Mobile Networks and Applications (2019) 24:1295–1305
https://doi.org/10.1007/s11036-018-1169-x

An Energy-Efficient and Congestion Control Data-Driven Approach for Cluster-Based Sensor Network

Syed Rooh Ullah Jan[1] · Mian Ahmad Jan[1] · Rahim Khan[1] · Hakeem Ullah[2] · Muhammad Alam[3] · Muhammad Usman[4]

Published online: 10 November 2018

Abstract

In Wireless Sensor Networks (WSNs), a dense deployment of sensor nodes produce data that contain intra-temporal and inter-spatial correlation. To reduce the intensity of correlation, we propose in-node data aggregation technique that eliminates redundancy in the sensed data in an energy-efficient manner. A novel data-driven approach is adopted to perform in-node data aggregation using an underlying cluster-based hierarchical network. Our proposed approach partially processes the data at each member node and forwards a fraction of the actual data, i.e., fused data, towards the cluster head. At each member node, the raw captured data is categorized into various classes, i.e., stratum. Each member node continuously senses the environment for temperature readings and compares them with the mean values of various strata. If the temperature reading is less than or greater than the mean value, it is compared with the existing Min/Max of that particular stratum. If in case, the new reading is less than/greater than the Min/Max of a particular stratum, it replaces these values, accordingly. Our proposed approach is computationally lightweight, energy-efficient and reduces the degree of correlation among the resource-constrained sensor nodes. As as a result, communication cost, packet collision and network congestion are reduced and the network lifetime is enhanced. The analytical results prove the validity and effectiveness of our proposed approach.

Keywords Wireless sensor networks · Stratified sampling · Data-driven approaches · In-network processing · Energy-efficiency · Congestion · Data redundancy

1 Introduction

Wireless Sensor Networks (WSNs) have felt their presence in almost every domain of daily life. The building blocks of these networks are densely-deployed miniature sensor nodes that are capable of sensing, processing, communication and coordination among each others to monitor a particular physical phenomenon. Each node senses the environment and processes any captured data, locally. Upon in-network data aggregation, the processed information is forwarded via gateways to the base station for further analysis, decision making and storage [1, 4]. The performance of these miniature and resource-constrained nodes is adversely affected due to in-node operations, i.e., communication, processing and sensing. Among these operations, data transmission consumes a much higher energy. Therefore, novel and energy-efficient in-node data proccessing techniques are required to reduce the communication cost of the network [2]. The limited energy issue poses a plethora of challenges for data aggregation that prevent the complete exploration of these networks. Moreover, the energy-constrained nature of the nodes are further worsen by the fact that they are deployed in a harsh, unattended, and sometimes human-inaccessible terrains [3]. Replacing or recharging their batteries is a cumbersome task and sometimes impossible. Although, various energy conservation and scavenging techniques are proposed in literature, data aggregation still remains an open research area due to the resource-starving nature of these networks.

Among the existing protocols of WSNs, cluster-based hierarchical routing protocols are the most efficient ones in

Strata is the plural of stratum

✉ Mian Ahmad Jan
mianjan@awkum.edu.pk

✉ Rahim Khan
rahimkhan@awkum.edu.pk

Extended author information available on the last page of the article.

term of data aggregation [22, 27, 29, 40]. These protocols partition the whole network into several clusters. Each cluster has a designated cluster head that is responsible to manage the nodes, i.e., member nodes, within that particular cluster. Each cluster head collects data from its member nodes within a cluster and broadcasts to the base station. These protocols have been widely studied by the research community to satisfy the scalability objective and achieve a prolonged network lifetime for WSNs. In cluster-based hierarchical WSNs, in-network data aggregation partially offloads the computation from cluster heads [23]. The member nodes locally process the data by using various data aggregation and fusion techniques. The use of in-network processing eliminates the need of transmitting raw redundant data to the base station via the cluster heads. Although, there exists numerous studies in cluster-based hierarchical WSNs for elimination of redundancy using various data aggregation techniques, the latter still remains an open research area. This is mainly due to the absence of computationally lightweight and energy conserving techniques in view of limited resources of the nodes.

In cluster-based hierarchical WSNs, energy conservation can be achieved using two different approaches, i.e., a low-duty cycling approach [5, 34] and a data-driven approach [8, 36]. The latter is achieved using in-network, also known as in-node processing. In this paper, we limit our discussion to data-driven approaches by proposing a novel and energy-efficient in-node data processing for a cluster-based hierarchical WSN [11]. Initially, at each member node, the real-time sensed data is divided into subgroups called stratum, similar to stratified sampling [15]. The startum that we have designed are based on the previous historical and archival data captured by [25]. Each member node constantly senses the environment for temperature readings and compare with a given stratum. In our proposed work, there are seven stratum, each one having a particular range for temperature readings. The proposed approach has two distinct but interrelated phases.

1. Comparison with the mean value of a particular stratum.
2. The new minimum (Min) and maximum (Max) values, obtained from the previous phase, are compared with the old values that already exists within that particular stratum in order to choose the Min/Max for that stratum.

During the first phase, upon sensing a new temperature reading, the sensor node compares it with the mean of a particular stratum to confirm whether the newly sensed value is within the Min/Max range. If this value is equal to the mean, it is simply discarded. If it is less than/greater than the mean value, the second phase initiates. In case, the new value is less than the mean value, a comparison is made with the Min of that stratum. If the new value is greater than the Min, it is discarded. However, if it is less than the Min, the new value replaces the existing Min and becomes the new Min. A similar comparison is made for the Max as well. If the new reading is less than the Max, it is discarded, otherwise, it becomes the new Max of that particular stratum. This process is repeated each time for a specific stratum whenever a new value is sensed. As a result, each stratum stores only two values, i.e., Min/Max, and are forwarded by each node to its respective cluster head after a certain period of time. Finally, data redundancy is examined at the cluster head level for each packet received from a member node. At this stage, the data of member nodes in close proximity to each other are compared. This eliminates the burden on cluster heads to compare the data received from all the member nodes. As a result, the energy consumption of the cluster heads is significantly reduced that extends the life time of the underlying network.

The main benefit of our proposed approach is that it performs in-node processing cost whenever possible, to conserve the communication cost. Our proposed approach significantly reduces the volume of data forwarded to the base station that in turn, reduces the network congestion [33] The main contributions of this paper are as follow.

1. We have proposed an energy-efficient and computationally lightweight in-node processing and aggregation technique. The design of our proposed method is simple to match the resource-constrained nature of sensor nodes.
2. The system model of our proposed method is validated analytically using various mathematical expressions. It achieves a higher quality of service (QoS) by controlling congestion to a greater extent and prolongs the network lifetime by performing energy-efficient in-node processing.

The rest of this paper is organized as follow. In Section 2, we provide the literature review of various approaches used for data aggregation in cluster-based hierarchical WSNs. In Section 3, we provide the network model of our proposed scheme. In Section 4, our proposed model for stratified sampling is discussed and its analytical analysis is provided. In Section 5, we provide the experimental results of our proposed scheme. Finally, the paper is concluded and future research directions are provided in Section 6.

2 Literature review

In WSNs, energy conservation is accomplished using duty cycling and data-driven approaches [9, 41, 42]. In duty cycling, the nodes switch between active and sleep modes, whenever required. Thus, energy of the nodes is preserved during the sleep mode [9]. In this paper, we restrict our discussion to data-driven approaches, only. These approaches utilize the storage and processing capabilities of the nodes by performing in-network processing that saves valuable energy of the nodes [42]. The use of in-network processing lowers the number of packets transmitted across the network by partially processing the data locally at the sensor nodes, prior to base station's transmission. Instead of sending raw data to the base station, a node exploits the correlation between the data packets and transmits a subset of data packets. Such transmission reduces data redundancy by eliminating the duplicate packets. In-network processing can be accomplished using various techniques such as data compression and data aggregation [39]. In data compression, the original input data stream is converted into a compressed stream [21, 24]. Data compression aims at exploring the compact representation of non-random data by reducing the number of packets forwarded towards the base station [35]. In WSNs, compression can be classified as sampling compression [19]. data compression [14, 37] and communication compression [32, 38]. Data aggregation, on the other hand, refers to in-network aggregation of raw data and routing them via multi-hop links. The gathered data is processed at the intermediate nodes, i.e., by fusing together sensor readings related to a similar event, in order to reduce the energy consumption and prolong network lifetime [13].

One the most fundamental and challenging issue in WSNs is data redundancy that exists in the sensor readings gathered by the nodes. This highly correlated data consume substantial amount of network's energy and causes congestion during transmission across the network. Moreover, if the data with such redundancy reaches the decision center, e.g., the base station, it can adversely affect the analysis and decision making process. Hence, it is a challenging issue that needs to be addressed to achieve efficient data aggregation. In literature, there exists various studies that use data aggregation to filter out redundant data in WSNs [28, 30]. In [17], the authors proposed an efficient data aggregation technique for cluster-based periodic wireless sensor networks (CPWSNs). Data aggregation is performed in two steps. Initially, the data is partially aggregated at the member nodes, i.e., local data aggregation, before forwarding it to their respective cluster heads. Next, repetition in the data set is eliminated at each cluster head using three methods, i.e., the jaccard function, one-way anova model and Bartlett test, and euclidean and cosine method. These methods are used to check similarities, conditional variance and dissimilarities between the data sets. Although, the proposed scheme reduces data redundancy, it does not take into account the complexity at the member nodes.

A dynamical message list based data aggregation (DMLDA) approach for cluster-based WSNs was proposed in [12]. DMLDA aimed at designing a computationally lightweight and high filtering efficient scheme in a distributed network architecture. The proposed scheme defined a special data structure at each aggregating node, i.e., cluster heads, which contains the history of data prior to its transmission. This data structure is used for data aggregation to filter out irrelevant and repeated data. The proposed scheme achieves data aggregation at cluster heads only and as such, ignores its importance at the basic level, i.e., at the member nodes. As a result, the cluster heads drain their energy much earlier. Moreover, the proposed scheme does not take into account the number of packets traversing the network. A cluster-based data aggregation technique was proposed in [16]. The proposed scheme eliminates data redundancy among the data sets with a minimum data latency. To achieve these goals, the proposed approach integrates K-Means algorithm at the first stage and a prefix frequency filtering (PFF) at the second stage. During the first stage, K-Means algorithm is employed that grouped together similar data sets into clusters. Next, a prefix frequency filtering (PFF) technique is employed at each cluster of the data set to reduce the number of comparisons being made by lowering the data latency. The simulation results show the validity of the proposed approach, however, it lacks energy evaluation and number of packet reduction during aggregation. Moreover, K-Means algorithm needs be applied on the prefixed as opposed to the whole data set to increase accuracy of the proposed approach.

To overcome the drawbacks of [16], a novel prefix-based data aggregation approach was proposed for a cluster-based network architecture [6]. Unlike the existing data aggregation schemes that rely on computing similarity values, the proposed approach filters data based on the quality of information. The proposed approach has two distinct phases. During the first phase, the member nodes compact their measurements according to jaccard similarity function to estimate the similarity between various sets of data. In the second phase, data aggregation is carried out at the cluster head level using frequency filtering technique. To improve accuracy and reduce error in measurements, the experimental results need to be carried out using various filtering techniques. A data fusion-average median sampling (DF-AMS) approach was proposed for WSNs in [20]. The proposed scheme combines data aggregation with sampling and operates in two stages. During the first stage, a sample of nodes among the member nodes within a cluster are selected, based on their residual energy. These sample nodes transmit data to their respective cluster heads. During the second stage, the cluster heads perform data aggregation on the raw data captured by the sample nodes, using the median and maximum method. A novel suffix-prefix filtering for data aggregation was presented in [18]. The proposed approach performs data aggregation at two levels:

node level and cluster head level. This approach is computationally lightweight, eliminates redundancy to a greater extent and reduces the volume of data transmitted across the network. Moreover, the proposed approach outperforms the existing prefix filtering techniques in term of various quality of service (QoS) metrics.

Although, there exists many studies to eliminate redundancy using various data aggregation techniques, the latter still remains an open research area due to the absence of computationally lightweight techniques in view of limited resources of the nodes. The main objective of our work is to introduce a comprehensive, yet simple and lightweight data aggregation scheme to meet the resource-constrained nature of sensor nodes. Unlike the existing data aggregation techniques that eliminate redundancy at the cluster head level, we eliminate it at the fine-grained level, i.e., at the member nodes. Our proposed scheme considers the energy of nodes while performing data aggregation and has reduced complexity in comparison to the existing schemes.

3 Network model

For our proposed scheme, we consider a cluster-based hierarchical architecture of Fig. 1. This type of network model is typically found in large-scale WSNs, partitioned by cluster-based protocols. In this figure, A, B and C are the source nodes that need to send a varying number of packets to a cluster head M. We assume that each node has the battery capacity of sending 1500 data packets, only. For the source nodes, the shortest path to the cluster head is the same, i.e., via node D, with minimum number of hop counts. If these source nodes transmit their data packets to node D prior to the cluster head, it will create an energy hot spot problem for the nodes near the host spot region. This will exhaust their energy earlier than expected due to the large volume of traffic that passes through this route towards the cluster head. On the other hand, if the same traffic is diverted through multiple paths, the situation may be different for node D. It is obvious that if the communication pattern is changed from shortest to multiple paths, the network lifetime will be prolonged due to the presence of relay nodes. We made the following assumptions during the evaluation of our proposed scheme.

1. We assume that the network dies whenever the first node dies, i.e., stability period.
2. The packet size is 128 bytes
3. Transmission cost of a packet is 1.064 μ joule.
4. The node sampling/sensing rate is 1 packet/second.

4 Our proposed scheme: a data-driven approach

In this section, we provide analytical analysis of our proposed model in Section 4.1, followed by a detailed description of the proposed scheme in Section 4.2.

4.1 Analytical analysis

In our proposed scheme, the member nodes of each cluster capture the data and place them according to various stratum. These stratum are defined within the buffer of each node. We have defined seven stratum, ranging from 26^{0}C to 32.99^{0}C, for the temperature readings captured by the member nodes. These seven stratum are designed based on the previous historical and archival data captured by [25]. However, temperature is a seasonal parameter that changes dynamically. Hence, the stratum storing these temperature values changes dynamically as well. Thus, the number of stratum may varies from few to many, based on the weather

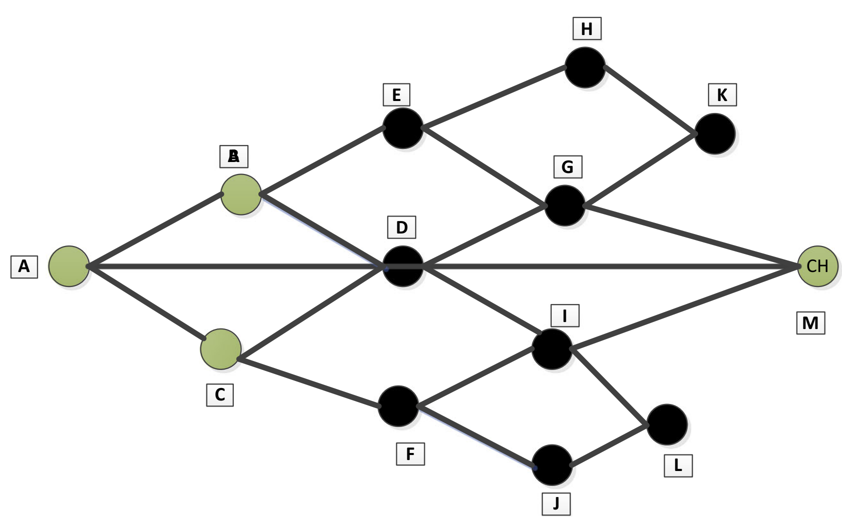


Fig. 1 Proposed network architecture

conditions and the readings captured in real-time from the sensing field. Once a new value is sensed by a member node, it is placed into the appropriate stratum, based on its reading.

Let each stratum be defined as follow.

$$S_1 = \{26.00...26.99\}$$
$$S_2 = \{27.00...27.99\}$$
$$S_3 = \{28.00...28.99\}$$
$$S_4 = \{29.00...29.99\}$$
$$S_5 = \{30.00...30.99\}$$
$$S_6 = \{31.00...31.99\}$$
$$S_7 = \{32.00...32.99\}$$

Although, our proposed scheme utilizes only seven stratum, it can be configured up to n^{th} stratum based on an application requirements, as shown in Eq. 1.

$$S_n = \beta...\beta\gamma. \tag{1}$$

Here, $\beta \in N$, where, $26 < \beta < 32$, $\gamma = 0.99$ and N is a natural number.

Next, we define a function f to take into account the changes incurred in each stratum with respect to time and temperature, as shown in Eq. 2. Here, α is the seasonal change in temperature and t is the time.

$$f = \sum_{i=1}^{n} S_i(t, \alpha). \tag{2}$$

For each stratum, we define Min $(S_n) = S_i$ and Max $(S_n) = S_j$, such that

$$S_i, \quad \alpha < m(S_i, S_j). \tag{3}$$

and

$$S_j, \quad \alpha > m(S_i, S_j). \tag{4}$$

Here, m is the mean of S_i, S_j. Based on Eqs. 3 and 4, the value of $f(S_n)$ is equal to 0, when α=m, i.e., at the critical point. Using the above calculations, we have

$$f(S_n) = \begin{cases} S_i, & if \ \ \alpha < m \ , \\ 0, & if \ \ \alpha = m \ , \\ S_j, & if \ \ \alpha > m \ . \end{cases} \tag{5}$$

Proof of Eq. 5

Let's consider the quadratic Taylor polynomial for a function f, centered at m, such that,

$$f(S_i) \approx f(m) + f'(m)(m-1) + f''(m)\frac{(m-2)^2}{2!}. \tag{6}$$

If f has a critical point at m, i.e., $f'(m) = 0$, this approximation reduces to

$$f(S_i) \approx f(m) + f''(m)\frac{(m-2)^2}{2}. \tag{7}$$

□

From Eq. 7, one can see that whenever $f(m) > 0$, i.e., nearer to m+1, f behaves like a parabola with vertex at $(m, f(m))$ and a positive leading co-efficient as shown in Fig. 2. Therefore, the critical point at m is the location of a relative minimum. Similarly, the function needs to have a relative maximum at m whenever, $f'(m) < 0$. Finally, when $f''(m) = 0$, it is clear that the test is inconclusive as the newly captured data α is equal to m.

However, additional insight into this situation can be provided by looking at higher degree Taylor polynomials. The cubic Taylor polynomial approximation for f centered at m is

$$f(S_i) \approx f(m) + f'(m)(m-1) + f''(m)\frac{(m-2)^2}{2!} + f'''(m)\frac{(m-3)^3}{3!}. \tag{8}$$

If $f'(m) < 0$ and $f''(m) = 0$, then the Taylor approximation reduce to

$$f(S_i) \approx f(m) + f'''(m)\frac{(m-3)^3}{3!} \tag{9}$$

Thus, nearer to m, the behavior of $f(S_i)$ is similar to that of the cubic function $C(m-3)^3$, where C is a real number. Using the above calculations, we can derive it for the n^{th} series.

$$f(S_i) \approx f(m) + f'(m)(m-1) + f''(m)\frac{(m-2)^2}{2!} + f'''(m)\frac{(m-3)^3}{3!} + ... + f^{(n)}(m)\frac{(m-n)^n}{n!}. \tag{10}$$

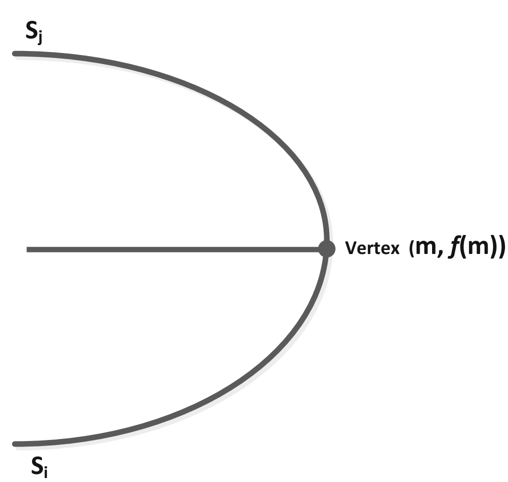


Fig. 2 Parabola of MinMax

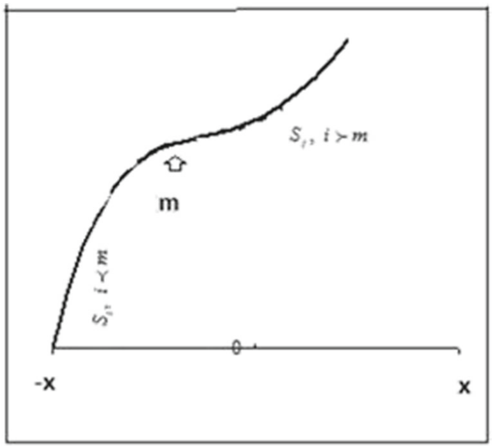


Fig. 3 Schematic diagram of the proposed analytical model

Based on Eqs. 6, 7, 8, 9 and 10, $f(S_n)$ yields

$$f(S_n) = \begin{cases} S_i, & if \ \alpha < m \\ 0, & if \ \alpha = m \\ S_j, & if \ \alpha > m \end{cases} \tag{11}$$

The whole procedure is graphically represented in Fig. 3.

4.2 Description of the proposed method

In our proposed approach, we considered a deployed sensor region A, that contains T_n nodes, where $T_n = \{1, 2, 3, ..., n\}$. These nodes are grouped together into equal-sized clusters T_c, where $T_c = \{C_1, C_2, C_3, ..., C_n\}$, using a balanced clustering algorithm [23]. Within a cluster, each member node captures the raw data by sensing the environment. At the node level, we defined various classes, i.e., stratum, based on common characteristics, i.e., $S_1, S_2, ..., S_n$. Each strata is dynamic in nature that changes according to the application requirements. We have designed these startum using the concept of stratified random sampling, a probability-based sampling method [10], for designing stratum. The stratified random sampling is operationally better than simple random sampling, especially when population units are homogeneous [31]. Our proposed scheme has two distinct but interrelated phases.

1. Comparison of the newly captured value with the mean value (m) of a stratum.
2. Comparing the value obtained from the previous step with either Min (S_i) or Max (S_j) value of that specific stratum.

During the first phase, whenever a new reading is captured by T_n, it goes into an appropriate stratum. To decide the range of a stratum, i.e., MinMax, the value of new reading is compared with m of that particular stratum. In the next phase, there are two probable conditions. The data may either be a new Min or Max that replace the current value within a specific stratum. For instance, if the new value obtained from the previous step is confirmed as Min, then, this value is compared with the existing Min value. The smaller among the two is chosen as the new Min value of that particular stratum. However, for the Max value, the new value is compared with the current Max value, and the greater among the two is chosen as the new Max value of that particular stratum. This process is repeated for each stratum. Moreover, the same procedures are repeated each time a new value is captured by a node. As a result, each stratum contains only two values, i.e., MinMax, which are forwarded from each stratum of a given member node to its respective cluster head at a predefined interval.

At the cluster head level, data redundancy is checked again among the data packets that it has received from its member nodes. However, the received packets from only those member nodes are compared that are in close proximity of each other. As a result, there is no need to check for data redundancy from all the member nodes. Hence, a substantial amount of processing energy of the cluster heads is conserved, as shown in the Fig. 4.

Our proposed approach achieves its goal by having a maximum of only two comparisons for real-time data, i.e., mean and MinMax [7, 26]. Thus, the performance and computational complexity of our proposed approach remain constant with an increase in the size of captured data. In existing data aggregation algorithms, the computational complexity increases with an increase in the size of captured data. Hence, the performance of a node is degraded, that shortens its lifetime. Our proposed approach lowers the network traffic, network congestion and packet collision rate. The stratum of each member node transmits only two packets after 1 minute, i.e., 14 packets from the strata of a single node. The proposed scheme has a very lower

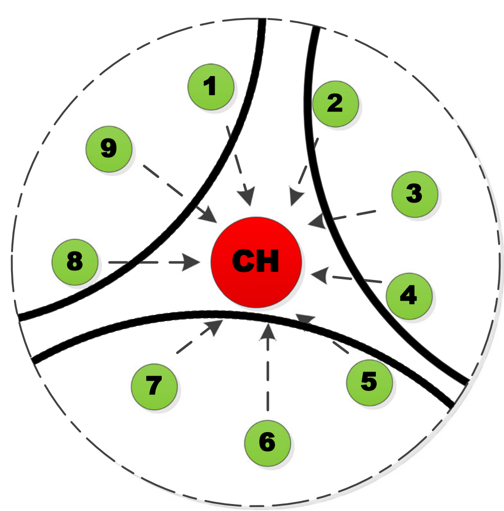


Fig. 4 Data aggregation at the cluster head level

complexity of $O(n)$, where n is so small that is equal to order 1 and is considered as $O(1)$. Hence, a considerable amount of energy is conserved. The detailed description of our proposed approach is shown in the Algorithm below.

Algorithm 1 In-node data aggregation

1. Input: α, β, m
2. OUTPUT: MinMax
3. BEGIN
4. **STEP 1** [For any newly captured α, assign stratum according to pre-defined range]
5. **STEP 2** [Match Stratum]
   $S_n = \beta ... \beta\gamma$
   if $(\alpha < mean)$
6. **STEP 3** [Compare α with Min for choosing minimum value]
   if $(\alpha < \min Si)$
   Min $Si = \alpha$ (New Min)
   else if $(\alpha > mean)$
7. **STEP 4** [Compare α with Max for choosing maximum value]
   if $(\alpha > \max S_i)$
   max Si=α (New Max)
   else if $(\alpha == mean)$
   do nothing
   exit loop

Based on our Algorithm, The data flow diagram of our scheme is presented in Fig. 5.

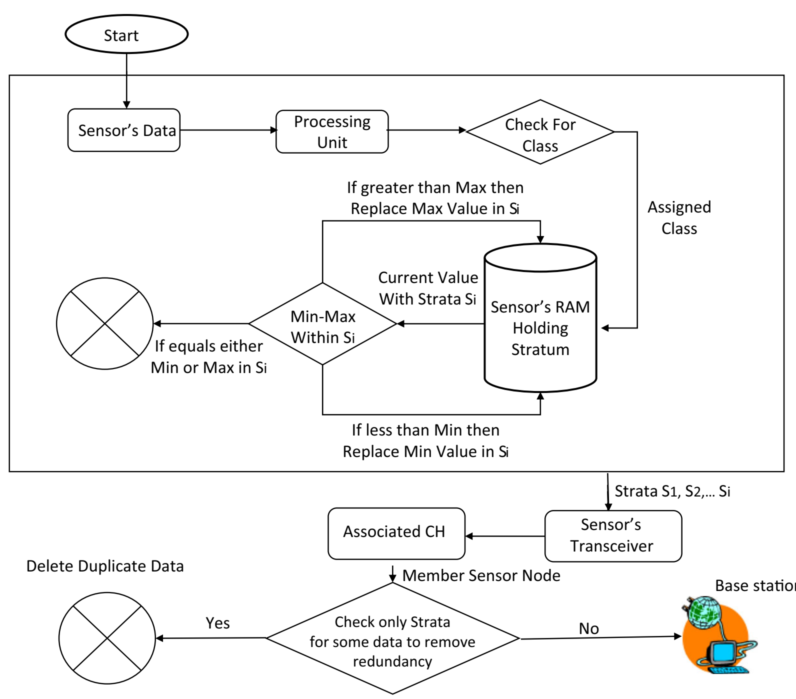
5 Experimental results

There are two main objectives of our proposed approach, i.e., energy efficiency and congestion control. To achieve these objectives, we performed in-network processing to conserve the communication cost. Upon analytical validation of our proposed technique in Section 4, we provide the experimental results here. In this section, we compared our scheme in terms of energy consumption, number of transmitted packets, processing cost and communication cost.

In Fig. 6, the energy consumption of a source node, i.e., member node, in presence and absence of our proposed approach is shown, with respect to time. The energy consumption of a node is reduced by 23% while employing our proposed aggregation technique. In this figure, the absence of our proposed scheme means that the source nodes do not employ any data aggregation and forward highly redundant and correlated data to the cluster heads. In absence of our proposed scheme, the energy consumption is almost double after 120 minutes of network deployment. The increase in energy consumption deteriorates the lifetime of WSNs.

In Fig. 7, the realistic scenarios of different number of refined packets (after performing data aggregation) with respect to various nodes are shown. In Fig. 7a and b, the number of nodes varies from 1 to 10 and the time varies from 1 minute to 60 minutes, respectively. The main objective was to study and validate our claim that our proposed technique reduces congestion to a greater extent by reducing the number of packets. Recall from Eqs. 3 and 4

Fig. 5 Data flow diagram



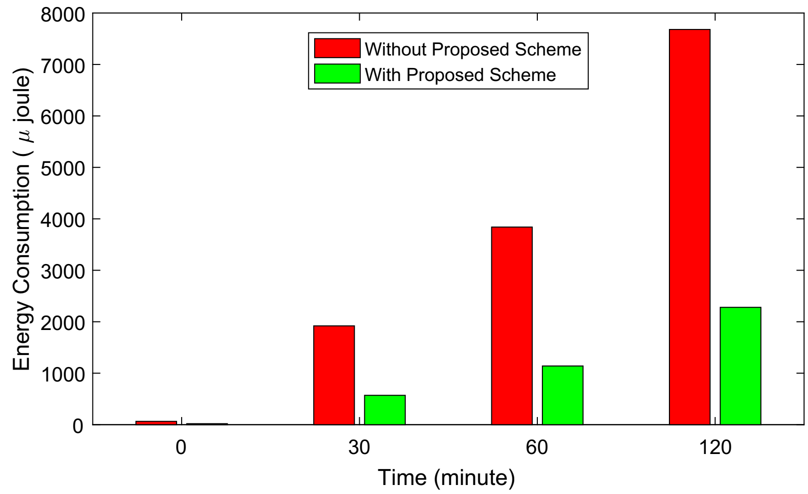


Fig. 6 Energy consumption vs. Time

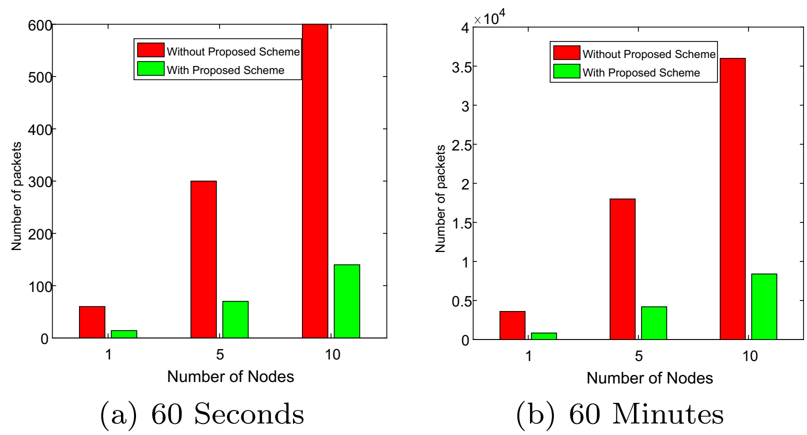


Fig. 7 Number of refined packets vs. number of nodes (in term of time)

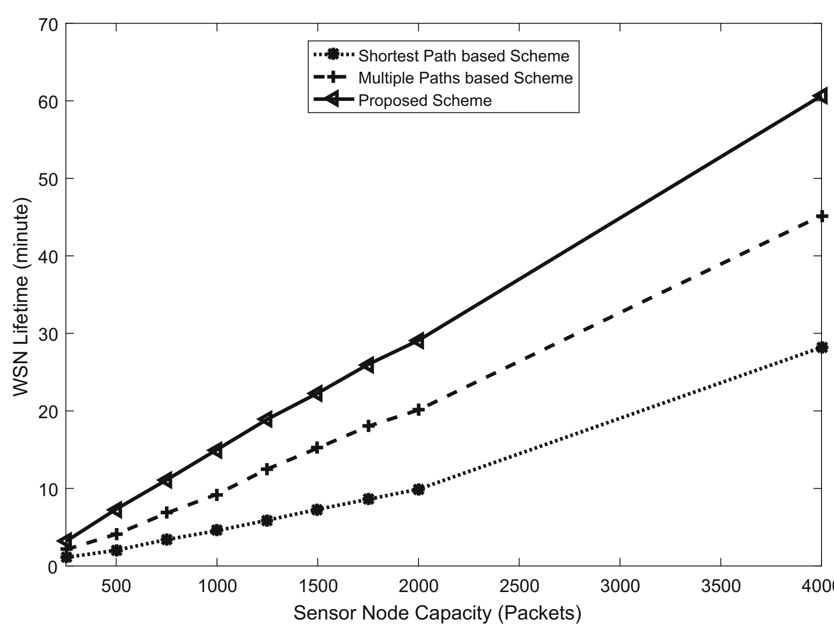
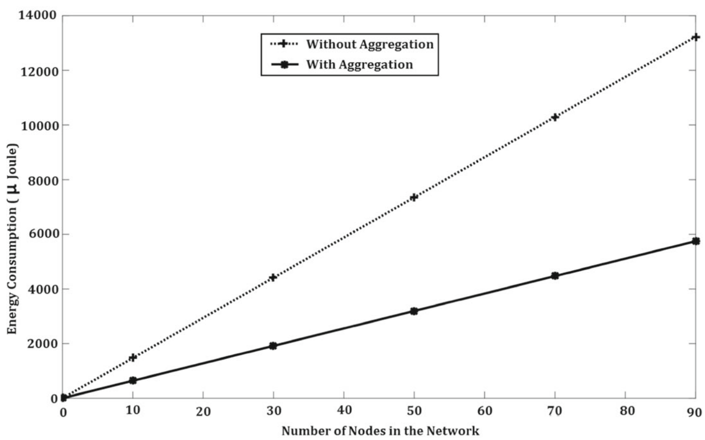
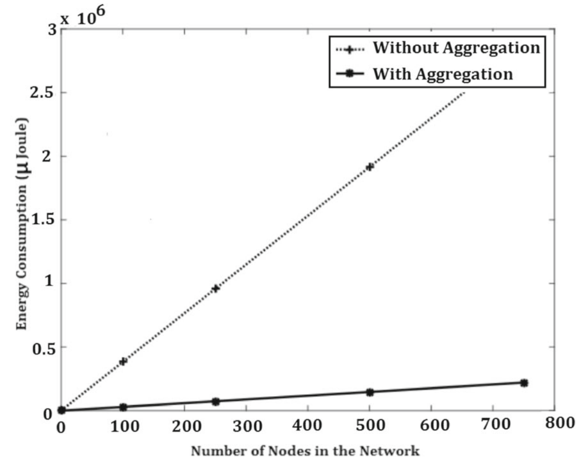


Fig. 8 WSN lifetime vs. sensor node capacity



(a) At 90 Nodes



(b) At 800 Nodes

Fig. 9 Energy consumption vs. varying number of nodes

in Sec. 4.1, only two packets from each stratum can be transmitted. This is in contrary to traditional approaches where each packet is forwarded towards the destination containing redundant data. This substantially reduces the network traffic, lowers congestion and ultimately bandwidth of the network.

In Fig. 8, the lifetime of WSN with respect to sensor nodes' capacity is shown. The network lifetime is the time interval at which the base station receives the last packet before the death of its first node. Our proposed scheme integrates the shortest path algorithm with our MinMax aggregation technique. In this figure, we compare our proposed scheme with two existing schemes, i.e., shortest path based scheme and multiple path based scheme. On y-axis, the lifetime of the network is shown, while on x-axis, the sensor node' capacity, i.e., number of packets that a node receives and transmits with the available battery power, is shown. We have investigated the WSN lifetime by assuming different capacities of a node. The evaluation results show that our proposed scheme is exceptionally better in terms of WSN lifetime as compared to the existing algorithms.

In Fig. 9, the energy consumption of our proposed scheme is shown for a varying number of nodes. In this figure, the energy consumption for a varying number of nodes is evaluated in presence and absence of data aggregation. In presence of our proposed data aggregation, the energy consumption is more than 3 times lower when the number of deployed nodes are 90 in the network, as shown in Fig. 9a. In case of 800 deployed nodes, the energy consumption is 18.8 times lower in comparison to absence of an aggregation approach, as shown in Fig. 9b. These results show that our proposed approach has substantially reduced the energy consumption with varying number of nodes. However, the proposed approach suffer from delay that occur due to in-node aggregation. Data is kept in the nodes for a predefined time. Thus, our proposed approach is not suitable for delay sensitive applications such as military application, where quick responses are required for critical and strategic decision making. However, this delay may be reduced if the sampling interval in reduced from 1 minute to 15 or 30 seconds, that will ultimately reduces the delay up to 50 % or more, improving efficiency of the proposed approach.

6 Conclusion

In this paper, we have proposed a novel and energy-efficient in-network data aggregation approach for a cluster-based hierarchical network. Each member node captures a temperature reading and compares it against a particular stratum. Based on its value, the Min and Max of that specific stratum is configured. The use of Min/Max permits a node to send only two values to the cluster head, after a pre-defined time interval. The proposed approach eliminates the burden of transmitting the whole of raw data from the member nodes to a base station via the cluster heads. Our approach is lightweight as it requires a maximum of two comparisons for real-time data, i.e., mean and Min/Max. Moreover, the transmission of only two values from each stratum lowers the communication cost, packet collision rate, network congestion and extends the network life time. Our proposed approach has a much better energy conservation with respect to time, reduced number of generated packets, better lifetime with respect to node capacity and reduced energy consumption with respect to number of deployed nodes. In future, we aim to evaluate our proposed approach in a real-world scenario using testbeds. We also aim to evaluate our approach for heterogeneous applications that collect diversified data. Moreover, the feasibility of the proposed scheme needs to be tested in tree and flat-based networks.

References

1. Akyildiz IF, Kasimoglu IH (2004) Wireless sensor and actor networks: research challenges. Ad hoc netw 2(4):351–367
2. Akyildiz IF, Vuran MC (2010) Wireless sensor networks, vol 4. Wiley, New York
3. Alam M, Albano M, Radwan A, Rodriguez J (2012) Context parameter prediction to prolong mobile terminal battery life. In: Mobile multimedia communications, pp 476–489. Springer, Berlin
4. Alam M, Rodriguez J (2010) A dual head clustering mechanism for energy efficient WSNs. In: Lecture notes of the institute for computer sciences, social informatics and telecommunications engineering, pp 380–387. Springer, Berlin
5. Anastasi G, Conti M, Di Francesco M, Passarella A (2009) Energy conservation in wireless sensor networks A survey. Ad hoc netw 7(3):537–568
6. Bahi JM, Makhoul A, Medlej M (2014) A two tiers data aggregation scheme for periodic sensor networks. Ad Hoc Sens Wirel Netw 21(1-2):77–100
7. Berl A, Gelenbe E, Di Girolamo M, Giuliani G, De Meer H, Dang MQ, Pentikousis K (2010) Energy-efficient cloud computing. Comput J 53(7):1045–1051
8. Biswas S, Das R, Chatterjee P (2018) Energy-efficient connected target coverage in multi-hop wireless sensor networks. In: Industry interactive innovations in science, engineering and technology, pp 411–421. Springer
9. Cheng L, Niu J, Luo C, Shu L, Kong L, Zhao Z, Gu Y (2018) Towards minimum-delay and energy-efficient flooding in low-duty-cycle wireless sensor networks. Comput Netw 134:66–77
10. De Vries PG (1986) Stratified random sampling. In: Sampling theory for forest inventory, pp 31–55. Springer
11. Dhand G, Tyagi SS (2016) Data aggregation techniques in wsn Survey. Procedia Comput Sci 92:378–384
12. Du T, Qu Z, Guo Q, Qu S (2015) A high efficient and real time data aggregation scheme for wsns. Int J Distributed Sens Netw 11(6):261381
13. Fasolo E, Rossi M, Widmer J, Zorzi M (2007) In-network aggregation techniques for wireless sensor networks: a survey. IEEE Wirel Commun 14(2):70–87
14. Ganjewar PD, Barani S, Wagh SJ, Sonavane SS (2018) Survey on data reduction techniques for energy conservation for prolonging life of wireless sensor network. Wirel Commun 10(2):17–25
15. Hansen MH, Hurwitz WN (1953) Sample survey methods and theory, vol I. Wiley, New York
16. Harb H, Makhoul A, Laiymani D, Jaber A, Tawil R (2014) K-means based clustering approach for data aggregation in periodic sensor networks. In: 2014 IEEE 10Th international conference on wireless and mobile computing, networking and communications (wimob), pp 434–441. IEEE
17. Harb H, Makhoul A, Tawbi S, Couturier R (2017) Comparison of different data aggregation techniques in distributed sensor networks. IEEE Access 5:4250–4263
18. Harb H, Makhoul A, Tawil R, Jaber A (2014) A suffix-based enhanced technique for data aggregation in periodic sensor networks. In: 2014 international wireless communications and mobile computing conference (IWCMC), pp 494–499. IEEE
19. Haupt J, Bajwa WU, Rabbat M, Nowak R (2008) Compressed sensing for networked data. IEEE Signal Proc Mag 25(2):92–101
20. Huy DV, Viet ND (2015) Df-ams: Proposed solutions for multi-sensor data fusion in wireless sensor networks. In: 2015 Seventh international conference on knowledge and systems engineering (KSE), pp 1–6. IEEE
21. Jan MA, Jan SRU, Alam M, Akhunzada A, Rahman IU (2018) A comprehensive analysis of congestion control protocols in wireless sensor networks. Mob Netw Appl 23:1–13
22. Jan MA, Nanda P, He X, Liu RP (2014) Pasccc: Priority-based application-specific congestion control clustering protocol. Comput Netw 74:92–102
23. Jan MA, Usman M, He X, Rehman AU (2018) Sams: A seamless and authorized multimedia streaming framework for wmsn-based iomt. IEEE Internet Things J. https://doi.org/10.1109/JIOT.2018.2848284
24. Kafi MA, Djenouri D, Ben-Othman J, Badache N (2014) Congestion control protocols in wireless sensor networks A survey. IEEE Commun Surv Tutorials 16(3):1369–1390
25. Khan R, Ali I, Suryani AM, Ahmad M, Zakarya M (2013) Wireless sensor network based irrigation management system for container grown crops in pakistan. World Appl Sci J 24(8):1111–1118
26. Laughlin SB, Sejnowski TJ (2003) Communication in neuronal networks. Science 301(5641):1870–1874
27. Liu H, Ma J, Huang W (2018) Sensor-based complete coverage path planning in dynamic environment for cleaning robot. CAAI Trans Intell Technol 3(1):65–72
28. Liu Z, Tsuda T, Watanabe H (2015) Traffic deduction exploring sensor data's intra-correlations in train track monitoring wsn. In: 2015 IEEE SENSORS, pp 1–4. IEEE
29. Ma J, Jiang X, Gong M (2018) Two-phase clustering algorithm with density exploring distance measure. CAAI Trans Intell Technol 3(1):59–64
30. Msechu EJ, Giannakis GB (2012) Sensor-centric data reduction for estimation with wsns via censoring and quantization. IEEE Trans Signal Process 60(1):400–414
31. Omona J (2013) Sampling in qualitative research: Improving the quality of research outcomes in higher education. Makerere J High Educ 4(2):169–185
32. Pattem S, Krishnamachari B, Govindan R (2008) The impact of spatial correlation on routing with compression in wireless sensor networks. ACM Trans Sens Netw (TOSN) 4(4):24
33. Rahman H, Ahmed N, Hussain I (2016) Comparison of data aggregation techniques in internet of things (iot). In: International conference on wireless communications, signal processing and networking (wiSPNET), pp 1296–1300. IEEE
34. Rout RR, Ghosh SK (2013) Enhancement of lifetime using duty cycle and network coding in wireless sensor networks. IEEE Trans Wirel Commun 12(2):656–667
35. Sayood K (2005) Introduction to data compression. Elsevier, Amsterdam
36. Singh VK, Kumar M (2018) In-network data processing in wireless sensor networks using compressed sensing. Int J Sens Netw 26(3):174–189
37. Uthayakumar J, Vengattaraman T, Dhavachelvan P (2018) A survey on data compression techniques: From the perspective of data quality, coding schemes, data type and applications. Journal of King Saud University-Computer and Information Sciences

38. Xiang L, Luo J, Vasilakos A (2011) Compressed data aggregation for energy efficient wireless sensor networks. In: 2011 8th annual IEEE communications society conference on sensor, mesh and ad hoc communications and networks (SECON), pp 46–54. IEEE
39. Xu X, Ansari R, Khokhar A, Vasilakos AV (2015) Hierarchical data aggregation using compressive sensing (hdacs) in wsns. ACM Trans Sens Netw (TOSN) 11(3):45
40. Yang H, Yu L (2017) Feature extraction of wood-hole defects using wavelet-based ultrasonic testing. J For Res 28(2):395–402
41. Yetgin H, Cheung KTK, El-Hajjar M, Hanzo LH (2017) A survey of network lifetime maximization techniques in wireless sensor networks. IEEE Commun Surv Tutorials 19(2):828–854
42. Zhu C, Wu S, Han G, Shu L, Wu H (2015) A tree-cluster-based data-gathering algorithm for industrial wsns with a mobile sink. IEEE Access 3(1):381–96

Affiliations

Syed Rooh Ullah Jan[1] · Mian Ahmad Jan[1] · Rahim Khan[1] · Hakeem Ullah[2] · Muhammad Alam[3] · Muhammad Usman[4]

Syed Rooh Ullah Jan
roohullahsyed@awkum.edu.pk

Hakeem Ullah
hakeem.ullah@awkum.edu.pk

Muhammad Alam
m.alam@xjtlu.edu.cn

Muhammad Usman
musman@swin.edu.au

1 Department of Computer Science, Abdul Wali Khan University Mardan, Mardan, Pakistan

2 Department of Mathematics, Abdul Wali Khan University Mardan, Mardan, Pakistan

3 Department of Computer Science and Software Engineering, Xi'an Jiaotong-Liverpool University (XJTLU), Suzhou, China

4 Department of Computer Science and Software Engineering, Swinburne University of Technology, Melbourne, Australia