CC_SCR: A Compression Cluster-based scheme in a Spatial Correlated Region for Wireless Sensor Networks

Sofiane Moad, Mario Rivero, Nizar Bouabdallah and Rami Langar
Email: sofiane.moad@inria.fr, mriveroa@ipn.mx, nizar.bouabdallah@inria.fr, Rami.Langar@lip6.fr

Abstract—In this paper, we propose a new Compression Cluster-based scheme in a Spatial Correlated Region (CC_SCR) for event-driven applications in wireless sensor networks. The main idea behind our proposal is to exploit the spatial correlation of such networks in order to reduce the size of the data packets that will be sent by means of data compression. The proposed clustering scheme is based on selecting a data value as reference while the rest of the active nodes transmit only the difference between their sensed value and this reference value. Hence, one major issue in the proposed mechanism is to appropriately select the reference node that achieves the highest reduction of the packet size among all active nodes. The CC_SCR protocol is evaluated by simulations and the results show that CC_SCR may reduce as much as 11 times the energy consumption compared to a classical clustering scheme.

keywords: data compression, clustering.

I. INTRODUCTION

Wireless sensor networks (WSNs) consist of spatially distributed autonomous devices, called sensors, that communicate in a wireless manner. Sensors cooperate together to monitor physical or environmental conditions such as temperature, sound, vibration and pressure. This technology has been originally developed for military applications such as battlefield surveillance. Recently, WSNs have been deployed in many other civilian application areas including environment applications such as fire detection in forest monitoring and health-care applications like monitoring the patient status.

Due to the limited characteristic of sensors, they should be able to self-organize, self-configure and should optimize the energy consumption to maximize the network’s lifetime. The network lifetime in WSNs refers to the period of time from the deployment of the sensor nodes to the instant when the network is considered unusable [1]. One way to extend the lifetime in WSNs is by means of data compression. By applying a suitable compression scheme, the consumed power is reduced at the transmission and processing tasks and thus extend the network’s lifetime. Also, by reducing the data packet size, less bandwidth is required for sending and receiving data. Hence, data compression is one effective method to utilize limited resources.

In terms of energy consumption, sensors consume energy for three main reason: sensing, processing and wireless communicating. The wireless communication refers to data transmission and reception. Among these three operations, it is known that the most power consuming task is data transmission. Approximatively 80% of power consumed in each sensor node is used for data transmission [2]. Thus by reducing the size of the data packets by means of compression, the energy consumed by the transmission and reception is also reduced. Indeed, a number of compression-based protocols have been developed to make these networks practical and efficient such as in [2].

In this paper, we propose a new Compression Cluster-based scheme in a Spatial Correlated Region (CC_SCR) for event-driven applications in wireless sensor networks. As WSNs are typically densely deployed over a sensor field [3], sensor nodes are typically very close to each other. Contrary to continuous monitoring applications, in Event Detection-Driven (EDD) applications, the active nodes are concentrated in a relatively small area. Therefore, the readings from these nodes are expected to be very similar. Building on this, we propose a clustering scheme that exploits this spatial correlation of the sensed data among the nodes to reduce the size of the data packets that will be sent. Specifically, in the proposed scheme, the Cluster Members (CMs) send only the difference between their readings and a reference data which corresponds to the value sensed by the selected CH. One important issue of the proposal is to select as Cluster Head (CH) the node that reduces the average packet size in the cluster. This issue is solved by allowing the BS to participate into the CH selection and also by means of considering the characteristics of the physical surveilled event in the CH selection.

The remainder of this paper is organized as follows. Section II presents a background of research related to this work. Section III exposes the network model, while section IV presents the performance evaluation. Finally, the paper concludes in section V with a summary of the main advantages of the proposed scheme.

II. RELATED WORK ON COMPRESSION SCHEMES

In the literature, there has been an increased interest in studying compression algorithms. On the other hand, many of these compression algorithms have been proposed for classical networks which are not suitable to be deployed in the WSNs context [4], [5]. The main reason is the limited memory size of sensor nodes. For example, the size of bzip2 is 219KB and the size of LZO is 220KB. Another reason is the limited processor speed of sensor nodes which is
around 4 – 8 MHz. Thus, embedding classical data compression schemes in these tiny nodes is very difficult and it is necessary to design a low-complexity and small size data compression algorithm for sensors. In general, there are two types of data compression algorithms: lossless and lossy algorithms. Lossless algorithms guarantee the data integrity during the compression/decompression process, while lossy algorithms ensure a higher compression ratio at the expense of allowing a loss of information. We review hereafter some of these compression techniques. A more detailed description of compression methods can be found in [6].

One compression algorithm for WSNs is the coding by ordering data compression technique [7]. In this algorithm, when data is combined at an aggregation node, some data is implicitly transmitted. The main idea behind this technique is to replace the data transmission of certain nodes by the order of compression data. The main idea behind this technique is the number of nodes sending a packet, \( n \), the number of sensor nodes dropped at the aggregation node, \( n_2 \). The data value of each packet can be any integer from a range of 0 to 23. If the aggregator node receives the data packets in the order \( n_0, n_1, n_2, n_3, n_4 \), then the value of packet \( n_4 \) is implicitly received as 0. Then, there are \( 4! = 24 \) possible ways of ordering the data packets. In the general case, the following inequality has to be satisfied to make this algorithm practical, \( m - l \geq (n-m+l) \times k^l \), where \( m \) is the number of nodes sending a packet, \( n \) is the total number of sensor nodes, \( k \) is the possible range data value, and \( l \) is the number of sensor nodes dropped at the aggregation node.

The pipelined in-network compression algorithm is discussed in [8]. The main idea behind this technique is to combine the data to make the data packets smaller than the original size. After collecting data from different nodes by the aggregator, it is stored for a certain amount of time. Data packets are then combined into one packet to minimize the data transmission. For example, consider that each data packet has the following form: \(<\text{data value, node ID, timestamp}>\). Then, the compressed data packet has the following form: \(<\text{shared prefix, suffix list, node ID list, timestamp list}>\). The shared prefix field is the most significant bits, which is the same for all the values. The suffix list field expresses the list of measured values excluding the shared prefix part. The node ID list is the list of node identifiers and the timestamp list is the list of timestamps. One advantage of this simple compression scheme is that the shared prefix system can be used for both node IDs field and timestamp fields. By doing so, more data compression can be achieved.

The compression scheme proposed in this paper is different from the schemes proposed in [7] and [8] since our proposed scheme considers the characteristics of the physical surveilled event and at the same time it takes advantage of the energy of the Base Station (BS) which participates into the CH selection. Specifically, our proposed scheme takes advantage of the fact that the nodes that sense a certain event are usually very close to each other in order to reduce the size of data packets communicated through the network. The BS then select an efficient CH that minimize the data transmission over the network.

III. NETWORK MODEL

We consider event detection driven wireless sensor applications. The center of the event is located in a random uniformly distributed point with coordinates \((x_{event}, y_{event})\) within the network’s area. The range of the event, i.e., the area range where sensors can detect the event is \( R_{event} \) meters, where \( R_{event} \in [1, \text{meters}] \). We also suppose that an event has a duration of \( T_{event} \) seconds. Additionally, in our model, only one event can be active inside the system’s area. The data value \( C \) at the center of the event is constant i.e., the stationary model in which the measured data do not change during the \( T_{event} \) seconds that the event is active.

We consider a WSN consisting of \( N \) sensors deployed over a vast field. We denote the i-th sensor node by \( n_i \) and the corresponding sensor node set \( S = \{n_1, n_2, ..., n_N\} \) where \( |S| = N \). Some assumptions about the sensor nodes and the underlying network model are now presented:

- Nodes are uniformly distributed in an \( A \times A \) area with \((x, y)\) coordinates. Nodes are homogenous, stationary and have the same capabilities. Each node is assigned a unique identifier ID.
- Nodes have two power controls to vary the transmission power which depends on the distance to the receiver. Each node \( n_i \) can reach any other node with a transmission range \( R_c \). The BS can be reached with transmission range \( R_t > R_c \).
- Nodes are characterized by their coverage range \( R_{cov} = R_{event} \). Only nodes within \( R_{event} \) range are considered as active nodes in the network and they are the only nodes performing as the source of the detected event. The other nodes are not considered in our analysis as they do not participate into the data reporting.
- CHs use the average operation as the aggregation to eliminate the data redundancy. Other aggregation techniques, such as those proposed in [9] can also be implemented.

The spatial correlation of the data from the different active nodes can be considered according to the following models:

1. Diffusion property [10].
2. Data is jointly gaussian with the correlation being a function of the distance [11].
3. Data is a function for their joint entropy [12].
4. Correlation is calculated from realistic environmental monitoring and testbeds.

In this paper, we use the diffusion property to model the spatially correlated data [13]. The model considered in this paper is the same as in [10] in which the data reading at a distance \( d \) from the center of the event is \( D_\alpha(d) \). Specifically, the data reading in any point at distance \( d \) from the center of the event is \( D = \frac{C}{(\alpha + \frac{d}{x})^\tau} \), where \( C \) is a constant representing the value at the center of the event, and \( \alpha \) is the diffusion parameter which depends on the particular environment and phenomenon surveilled, e.g., for light \( \alpha = 2 \), heat = \( \alpha = 1 \).
Fig. 1 shows the data reading using the aforementioned model, with different values of \( \alpha \) and \( C = 250 \). On one hand, when \( \alpha \geq 1 \), we observe a relatively big difference between the value at the center of the event and the values observed at distance \( d \) far away from the center of the event. On the other hand, when \( \alpha < 1 \ (\alpha = 0.1, 0.01 \) and \( 0.001 \)), the data readings away from the center of the event are very similar. In our study, we are interested, specifically, on the type of events where data values are highly correlated.

![Graph showing variation of data reading with distance from the event center.](image)

We use Henizelman’s energy consumption model [14]. Specifically, the energy consumed to transmit a message is

\[
E_{tx}(sz, d) = \begin{cases} 
    sz \times E_{elec} + sz \times E_{fs} \times d^2, & \text{if } d \leq d_0, \\
    sz \times E_{elec} + sz \times E_{mp} \times d^4, & \text{otherwise.}
\end{cases}
\]

where \( sz \) is the data packet size in bits, \( E_{elec} \) is the energy consumed due to the transmitter/receiver circuitry, \( E_{mp} \) is the energy consumed by the transmitter amplifier and \( d_0 = \sqrt{\frac{E_{tx}}{E_{mp}}} \) is the distance threshold between the transmitter and the receiver over which the multi-path fading channel model is used. The energy consumed to receive a message is \( E_{rx}(sz) = sz \times E_{elec} \). \( E_{DA} \) is the energy consumed to aggregate the data. Next, the classical protocol and \( CC_{SCR} \) are described.

A. Classical clustering protocol

A classical (or a standard) clustering process is composed by two phases: set up phase and steady state phase. When an event occurs in a random (uniformly distributed) point of the network, nodes inside the event area wake up and they start the clustering process. At the beginning of this phase, active nodes compete among each other to become CH. Specifically, active nodes transmit their control packet to the BS according to the specified random medium access protocol. In this paper, the Carrier Sense Medium Access (CSMA) control protocol is used. The control packet only comprises the node’s ID and no data are transmitted at this point. The first node that successfully transmits this packet becomes the CH. All nodes involved in the event reporting immediately send their signaling message to the BS. Therefore, the BS selects the first node that transmitted successfully the signaling message and it broadcasts a signaling message over the network for a CH notification. Thus the rest of the nodes become CMs. In the steady phase, CMs send their data in a scheduled fashion using a Time Division Multiple Access (TDMA) protocol. The CH aggregates the data values received from its CMs with its own data and sends the resulted data to the BS.

B. Proposed compression clustering protocol

The proposed clustering \( CC_{SCR} \) process is also composed of the same two phases, namely: set up phase and steady state phase. As in the classic protocol, the set up phase occurs whenever an event occurs in a region of the network. However, in \( CC_{SCR} \), the active nodes send their first measured data value to the BS, i.e., they no longer send just their control packet. Instead, active nodes send a data packet. The reason for this is that, this sensed data is used for the CH selection procedure. Indeed, this entails an extra energy consumption at the set up phase compared to the classical protocol. However, this first data transmission allows important energy saving in the steady state phase.

It is important to notice that, \( CC_{SCR} \) is best suited for environments where the event conditions are fairly stable during the event duration. This is due to the fact that the CH is chosen according to the first sensed data. Hence, if the event conditions suffer from a high variation, the originally selected CH may no longer render acceptable energy savings. Example of such applications includes a fire surveillance forest, in which when a fire occurs in a region, the temperature remains stationary for the duration of that fire in this region. Another example of application includes target tracking. In this kind of application, the target is the source of the measured data at sensor nodes, such as the light or temperature. Here, the measured data remains the same whenever the target stays in the same place and hence the sensor nodes sense the same measured data during the presence of the target. In what follows, we describe the set up and the steady state phase of the proposed algorithm.

- In the set up phase, after reception of the first data packets of all active nodes, the BS calculates the difference between the data of node \( n_i \) and the data of node \( n_j \) (\( i \neq j \), and \( i \leq N, j \leq N \)). Afterwards, these differences are summed over. We call this sum of the data value differences \( S_i \). Then, the BS selects as CH the node which minimizes the total difference calculated value \( S_i \) between each node \( n_i \) and node \( n_j \) (\( i \neq j \), and \( i \leq N, j \leq N \)). Finally, the BS broadcasts a control message to the active nodes in order to notify the node selected as CH. Therefore, the rest of the nodes consider themselves as CMs. Note that there is no need for the CMs to send any extra packets since the BS already knows the active nodes.
- In the steady state phase, CMs send the difference between their sensed data and the CH’s data value, which corresponds to a compressed value, called \( \Delta_i \), rather than the complete data packet, \( value_{CM_i} \). Therefore,
\[ \Delta_i = |\text{value}_{CM_i} - \text{value}_{CH}| \] represents the difference between the \(i\)-th CM’s data value \(\text{value}_{CM_i}\), and the corresponding CH data value \(\text{value}_{CH}\). In order to perform this compression, the CH sends its complete sample data value to the CMs at the beginning of each event occurrence. Therefore, the CMs send only the \(\Delta_i\) to the CH. The main advantage of the proposal scheme is that the \(S_i\) calculation is centralized at the BS, which is not energy constrained.

1) Example: To illustrate the protocol operation, let us consider the following example presented in Fig. 2(a) (the Fig. 2(b) shows the case of a classical scheme). We consider five active nodes \(n_1, n_2, n_3, n_4\) and \(n_5\) in the event region \(e\) which covers a region of range \(R_{\text{event}}\). We consider in this example the temperature as the sensed measurement value. Nodes \(n_1, n_2, n_3, n_4\) and \(n_5\) sense the value of \(20^\circ\), \(22^\circ\), \(19^\circ\), \(20^\circ\) and \(15^\circ\), respectively and they send the values to the BS. When the BS receives the data values, it calculates the \(S_i\) value for each node \(n_i\). The node which has the smallest \(S_i\) is considered as the CH.

For node \(n_5\):
\[ |15 - 20| = 5; |15 - 22| = 7; |15 - 19| = 4; |15 - 20| = 5 \]
The BS calculates \(S_5 = |15 - 20| + |15 - 22| + |15 - 19| + |15 - 20| = 21.\]

Finally, the BS selects either the node \(n_4\) or node \(n_1\) as a CH. Both nodes minimize the total difference value measured. The other nodes become CMs. During the steady phase, CM nodes send the \(\Delta_i\) value to the CH rather than their complete value. In this example, \(n_3\) sends the sample value of 2 rather than the complete sample value of 22. Note that the compression data sent in our scheme involves sending less coded bits compared to a complete data that is sent in the classical scheme. Therefore, a considerable energy saving is achieved in our scheme as shown in the next section.

### IV. PERFORMANCE EVALUATION

We use TinyOS [15] as a simulation tool. The parameters used for this set of results are presented in Table I.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(E_{fs})</td>
<td>10pJ/bit/m²</td>
</tr>
<tr>
<td>(E_{mp})</td>
<td>0.001nJ/bit/m²</td>
</tr>
<tr>
<td>(F_{elec})</td>
<td>50nJ/bit</td>
</tr>
<tr>
<td>(E_{DA})</td>
<td>5nJ/bit/signal</td>
</tr>
<tr>
<td>Signaling packet length (S)</td>
<td>24bit</td>
</tr>
<tr>
<td>Data value at the center of the event (C)</td>
<td>250°</td>
</tr>
<tr>
<td>Initial energy per node (E_0)</td>
<td>10J</td>
</tr>
<tr>
<td>(T_{\text{event}})</td>
<td>200 second</td>
</tr>
<tr>
<td>(R_{\text{event}})</td>
<td>600meters</td>
</tr>
<tr>
<td>(R_{c})</td>
<td>1000meters</td>
</tr>
<tr>
<td>(R_{e})</td>
<td>4000meters</td>
</tr>
<tr>
<td>Area A</td>
<td>100 \times 100m²</td>
</tr>
</tbody>
</table>

Fig. 3 shows the average energy consumed in the network per unit of time for different number of nodes. In this case the number of simulated events is 20. The results clearly demonstrate that our proposal outperforms the classical scheme. From that figure, we can see that the energy consumption increases with the increase of the number of nodes in the system. Indeed, when the number of nodes in the network is high, the number of nodes that sense the event is also high. Hence, the number of packet transmissions (both control and data packets) is much higher than for the case where just a few nodes are active per event. The main reason for the better performance of the proposed protocol is that only the difference \(\Delta_i\) is transmitted rather than the complete data packet during the steady state. This difference between the classical and the proposed protocol increases for higher network densities. Since in this case, the nodes are closer to each other, which in entails a higher correlation degree among their sensed data. As a result smaller packet size is observed. Conversely, for the classical scheme, since the packet size is fixed, a higher network density increases the number of packets transmitted, consuming thus a lot of energy.

Fig. 4 shows the average energy consumed over time for 60 nodes in the classical, \(CC_{SCR}\) and with using one single
hop to reach the BS. In order to explore the benefits of the clustering architecture, a scenario where all the nodes transmit directly to the BS is presented. For the proposed scheme, all active nodes transmit their initial packet to the BS in order to choose the reference node (note that in this case there is no CH). Then, the BS selects the node that minimizes the data differences as explained in the previous section and then transmits a control packet indicating the ID of the reference node. Then, for the data transmission, the active nodes only transmit their difference $\Delta_i$ directly to the BS. The results demonstrate clearly that $CC_{SCR}$ conserves more energy compared to the classical scheme. Also, it is clear that the choice of the clustering scheme offers more energy savings than the single hop scheme. The gain ratio presented in Fig. 5 may reach up to 11 times more energy conservation than the classical scheme and 119 times more energy conservation than the single hop scheme, which are high benefit ratios.

Fig. 6 shows the average energy consumed for different values of the $T_{event}$ period. Note that increasing $T_{event}$ also increases the period of the steady state phase and thus the number of active nodes. Note that by increasing the number of active node, the energy consumption also increases. Observe for instance that the energy consumption when $R_{event} = 30$ is less than the consumption when $R_{event} = 60$ and 90. In each scenario, we can see that enabling our compression scheme reduces the energy consumption over the network and therefore extends the network lifetime.

Fig. 8 shows the average energy consumed for different values of the $R_{event}$ region. When $R_{event}$ is varied, the number of active nodes per event is modified accordingly. Fig. 7 shows the number of active nodes per event. It can be seen that the average number of active nodes for both the classical and the proposed scheme is approximately the same. Indeed, the proposed mechanism has no impact on the
number of data reported. Therefore, an increase of the energy consumption is observed, as shown in Fig. 8. On the other hand, enabling our compression scheme reduces the energy consumption over the network (see Fig. 8) and therefore extends the network lifetime. It is important to note that the proposed mechanism is particularly energy efficient for high event duration times. This is due to the fact that as the event duration increases, the CMs in the classical scheme transmit many full length packets while for the proposed mechanism, the CMs also transmit many packets but with a much smaller length. This enables a slight increase of energy consumption for the proposed mechanism while for the classical scheme there is an important augmentation in the energy consumption when the event duration increases.

![Fig. 8. Average energy consumption vs $T_{\text{event}}$.](image)

Fig. 8 shows the average energy consumption in $CC_{\text{SCR}}$ when the aggregation technique is enabled at the CH compared to the case where no aggregation is performed. The results show clearly that the aggregation technique conserves more energy. This is expected since when the CH aggregates the data of its CMs, the CH only transmits one single packet to the BS, unlike the case where no aggregation is used since the CH transmits each data value received from CMs.

![Fig. 9. Energy consumed with aggregation and no aggregation.](image)

Fig. 9 shows the average energy consumption in $CC_{\text{SCR}}$ when $T_{\text{event}}$ is 200, 300, and 400 for both classical and proposed protocols. The figure clearly shows that the proposed mechanism conserves more energy compared to the classical scheme.

V. CONCLUSION

In this paper, we have proposed a new compression technique in event-driven applications for WSNs. The proposed clustering scheme is based on selecting the node that reduces the packet size among all active nodes in the system. The BS selects the node which minimizes the total amount of data as a CH, therefore it increases the efficiency of the compression technique by sending only the difference, rather than the complete data value to the CH. By varying different parameters of the system, simulation results show that considering the spatial correlation in the communication of WSNs achieves significant energy conservation compared to a classical clustering scheme. The ratio benefit may reach 11 time compared to the classical scheme.

REFERENCES