposed model by Heinzelman [31]. According to this model, the energy consumption rate depends on the distance between transmitter and receiver. If this distance is less than the threshold d_0, the free space model is employed. Otherwise, the amount of consumed energy to transmit data packets is derived using multipath fading channel model. Accordingly, the required energy to transmit l bits of data is formally calculated as:

$$E_{Tx}(l,d_{tr})=\begin{cases} l\left(E_{elec}+\varepsilon_{fs}\,d_{tr}^{2}\right) & d_{tr}<d_0 \\ l\left(E_{elec}+\varepsilon_{mp}\,d_{tr}^{4}\right) & d_{tr}\geq d_0 \end{cases} \quad (1)$$

where d_{tr} and E_{elec} stand for the distance between the transmitter and receiver, and energy depletion of the electronic circuit, respectively. Parameters ε_{fs} and ε_{mp} denote the energy consumption of amplifier in the free space and multipath fading channel models, respectively. Moreover, the required energy to receive l bits of data is computed as:

$$E_{Rx}(l)=l\,E_{elec} \quad (2)$$

The list of utilized notations is given in Table 1.

Table 1. The employed notations in this paper

Notation	Definition
V	Set of the sensors
E	Set of the links among the sensors
e_i	Remaining energy of node v_i
N_i	Neighboring set of node v_i
α	Percentage of the malicious nodes
t_{ij}	Trust of node v_i to node v_j
ts_i	Current trust level of the sink to sensor v_i
γ	Percentage of valuable data
m	Number of replications of valuable data packets
NH_i	Set of the neighbors of v_i which are closer to the sink
CL	Set of the clusters
C_j	Cluster number j
CH_j	CH of C_j
CN_j	Set of neighboring CHs of CH_j that are closer to the sink
$\|.\|$	Size of a given set

4. The Proposed Algorithm

The proposed algorithm makes a tradeoff between reliability and energy efficiency. To this end, it follows different approaches for handling valuable and regular data. The regular data is gathered in an energy-efficient manner. On the other hand, the main concern in gathering the valuable data is reliability. The proposed approaches for handling these types of data are explained in Sections 4.1 and 4.2, respectively. Figure 1 depicts the proposed data gathering scheme. Section 4.3 justifies the importance of handling regular data in an energy-efficient manner. In Section 4.4, a lightweight IDS for detecting malicious nodes is given.

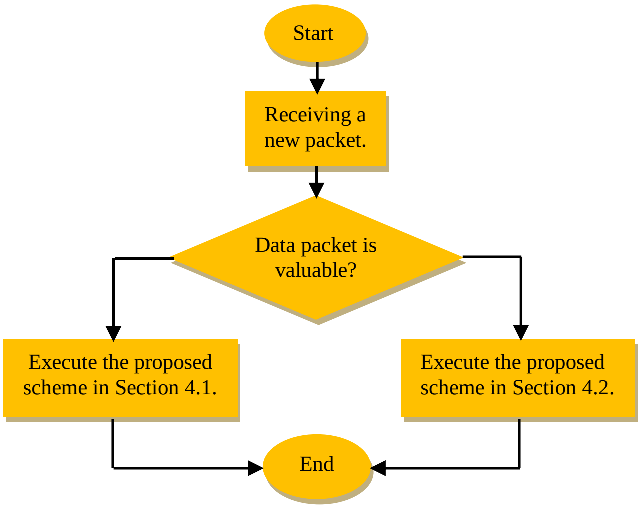
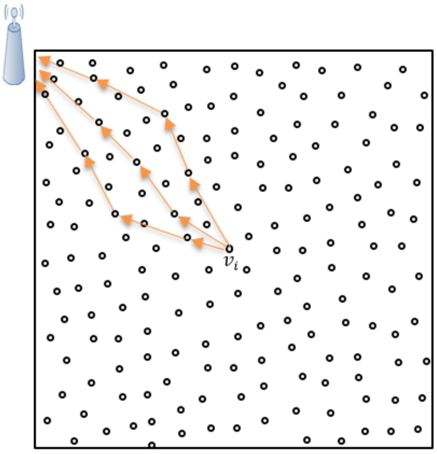

Start
Receiving a new packet.
Data packet is valuable?
Execute the proposed scheme in Section 4.1.
Execute the proposed scheme in Section 4.2.
End

Figure 1. Flowchart of data gathering

4.1 Valuable Data Gathering

The proposed scheme for valuable data gathering provides reliability at the expense of energy depletion. To send a valuable data packet, sensor v_i selects m of the most trustworthy nodes from NH_i. Each selected node forwards one copy of the packet toward the sink. Figure 2 illustrates the application of the proposed scheme, where m is set to 3. This figure shows the paths to transmit a given packet from node v_i to the sink. It is obvious that the next-hops may vary in different rounds, which results in various transmission paths.



Sink: , sensor node:○

Figure 2. Transmission of valuable data from v_i to the sink (m=3)

4.2 Regular Data Gathering

The proposed approach for gathering regular data exploits clustering scheme, in which nodes are divided into some clusters. To construct the clusters, we employ the LEACH algorithm [16]. This algorithm balances energy consumption through raising the possibility of becoming CH for all sensor nodes. For this purpose, each node v_i is selected as CH in a probabilistic manner. The node chooses a random number within the range of [0-1]. If this number is less than the given threshold in (3), sensor node v_i becomes CH in the next round.

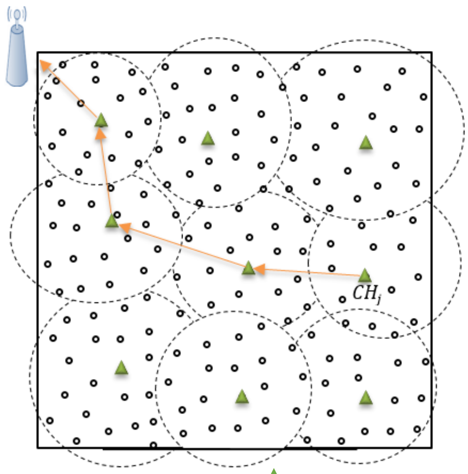
$$Th(v_i)=\begin{cases}\dfrac{|CL|}{|V|-|CL|\left(r \bmod \dfrac{|V|}{|CL|}\right)} & \forall v_i \in NCH \\ 0 & o.w.\end{cases} \quad (3)$$

where $Th(v_i)$ is the derived threshold for node v_i, and r refers to the round number. In addition, NCH indicates the sensor nodes that are not chosen as CH in the last $|V|/|CL|$ rounds.

In the proposed scheme, the aggregated data is forwarded toward the sink using a multi-hop manner. More specifically, CH_j forwards data to one of the neighboring CHs in CN_j. Figure 3 presents the overall scheme of regular data gathering in the proposed algorithm.

To have an effective data gathering scheme, the next-hop of CH_j should be determined properly. Two main measures to select CH_k as the next-hop are listed below:

- Energy efficiency: Comprising the required energy to transmit data to CH_k, and the residual energy of this node (i.e., e_k).



CH: ▲

Figure 3. Transmission of regular data from cluster C_j to the sink

- Reliability: The trust level of CH_j to CH_k, namely t_{jk}.

Using these criteria, the preference of CH_k to be selected as the next-hop of CH_j, namely nh_{jk}, is defined as:

$$nh_{jk}=\frac{e_k\,t_{jk}}{E_{Tx}\left(k,d_{jk}\right)} \quad (4)$$

The probability of selecting CH_k as the next-hop of CH_j, which is denoted by ph_{jk}, is computed as:

$$ph_{jk}=\frac{nh_{jk}}{\sum_{CH_l\in CN_j} nh_{jl}} \quad (5)$$

4.3 Energy Exhaustion Analysis

This section devotes to investigating the energy consumption of the proposed scheme. To this end, we consider three scenarios as follows:

- Case 1: All data is gathered using the proposed approach in Section 4.1. This scheme brings about high reliability and energy consumption. In this case, the required energy to send a l-bit message from sensor v_i to the sink, namely $E1_i$, is delivered as:

$$E1_i=m\sum_{(v_p,v_q)\in path_i}\left(E_{Tx}\left(l,d_{pq}\right)+E_{Rx}(l)\right) \quad (6)$$

where $path_i$ presents the data transmission path from v_i to the sink. For the ease of comparison, we assume that the distances of the successive nodes on $path_i$ are the same, which is denoted by d. Accordingly, (6) is rewritten as:

$$E1_i=ml\left|path_i\right|\left(E_{Tx}(1,d)+E_{Rx}(1)\right) \quad (7)$$

- Case 2: All data is gathered using the proposed approach in Section 4.2. This strategy requires less energy in comparison to the first case at the expense of reliability. The consumed energy to deliver a l-bit message from sensor v_i to the sink is computed as:

$$E2_i = \frac{1}{|C_j|}\left(\begin{array}{l} \sum\limits_{v_i \in C_j} \left(E_{Tx}\left(l, d_{ij}\right) + E_{Rx}\left(l\right)\right) + \\ \sum\limits_{\substack{(CH_p, CH_q) \\ \in path_j}} \left(E_{Tx}\left(g_j, d_{pq}\right) + E_{Rx}\left(g_j\right)\right) \end{array} \right) \quad (8)$$

where $E2_i$ represents the amount of consumed energy, and g_j stands for the size of the aggregated packet by CH_j. In (8), the first term presents the required energy for data gathering within C_j, and the latter term denotes the consumed energy to transmit the aggregated data toward the sink. To simplify (8), we assume that the distances among all successive nodes on $path_j$ and among cluster members and CH_j are equal to d. In addition, g_j can be computed as [32]:

$$g_j = \frac{l|C_j|}{|C_j|cf - cf + 1} \quad (9)$$

where cf is the correlation factor. According to the above discussion, (8) can be reformulated as:

$$\begin{aligned} E2_i &= \left(E_{Tx}\left(l, d\right) + E_{Rx}\left(l\right)\right) + \\ &\quad \frac{|path_j|}{|C_j|}\left(E_{Tx}\left(g_j, d\right) + E_{Rx}\left(g_j\right)\right) \qquad (10) \\ &= l\left(1 + \frac{|path_j|}{|C_j|cf - cf + 1}\right)\left(E_{Tx}\left(1, d\right) + E_{Rx}\left(1\right)\right) \end{aligned}$$

- Case 3: The proposed algorithm, which uses the given approaches in Sections 4.1 and 4.2 for handling valuable and regular data, respectively. Assuming γ percentage of the sensed data is valuable, the exhausted energy to transmit an l-bit message from node v_i to the sink, namely $E3_i$, is stated as:

$$\begin{aligned} E3_i &= \gamma l m |path_i|\left(E_{Tx}\left(1, d\right) + E_{Rx}\left(1\right)\right) + \\ &\quad (1-\gamma)k\left(1 + \frac{|path_j|}{|C_j|cf - cf + 1}\right)\left(E_{Tx}\left(1, d\right) + E_{Rx}\left(1\right)\right) \qquad (11) \\ &= l\left(\gamma m |path_i| + (1-\gamma)\left(1 + \frac{|path_j|}{|C_j|cf - cf + 1}\right)\right) \\ &\quad \left(E_{Tx}\left(1, d\right) + E_{Rx}\left(1\right)\right) \end{aligned}$$

Equations (7), (10)-(11) are evaluated under γ=0.1, cf=0.1, m=3, $|path_i|$=3, $|path_j|$=2, and $|C_j|$=20 as:

$$E1_i = 9l\left(E_{Tx}\left(1, d\right) + E_{Rx}\left(1\right)\right) \quad (12)$$

$$E2_i = 1.7l\left(E_{Tx}\left(1, d\right) + E_{Rx}\left(1\right)\right) \quad (13)$$

$$E3_i = 2.4l\left(E_{Tx}\left(1, d\right) + E_{Rx}\left(1\right)\right) \quad (14)$$

From (12)-(14), we can see that the first scenario needs a high amount of energy for data transmission. On the other hand, the last scenario, i.e., our proposed approach, makes a tradeoff between reliability and energy exhaustion.

4.4 IDS

Our proposed algorithm enhances the reliability of valuable data gathering through sending m copies of valuable packets to the sink. In the following, we investigate the impact of m on the probability of valuable packet delivery. For this purpose, we define the delivery probability of a generated packet by node v_i, namely pe_i, as:

$$pe_i = 1 - \left(1 - \left(1 - \alpha\right)^{|path_i|}\right)^m \quad (15)$$

This equation is derived as follows. The term (1-α)^|$path_i$| presents the probability that the generated valuable packet by sensor v_i over a given path is delivered to the sink. Accordingly, (1-(1-α)^|$path_i$|)^m denotes the probability of packet forwarding failure over all paths.

Figure 4 presents the impact of m on pe_i. In this figure, |$path_i$| is assumed to be 3, which is equal to average hop count distance from the sensor nodes to the sink. It is derived from this figure that pe_i decreases rapidly by raising m. Parameter pe_i drops at a higher rate in the WNSs with more malicious nodes. For example, it drops from 0.92 to 0.86 when α is increased from 0.1 to 0.2 under m=2. From this figure, we can derive that the suitable amount of m is equal to 3 when α is less than 0.2.

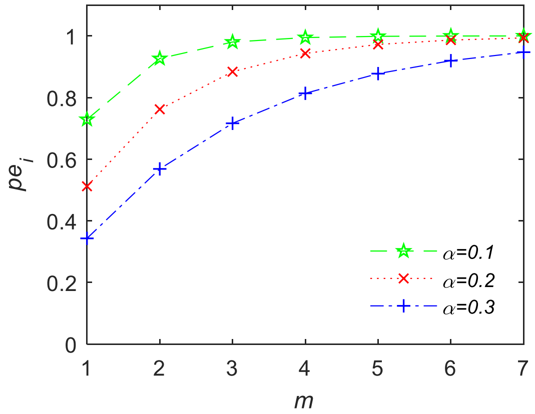


Figure 4. Achieved reliability for valuable data transmission versus different values of m

The main problem is to enhance the probability of regular data delivery. The probability of packet dropping in this case, namely pn_i, is computed as:

$$pn_i = \left(1 - \alpha\right)^{|path_j| + 1} \quad (16)$$

From (16), it is obvious that pn_i is exponentially depended on |$path_j$|. Therefore, the regular data delivery rate becomes very low.

To tackle this problem, we exploit a low overhead IDS to identify malicious nodes. In the proposed scheme, each cluster member in C_j monitors its corresponding CH_j and reports malicious activities of the CH to the sink. The sink decides about maliciousness of CH_j based on the submitted reports by the members of C_j. The malicious activities of CH_j can be categorized as follows:

- Incorrect data aggregation: We use the proposed aggregation model in [16], where the CH omits duplicated sensed data throughout the cluster. Each cluster member v_i ∈ C_j monitors the sent data by CH_j to find its own reported data. The elimination of this report from the aggregated data is considered as misbehavior.
- Packet dropping: CH_j is responsible for forwarding the data of further clusters to the sink. The transmitted data to CH_j is also heard by a number of member nodes in C_i. These

nodes monitor data forwarding by CH_j to find out the percentage of dropped data packets.

According to the above discussion, the trust of node v_i to node CH_j in round r, namely $t_{ij}(r)$, is calculated as:

$$t_{ij}(r) = w_f \frac{FP_j(r)}{RP_j(r)} + (1 - w_f)\frac{TG_{ij}(r)}{AG_j(r)} \qquad (17)$$

where $FP_j(r)$ and $RP_j(r)$ present the number of forwarded and received packets by CH_j in round r, respectively. Variable $TG_{ij}(r)$ shows the number of sent packets by member node v_i in round r that is included in the aggregated data by CH_j. In addition, the number of times that CH_j performs aggregation in round r is presented by $AG_j(r)$. Parameter w_f denotes the importance of data forwarding against data aggregation.

The sensor node v_i also monitors its non-CH neighboring nodes to ensure their trustworthiness. To this end, node v_i keeps track of the number of forwarded valuable data packets by sensor node v_j and computes $t_{ij}(r)$ as:

$$t_{ij}(r) = \frac{FP_j(r)}{RP_j(r)} \qquad (18)$$

At the end of the round, node v_i computes t_{ij} for each node $v_j \in N_i$ as:

$$t_{ij} = \beta t_{ij}(r) + (1 - \beta) t_{ij} \qquad (19)$$

where β is a constant parameter that determines the importance of trust level to node v_j in the current round.

Each node sends its trust value to the neighboring nodes to the sink. For each senor node v_i, the sink computes ts_i as follows:

$$ts_i(r) = \frac{\sum_{v_j \in N_i} t_{ij}(r)}{|N_i|} \qquad (20)$$

$$ts_i = \beta ts_i(r) + (1 - \beta) ts_i \qquad (21)$$

where $ts_i(r)$ is the trust value of node v_i in round r. After computing the trust values of the nodes, the sink considers a given node v_i as malicious if its trust level, i.e., ts_i, is less than the predefined threshold.

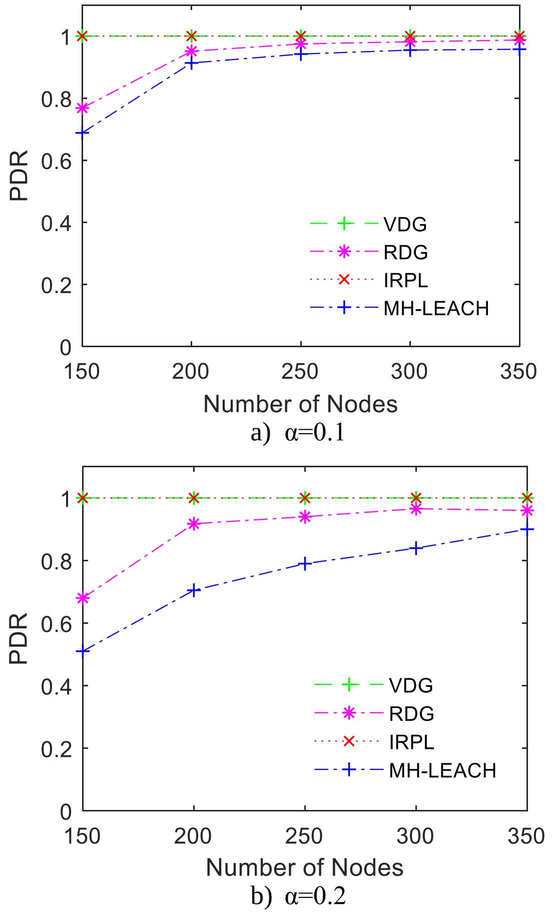
5. Performance Analysis

In this section, we evaluate the effectiveness of the proposed algorithm. For this purpose, we compare it with MH-LEACH [16] and IRPL [9], which follow clustering and multipath routing schemes, respectively. The considered algorithms are implemented using OMNET++ [33]. The employed criteria for comparison are the packet delivery ratio (PDR) and total energy exhaustion to forward the generated data packets toward the sink, which are measured after 300 seconds.

In the performed simulations, the dimensions of the sensor field are assumed to be 500m×500m. The sink is fixed at the upper left corner of the sensor filed. The sensors are located in random positions, where their number is varied within the range of [150-350]. The initial energy of each node is assumed to be 0.5J. To investigate the impact of the percentage of malicious nodes on the performance, α is varied within the range of [0.1-0.3]. In addition, γ is set to 10%. Finally, m is fixed at 3 to achieve satisfactory reliability.

5.1 Impact of the number of sensors

In the performed experiments in this section, we compare the resultant PDR and energy consumption by the considered algorithms versus different number of sensor nodes. To this end, the number of sensors is varied from 150 to 350. Figure 5 depicts the derived PDR using the considered algorithms by increasing the number of nodes. As regular and valuable data have different characteristics, these data types are evaluated separately in Figure 5 and the other figures. For the ease of explanation, in the following the valuable and regular data gathering approaches are shown by VDG and RDG, respectively.



a) α=0.1

b) α=0.2

Figure 5. Achieved PDR using different number of nodes

It is derived from Figure 5 that both VDG and RDG improve PDR in comparison to MH-LEACH using different number of nodes. The superiority of RDG is due to employing the IDS to detect malicious nodes. VDG also exploits multi-copy data forwarding and therefore, its PDR is near to optimum. As it is shown in the figure, the performance gap among RDG and MH-LEACH increases by raising α. For example, the difference between achieved PDR by MH-LEACH and RDG increases from 5% to 20% when α is raised from 0.1 to 0.2 for 200-node WSNs. This outcome highlights the importance of the IDS, which omits malicious nodes from the network. The other point is that the obtained PDR by RDG and MH-LEACH increases with raising the number of sensors. This is due to that, using more sensors increases the node density. Therefore, each CH has more choices for selecting the next-hop CH toward the sink, which reduces the probability of packet dropping.

The energy consumption measure is considered in Figure 6, where α is set to 0.1. It is derived from this figure that IRPL exhausts the highest amount of energy among the considered algorithms. This is due to that this algorithm applies multi-copy data forwarding to all data packets, either regular or valuable. On the other hand, our algorithm uses multi-copy data forwarding for only valuable data. Therefore, energy consumption is reduced considerably. From this figure, we can see that the gap between RDG and MH-LEACH is acceptable. The amount of consumed energy by the proposed algorithm is 28% higher than that of MH-LEACH on average. These results justify that we achieve our design goals, comprising reliability and energy efficiency.

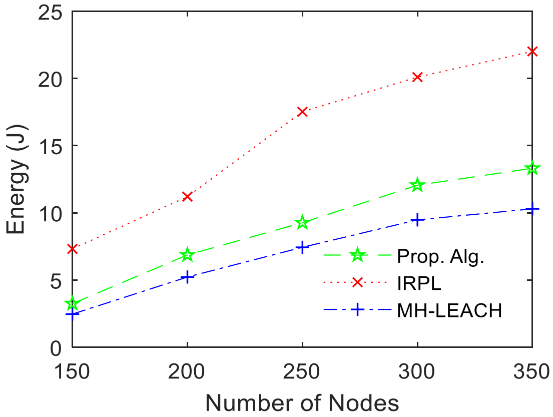


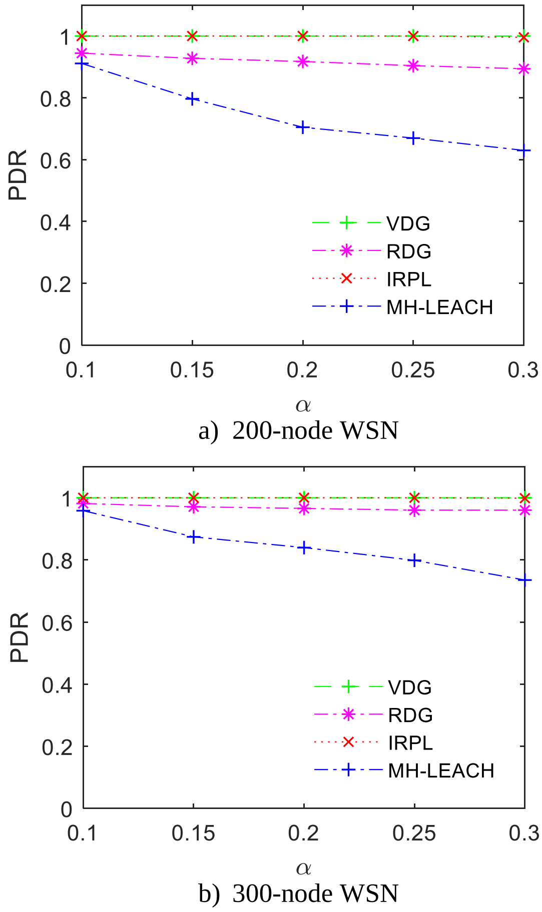
Figure 6. Consumed energy versus different number of sensor nodes

5.2 Impact of α on the Reliability

Parameter α has a great impact on network reliability. Therefore, in this section we study the impact of this parameter on the achieved PDR by the considered algorithms. Figure 7 illustrates the resultant PDR of the considered algorithms against different amounts of α. According to the given results in this figure, the achieved PDR by VDG and IRPL are not seriously affected by increasing α. On the other hand, RGD and MH-LEACH are sensitive to the variations of α. In these algorithms, the PDR is diminished by raising the number of malicious nodes. However, RDG is more robust against misbehaving activities compared to MH-LEACH. In this scheme, the PDR is dropped to 92% in 200-node WSNs on average. On the other hand, in MH-LEACH 26% of the generated packets are dropped on average. The superiority of RDG is due to that, the employed IDS in our design detects malicious nodes. Therefore, they are not selected as CH and cannot drop the generated packets.

6. Conclusion

This paper dealt with the problem of reliable and energy-efficient data gathering in WSNs. The proposed algorithm offered different strategies for gathering valuable and regular data. The valuable data is gathered reliably. To this end, the sender node constructs several copies of the valuable data packets and sends them to the sink over trustworthy paths. The disadvantage of this scheme is its high energy consumption. However, its overhead is reasonable because the rate of generating valuable data is very low. In addition, the regular data is gathered using the clustering scheme. Using this technique diminishes energy exhaustion considerably. On the other hand, it does not take into account the reliability criterion. As losing some regular data is tolerable, this shortcoming does not lead to a severe problem. We also included a lightweight IDS in our design to discover malicious nodes. Simulation results validated the effectiveness of the proposed approach in comparison to the existing solutions.



a) 200-node WSN

b) 300-node WSN

Figure 7. Achieved PDR versus different values of α

References

[1] J. Yick, B. Mukherjee, D. Ghosal, "Wireless sensor network survey," Computer Networks, Vol. 52, No. 12, pp. 2292-2330, 2008.

[2] M.A. Mahmood, W.K.G. Seah, I. Welch, "Reliability in wireless sensor networks: A survey and challenges ahead," Computer Networks, Vol. 79, pp. 166-187, 2015.

[3] E. Felemban, C-G Lee, E. Ekici, "MMSPEED: Multipath multi-speed protocol for QoS guarantee of reliability and timeliness in wireless sensor networks," IEEE Transactions on Mobile Computing, Vol. 5, No. 6, pp. 738-754, 2006.

[4] T. Le, W. Hu, P. Corke, S. Jha, "ERTP: Energy-efficient and reliable transport protocol for data streaming in wireless sensor networks," Computer Communications, Vol. 32, pp. 1154-1171, 2009.

[5] M.K. Marina, S.R. Das, "Adhoc on-demand multipath distance vector routing," Wireless Communications and Mobile Computing, Vol. 6, No. 7, pp. 969-988, 2006.

[6] S. Chakraborty, S. Chakraborty, S. Nandi, S. Karmakar, "Fault resilience in sensor networks: Distributed node-disjoint multi-path multi-sink forwarding," Journal of Network and Computer Applications, Vol. 57, pp. 85-101, 2015.

[7] A. Attir, Y. Challal, A. Hadjidj, A. Bouabdallah, "Braided disjoint branch routing protocol for WSNs," in: Proc. 8th IEEE International Conference on Broadband and Wireless

Computing, Communication and Applications (BWCCA'13), Compiegne, France, pp. 106-113, 2013.

[8] X. Sun, H. Chen, X. Wu, X. Yin, W. Song, "Opportunistic communications based on distributed width-controllable braided multipath routing in wireless sensor networks," Ad Hoc Networks, Vol. 36, pp. 349-367, 2016.

[9] Z. Wang, L. Zhang, Z. Zheng, J. Wang, "Energy balancing RPL protocol with multipath for wireless sensor networks," Peer-to-Peer Networking and Applications, Vol. 11, No. 5, pp. 1085-1100, 2018.

[10] K. Suganthi, B. Vinayagasundaram, J. Aarthi, "Randomized fault-tolerant virtual backbone tree to improve the lifetime of wireless sensor networks," Computers & Electrical Engineering, Vol. 48, pp. 286-297, 2015.

[11] M.A. Mahmood, W.K.G. Seah, "Event reliability in wireless sensor networks," in: Proc. 7th IEEE International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP'11), Adelaide, SA, Australia, 2011.

[12] Y. Xue, B. Ramamurthy, Y. Wang, "LTRES: A loss-tolerant reliable event sensing protocol for wireless sensor networks," Computer Communications, Vol. 32, No. 15, pp. 1666-1676, 2009.

[13] X-X. Ding, M. Ling, Z-J. Wang, F-L. Song, "DK-LEACH: An optimized cluster structure routing method based on LEACH in wireless sensor networks," Wireless Personal Communications, Vol. 96, No. 4, pp. 6369-6379, 2017.

[14] M.O. Oladimeji, M. Turkey, S. Dudley, "HACH: Heuristic algorithm for clustering hierarchy protocol in wireless sensor networks," Applied Soft Computing, Vol. 55, pp. 452-461, 2017.

[15] P.C. Srinivasa Rao, P.K. Jana, H. Banka, "A particle swarm optimization based energy efficient cluster head selection algorithm for wireless sensor networks," Wireless Networks, Vol. 23, No. 7, pp. 2005-2020, 2017.

[16] E. Alnawafa, I. Marghescu, "MHT: Multi-hop technique for the improvement of leach protocol," in: Proc. 15th IEEE RoEduNet Conference: Networking in Education and Research, Bucharest, Romania, 2016.

[17] T. Bhatia, S. Kansal, S. Goel, A.K. Verma, "A genetic algorithm based distance-aware routing protocol for wireless sensor networks," Computers & Electrical Engineering, Vol. 56, pp. 441-455, 2016.

[18] R.S.Y. Elhabyan, M.C.E. Yagoub, "Two-tier particle swarm optimization protocol for clustering and routing in wireless sensor network," Journal of Network and Computer Applications, Vol. 52, pp. 116-128, 2015.

[19] A. Bradai, K.D. Singh, A. Rachedi, T. Ahmed, "EMCOS: Energy-efficient mechanism for multimedia streaming over cognitive radio sensor networks," Pervasive and Mobile Computing, Vol. 22, pp. 16-32, 2015.

[20] B.P. Laxmi, A. Chilambuchelvan, "GSR: Geographic secured routing using SHA-3 algorithm for node and message authentication in wireless sensor networks," Future Generation Computer Systems, Vol. 76, pp. 98-105, 2017.

[21] S. Ali, A. Fakoorian, H. Taheri, "Optimum Reed-Solomon erasure coding in fault tolerant sensor networks," in: Proc. 4th IEEE International Symposium on Wireless Communication Systems (ISWCS'07), Trondheim, Norway, 2007.

[22] B. Sun, C. Gui, Y. Song, H. Chen, "Novel network coding and multi-path routing approach for wireless sensor network," Wireless Personal Communications, Vol. 77, pp. 87-99, 2014.

[23] X. Ding, X. Sun, C. Huang, X. Wu, "Cluster-level based link redundancy with network coding in duty cycled relay wireless sensor networks," Computer Networks, Vol. 99, pp. 15-36, 2016.

[24] H. Nabizadeh, M. Abbaspour, "IFRP: an intrusion/fault tolerant routing protocol for increasing resiliency and reliability in wireless sensor networks," International Journal of Ad Hoc and Ubiquitous Computing, Vol. 14, No. 1, pp. 52-69, 2013.

[25] L. Sánchez-Casado, G. Maciá-Fernández, P. García-Teodoro, R. Magán-Carrión, "A model of data forwarding in MANETs for lightweight detection of malicious packet dropping," Computer Networks, Vol. 87, pp. 44-58, 2015.

[26] O.A. Wahab, H. Otrok, A. Mourad, "A cooperative watchdog model based on Dempster-Shafer for detecting misbehaving vehicles," Computer Communications, Vol. 41, pp. 43-54, 2014.

[27] J.A.F.F. Dias, J.J.P.C. Rodrigues, F. Xia, C.X. Mavromoustakis, "A cooperative watchdog system to detect misbehavior nodes in vehicular delay-tolerant networks," IEEE Transactions on Industrial Electronics, Vol. 62, No. 12, pp. 7929-7937, 2015.

[28] A.M. Shabut, K. Dahal, I. Awan, "Friendship based trust model to secure routing protocols in mobile ad hoc networks," in: Proc. IEEE International Conference on Future Internet of Things and Cloud (FiCloud'14), Barcelona, Spain, 2014.

[29] H. Xia, J. Yu, C-L. Tian, Z-K. Pan, E. Sha, "Light-weight trust-enhanced on-demand multi-path routing in mobile ad hoc networks," Journal of Network and Computer Applications, Vol. 62, pp. 112-127, 2016.

[30] C. Titouna, M. Gueroui, M. Aliouat, A. Adamou Abba Ari, A. Amine, "Distributed fault-tolerant algorithm for wireless sensor network," International Journal of Communication Networks and Information Security, Vol. 9, No. 2, 2017.

[31] W.B. Heinzelman, A.P. Chandrakasan, H. Balakrishnan, "Application specific protocol architecture for wireless microsensor networks," IEEE Transactions on Wireless Communications, Vol. 1, pp. 660-670, 2002.

[32] G. Sudha A. Kumar, G. Manimaran, Z. Wang, "End-to-end energy management in networked real-time embedded systems," IEEE Transactions on Parallel and Distributed Computing, Vol. 19, No. 11, pp. 1498-1510, 2008.

[33] OMNET++, https://www.omnetpp.org.