Scientific Research Publishing

Information Driven Data Gathering for Energy Efficient Wireless Sensor Network

K. Anna Raja[1], L. R. Karlmarx[2]

[1]Department of Electrical and Electronics, PTR College of Engineering and Technology, Madurai, India
[2]Department of Electronics and communication, Thiagarajar College of Engineering, Madurai, India
Email: annaraja.k@gmail.com, lrkarlmarx@tec.edu

Abstract

Large scale dense Wireless Sensor Networks (WSNs) have been progressively employed for different classes of applications for the resolve of precise monitoring. As a result of high density of nodes, both spatially and temporally correlated information can be detected by several nodes. Hence, energy can be saved which is a major aspect of these networks. Moreover, by using these advantages of correlations, communication and data exchange can be reduced. In this paper, a novel algorithm that selects the data based on their contextual importance is proposed. The data, which are contextually important, are only transmitted to the upper layer and the remains are ignored. In this way, the proposed method achieves significant data reduction and in turn improves the energy conservation of data gathering.

Keywords

Wireless Sensor Network, Energy Efficiency, Data Reduction, Temporal Correlation

1. Introduction

Wireless Sensor Networks (WSNs) [1]-[4] can be defined as a cooperative network of small, battery-operated. These networks have two functions: the main goal of this network is monitoring their surroundings for local data and for forwarding the gathered data to a sink node using typically multihop communication. Then this sink node is liable for processing all the acknowledged data from numerous source nodes and writing them to an observing facility. This network architecture permits a number of innovative observing based applications in numerous areas such as ecological, medical, engineering, and military.

One of the main confines of the WSNs is the battery activated environment of their

sensor nodes. It creates this kind of network exceedingly energy guarded. Therefore, energy convertible is one of the key concerns of rules and applications in WSNs. As communication between the nodes is one of the key sources of energy intake, most of the rules in WSNs try to evade or delay communication, until it is certainly essential [5]-[11]. However, by undertaking so, invalid and/or unfinished information is usually attained by the sink node varieties, the fundamental claim neither dependable nor beneficial.

In this present work, besides delaying or evading communication, clarification is projected which acquires the most out of every vital communication. It is completed by manipulating both the great density of WSN nodes and the experimental resemblance of nearby congregated data. This methodology not only outspreads the lifespan of a WSN but also offers near real-time evidence about the observed area.

2. Related Work

Generally, instead of having all sensor nodes reporting the same data, it is more efficient to select a few representative nodes to inform the sink node about the detected event [12]. A representative node reports the event information of a given area on behalf of a group of nodes and it collects similar information in the same area.

Sensor readings about the environment are typically periodic [13]; a result, the time-ordered sequence of sensed data constitutes a time series. Due to the nature of the physical phenomenon, there is a significant temporal correlation among each consecutive observation of a sensor node and gathered data are usually similar over a short-time period [14]. Hence, in this case, if the current reading is within an acceptable error threshold regarding the last reported reading sensor nodes do not need to transmit their readings. The sink node can just assume that any unreported data are unchanged from the previously received ones. The degree of correlation among successive sensor quantities may fluctuate depending on the characteristics of the occurrence.

Akyildiz *et al.* [15] analysed the bonding between the reliability of event detection and spatial location of the sensor nodes in the event area. Their solution estimates the number of sensor nodes (representative nodes) required to send the detected event to the sink in order to get reliable event information. Each representative node represents a spatially correlated group of nodes. Though their solution has achieved overall energy gain, it has failed to consider the remaining energy during the selection of the representative nodes. It is an assumption that it should not be neglected in a WSN because of hardware constraints. If a representative node works in the correlation region for a long period of time, it will spend more energy because of the number of transmitted messages compared to the other nodes.

Yoon and Shahabi [16] have proposed a new mechanism for spatial correlation in WSNs. This mechanism, called Clustered Aggregation Technique (CAG). It creates clusters of nodes with similar sensing values and only one node inside the cluster informs its reading to the Sink node whereas the other nodes ignore their readings. The CAG algorithm is classified into two phases: query and response. In the query phase,

the data-centric clusters are created according to a user-specified error thresholds. Nodes that have sensed values smaller than this threshold belong to the same cluster. In the response phase, just one node per cluster (cluster-head) sends its sensed value to the sink node by notifying the detected event. The authors have showed that the proposed mechanism reduces significantly the number of transmitted messages during the data collection. However, during the query phase, the CAG algorithm uses a flooding-based protocol to disseminate the query to all sensor nodes, and it is not needed in most scenarios. Moreover, the maintaining of the data-centric clusters is a difficult problem.

3. Methodology

Temporal correlation is the change pattern of current sensor readings and it is equal or similar to the reading observed at previous times. In the proposed solution, energy is not saved by delaying or suppressing messages. Where as it is saved by combining the correlated information in order to make a better use of the data communication. The proposed algorithm, is named as IDDG (information driven data gathering). In that algorithm, sensor nodes are clustered under a spatial correlation approach where as the leader and representative nodes process a temporal suppression technique. The leader node generates a representative value for the whole cluster based on the data received by the representative nodes. These form a subset of all nodes sensing the same event.

One of the key aspects of the proposed solution is that it dynamically adjusts itself according to both the event characteristics and the residual energy of the sensing nodes, which are commonly ignored by current proposed solutions. Thus, in the IDDG algorithm 1) residual energy of sensor nodes is balanced, 2) energy consumption is reduced by eliminating redundant notifications, and 3) both the number of representative nodes and error threshold are dynamically adjusted according to the event characteristics and accuracy requirements. In order to provide a better understanding of the behavior of the proposed algorithm in different scenarios, we present an extensive set of experiments are presented and they show the need for new algorithms (specially for real-time applications). They also clearly show the good performance of the proposed algorithm compared to the related approaches.

4. Temporal Correlation

Data collected from a specific sensor at different time intervals may be correlated, if the set of collected data varies. This is called as temporal correlation. Because of the environment of the physical occurrence, there is an important temporal correlation between each successive reflection of a sensor node. For example, in a daily sampling of temperature performed at each minute, the temperature may not change significantly. In this case, it is not necessary to report the new sampling at each minute, since the last reported sampling corresponds to the actual one.

Vuran *et al.* [17] proposed a new framework to create data centric protocols that explore the nature of the physical phenomenon observed by a WSN. The main motto of the framework is to incorporate temporal correlation among consecutive observations

of the phenomenon is order to reduce the cost of communication. The authors have also explored spatial correlation by showing that the nearby nodes tend to have the same observed data. The proposed framework can be used in two ways: 1) to develop efficient protocols, and 2) to develop reliable sensed information reporting in WSN.

Deligiannakis and Kotidis have proposed a framework based on temporal correlation that uses a Self-Based Regression (SBR) algorithm [18] to decrease the number of transmitted messages required to monitor a physical phenomenon. Processing the observed data before sending them to the sink node is the function of the SBR algorithm. The framework stores the sensed information in buffer. When the buffer is full, the SBR algorithm processes the data to find representative information. The authors identified that by just sending the representative information, the sink node can reconstruct the observed event without losing accuracy. Whereas, the main drawback of such an approach is the waiting time until the buffer fills up. In this case, the sink node can receive outdated information about the sensed event.

Definition (Temporal suppression). Each source node keeps the last reported reading. When current reading (S_{new}) is available, S_{new} is compared to the last reported reading (S_{old}). The current reading of a source node is reported if the given relative threshold is greater than the temporal coherency tolerance (*tct*), *i.e.*

$$\left(\frac{\left| \left(S_{new} - S_{old} \right) \right|}{\left(S_{old} \right)} \right) \times 100 > tct \,,$$

where *tct* is the percentage of temporal coherency tolerance. Otherwise, the value Rnew is suppressed.

Overview of the PROPOSED Algorithm

The main idea of the proposed IDDG algorithm is to manage the energy consumption of nodes that detected an event by eliminating redundant notifications. The proposed algorithm considers the following roles to perform data routing:

Member node: A node that currently detects one or more events. When its sensed data are redundant, it will not report the gathered data.

- Representative node: A node that detects an event and reports the gathered data to a coordinator representing not only itself but all nearby nodes with similar readings, while applying temporal suppression.
- Coordinator node: A node that detects the event and is responsible for gathering all the event data sent by representative nodes. It processes the received data and sends the result towards the sink node.
- Relay node: A node that forwards data towards the sink node.
- Sink node: The gateway between the WSN and the monitoring facility.

The proposed algorithm uses shortest routes (in Euclidean distance) in two different levels for forwarding the gathered data towards the sink node. In the first level, representative nodes use shortest routes to forward data towards the coordinator node. In the second level, the coordinator nodes use shortest routes to forward the data towards the sink node. Figure 1 shows two examples of the routing structure obtained by the
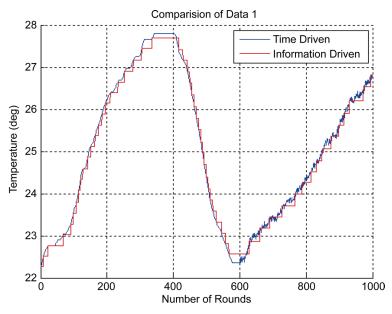
Figure 1. Comparison of data.

proposed algorithm (the gray field indicates the event area, the cells represent the regions of correlation and the red dotted line shows the shortest route).

The key objective of the projected algorithm is to decrease energy consumption in data gathering, while preserving both data accuracy and real-time reporting. To achieve this goal, IDDG dynamically changes the size of the correlation region and the value of the coherency tolerance according to the event characteristics. For this, an event area is divided into cells. Each cell defines a correlation region and nodes within each cell are assumed to be spatially correlated. Only one node within a cell notifies the sensed information, if and only if, the given relative error threshold is greater than the temporal coherency tolerance. This last node is the representative node of the cell. The value of the representative selection probability is then calculated by multiplying the cluster count factor with the relative energy level. Formally,

$$s_i = f_c * \frac{e_i(t) * \left(\left\| nbr(p_i) \right\| + 1 \right)}{e_i(t) + \sum_{p_j \in nbr(p_i)} e_j(t)}.$$

This enables us to favor nodes with higher energy levels for representative node. Here $e_i(t)$ is the energy level of node i at the instant t, $nbr(p_i)$ indicates the number of neighbors of node i. f_c is the representative fraction.

Cells are independent from each other. Hence, the change of representative nodes in one cell does not require any reconfiguration. The change of a representative node in each cell is performed to balance the energy consumption of spatially correlated nodes, while temporal suppression is applied to reduce the reporting of redundant data. During cell formation, a representative node pi takes the following concrete actions. It first creates a correlation matrix C_i such that for any two nodes in the cluster C_i, say p_u and p_v, $C_i[p_u, p_v]$ is equal to the correlation between the series $D_i[p_u]$ and $D_i[p_v]$, formally

$$\frac{\left(D_i\left[p_u\right]-E\left[D_i\left[p_u\right]\right]\right)*\left(D_i\left[p_v\right]-E\left[D_i\left[p_v\right]\right]\right)^{\mathrm{T}}}{L*\sqrt{Var\left(D_i\left[p_u\right]\right)}*\sqrt{Var\left(D_i\left[p_v\right]\right)}}$$

where, L is the length of the series and T represents matrix transpose.

Since correlation regions are independent, their resizing does not require any additional communication among the nodes within the event areas in order to compute the new cell they belong to. Furthermore, each node performs temporal suppression locally without communicating with its neighbors. The proposed spatio-temporal correlation approach is adaptive and scalable regarding events of different intensities, as it will be shown during its evaluation.

5. Results and Discussion

The proposed IDDG algorithm is evaluated on MATLAB platform. The evaluation involves a network of 100 nodes that are randomly deployed over a region of 100×100 m space. All the nodes are powered with a uniform battery source and are equipped with RF transceiver of same power level. The sensor nodes are defined for collecting temperature data from the environment.

Figure 1 depicts the original and information driven based collected data. The data are communicated only when there is a change in data direction or data pattern. As the measured data are highly self correlated, the information driven approach reaps the data reduction benefit to the fullest extent. From the Figure 1 it is observed that the there is a minimal deviation between the collected data and the original data.

Figure 2 indicates the actual deviation between the collected data and original data at every data collection round. It is proved that the collected data have minimal deviation
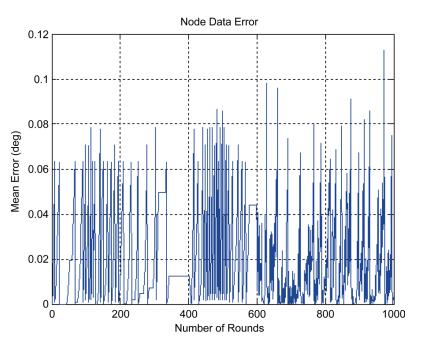

Figure 2. Node data error.

from the original data. The deviation is also limited to the threshold error value set by the application.

Figure 3 shows the comparison of lifetime between time driven and information driven approaches. From the figure, it is observed that the proposed information driven approach has elongated the lifetime of the network to multiple folds. The time driven approach steady state lifetime of the network is approximately 200 - 300 rounds. Whereas in the information driven approach, steady state lifetime is increased to 1200 rounds which is six times larger than the existing time driven approach. In the same way, the complete lifetime of the network is also improved. The complete lifetime of a time driven approach is only 800 rounds. In the information driven approach, the same is elongated to 4500 rounds. This longer life of the network is attributed to the information driven data collection where only the important data are transmitted, and the rest are not transmitted. Since the critical data are always communicated, this approach maintains optimal data quality with high amount of energy conservation.

Figure 4 indicates an energy consumption of the network under two different data gathering approaches. Since time driven data gathering involves continuous data transmission, the energy consumption is also high. At the end of 1000 data collection rounds, the total energy consumption of network is above 0.0325 J. At the same round, if the data are collected through information driven approach, the total energy consumption is kept between 0.005 J and 0.01 J. This massive reduction in energy consumption happens due to the high level of data reduction in the information driven approach.

In Figure 5, the amount of data transmission between the two approaches is compared. The time driven approach transmits 60 kb of data over the span of 1000 rounds. Over the same time period, the proposed approach has reduced the total transmission
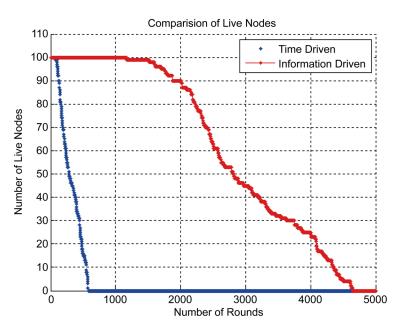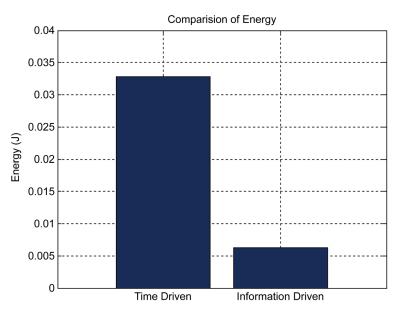

Figure 3. Comparison of live nodes.
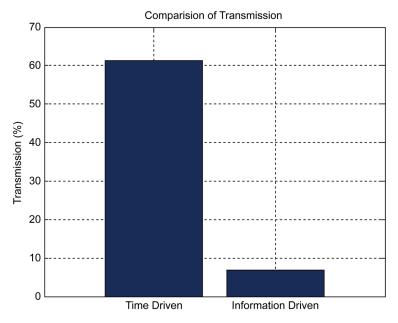
Figure 4. Comparison of energy.

Figure 5. Comparison of transmission.

to the meagre data transmission of 8 kb. This reduction in data transmission also results in reduced data collision and improves the data throughput of the network. The proposed approach also necessitates minimal bandwidth requirement for a given scale of the network.

6. Conclusion

The proposed IDDG, an algorithm for energy-aware data forwarding in WSNs, takes full advantage of temporal correlation mechanisms to save energy while maintaining

real-time, accurate data report towards the sink node. In the current literature of temporal correlation algorithms, most of the proposed studies do not consider the energy dissipation during data collection to better choose the representative nodes. In this work, there is a deeper manager, an energy-aware temporal correlation mechanism in which nodes, detecting the equivalent event, are dynamically assembled in correlated constituencies and a representative node is nominated at every correlation region for spotting the occurrence. The complete region of sensors per incident is effectively a set of evocative nodes which achieves the task of data collection and temporal correlation.

References

[1] Akyildiz, I.F., Su, W., Sankarasubramaniam, Y. and Cyirci, E. (2002) Wireless Sensor Networks: A Survey. *Computer Networks*, **38**, 393-422.
http://dx.doi.org/10.1016/S1389-1286(01)00302-4

[2] Estrin, D., Girod, L., Pottie, G. and Srivastava, M. (2001) Instrumenting the World with Wireless Sensor Networks. 2001 *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, 7-11 May 2001, 2033-2036.

[3] Pottie, G.J. and Kaiser, W.J. (2000) Wireless Integrated Network Sensors. *Communications of the ACM*, **43**, 51-58. http://dx.doi.org/10.1145/332833.332838

[4] Romer, K. and Mattern, F. (2004) The Design Space of Wireless Sensor Networks. *IEEE Wireless Communications*, **11**, 54-61. http://dx.doi.org/10.1109/MWC.2004.1368897

[5] Villas, L.A.,. Guidoni, D.L., Araújo, R.B., Boukerche, A. and Loureiro, A.A. (2010) A Scalable and Dynamic Data Aggregation Aware Routing Protocol for Wireless Sensor Networks. *Proceedings of the* 13*th ACM International Conference on Modeling, Analysis, and Simulation of Wireless and Mobile Systems*, Bodrum, Turkey, 17 - 21October 2010, 110-117.

[6] Olariu, S., Xu, Q. and Zomaya, A. (2004) An Energy-Efficient Self-Organization Protocol for Wireless Sensor Networks. *Proceedings of the* 2004 *Intelligent Sensors, Sensor Networks and Information Processing Conference* (*ISSNIP*), Melbourne, 14-17 December 2004, 55-60. http://dx.doi.org/10.1109/issnip.2004.1417437

[7] Boukerche, A. (2008) Algorithms and Protocols for Wireless Sensor Networks. Wiley-IEEE Press.

[8] Nakamura, E.F., de Oliveira, H.A.B.F., Pontello, L.F. and Loureiro, A.A.F. (2006) On Demand Role Assignment for Event-Detection in Sensor Networks. 11*th IEEE Symposium on Computers and Communications*, Washington DC, 26-29 June 2006, 941-947.

[9] Villas, L.A., Boukerche, A., Araujo, R.B. and Loureiro, A.A. (2009) A Reliable and Data Aggregation Aware Routing Protocol for Wireless Sensor Networks. *Proceedings of the* 12*th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, Tenerife, Canary Islands, Spain, 27-30 October 2009, 245-252.

[10] Fan, K.-W., Liu, S. and Sinha, P. (2007) On the Potential of Structure-Free Data Aggregation in sensor Networks. *IEEE Transactions on Mobile Computing*, **6**, 929-942.
http://dx.doi.org/10.1109/TMC.2007.1011

[11] Leandro, R.B.A., Villas, A., Boukerche, A. and Loureiro, A.A.F. (2010) Highly Dynamic Routing Protocol for Data Aggregation in Sensor Networks. 15*th IEEE Symposium on Computers and Communications*, Riccione, 22-25 June 2010, 496-502.

[12] Incel, O.D. and Krishnamachari, B. (2008) Enhancing the Data Collection Rate of Tree-Based Aggregation in Wireless Sensor Networks. 5*th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, San Francis-

co, 16-20 June 2008, 569-577. http://dx.doi.org/10.1109/sahcn.2008.74

[13] Lu, G., Krishnamachari, B. and Raghavendra, C.S. (2007) An Adaptive Energy-Efficient and Low-Latency MAC for Tree-Based Data Gathering in Sensor Networks. *Wireless Communications and Mobile Computing*, **7**, 863-875. http://dx.doi.org/10.1002/wcm.503

[14] Pham, N.D., Le, T.D. and Choo, H. (2008) Enhance Exploring Temporal Correlation for Data Collection in WSNs. *IEEE International Conference on Research, Innovation and Vision for the Future*, Ho Chi Minh City, 13-17 July 2008, 204-208. http://dx.doi.org/10.1109/rivf.2008.4586356

[15] Akyildiz, I.F., Vuran, M.C. and Akan, B. (2004) On Exploiting Spatial and Temporal Correlation in Wireless Sensor Networks. *Proceedings of WiOpt* 2004: *Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks*, University of Cambridge, 24-26 March 2004, 71-80.

[16] Yoon, S. and Shahabi, C. (2005) Exploiting Spatial Correlation towards an Energy Efficient Clustered Aggregation Technique (CAG) [Wireless Sensor Network Applications]. *IEEE International Conference on Communications*, **5**, 3307-3313.

[17] Vuran, M.C. and Akyildiz, I.F. (2006) Spatial Correlation-Based Collaborative Medium Access Control in Wireless Sensor Networks. *IEEE/ACM Transactions on Networking*, **14**, 316-329. http://dx.doi.org/10.1109/TNET.2006.872544

[18] Deligiannakis, A., Kotidis, Y. and Roussopoulos, N. (2004) Compressing Historical Information in Sensor Networks. *Proceedings of the* 2004 *ACM SIGMOD International Conference on Management of Data*, Paris, France, 13-18 June 2004, 527-538. http://dx.doi.org/10.1145/1007568.1007628