Volume 5, Issue 1, January 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A Survey on Data Collection Techniques in Wireless Sensor Networks

[1]Anamika Pandey*, [2]Suman Srivastava, [3]Ashok Shaky
[1, 2] Student of Master of Technology, [3]Assistant Professor
Computer Science & Engineering, UPTU, Lucknow, UP, India

Abstract— In wireless sensor network the most challenging task is to reduce energy consumption. Due to resource restricted sensor nodes, it is valuable to reduce the amount of data transmission so that the average sensor lifetime and the overall bandwidth utilization are improved. The main goal of data collection algorithms is to gather and aggregate data in an energy efficient manner so that network lifetime is enhanced. In this paper to reduce energy consumption we discuss the data collection approaches and various performance measures of the data collection in the network.

Keywords— Data Collection, Energy Consumption, Lifetime, Sink, Wireless Sensor Network.

I. INTRODUCTION

Wireless sensor networks are usually composed of hundreds or thousands of inexpensive, low-powered sensing devices with limited memory, computational, and communication resources [5, 6]. The sensors coordinate among themselves to form a communication network such as a single multi-hop network or a hierarchical organization with several clusters and cluster heads. The sensors could be scattered randomly in harsh environments such as a battlefield or deterministically placed at specified locations. The sensors periodically sense the data, process it and transmit it to the base station. The frequency of data reporting and the number of sensors which report data usually depends on the specific application.

The main objective of data aggregation is to increase the network lifetime by reducing the resource consumption of sensor nodes (such as battery energy and bandwidth). Data aggregation is defined as the process of aggregating the data from multiple sensors to eliminate redundant transmission and provide fused information to the base station. Data aggregation usually involves the fusion of data from multiple sensors at intermediate nodes and transmission of the aggregated data to the base station (sink).

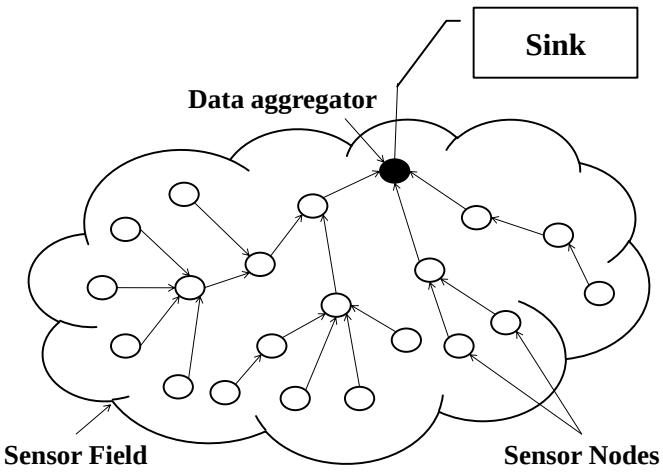

Figure 1 Data Aggregation in Wireless Sensor Network

An example of data aggregation scheme is presented in above figure where a group of sensor nodes collect information from a target region. When the base station queries the network, instead of sending each sensor node's data to base station, one of the sensor nodes, called data aggregator, collects the information from its neighboring nodes, aggregates them (e.g., computes the average), and sends the aggregated data to the base station.

II. CLUSTERING

Clustering is a well-established technique for reducing data collection costs in WSNs. In this technique, sensor nodes are grouped into disjoint sets, with each set managed by a designated Cluster-Head (CH), selected from among the sensor

nodes. The cluster members send their collected observations (which are likely to be highly correlated) to their CH. The CH suppresses the local redundancies and communicates the compressed data to the sink possibly via multi-hop transmission.

III. DATA COLLECTION TECHNIQUES

A. Directed Diffusion

This is a data centric approach, in this method [3] all the communication is for the named data. In directed diffusion method sink broadcast its request to its neighbour. The neighbour cached the interests and the neighbour data which matching to the interest will respond back to sink. The human operator's query would be transformed into an interest that is diffused towards nodes in regions an interest that is diffused towards nodes in regions X or Y. When a node in that region receives an interest, it activates its sensors which begin collecting information about pedestrians. When the sensors report the presence of pedestrians, this information returns along the reverse path of interest propagation. Intermediate nodes might aggregate the data, e.g., more accurately pinpoint the pedestrian's location by combining reports from several sensors. An important feature of directed diffusion is that interest and data propagation and aggregation are determined by localized interactions (message exchanges between neighbours or nodes within some vicinity).

Directed diffusion is significantly different from IP-style communication where nodes are identified by their endpoints, and inter-node communication is layered on an end-to-end delivery service provided within the network. In this approach the number of transmissions will be reduced because of interested neighbor will be responded. The disadvantage of this method is that it is not suitable for long term large scale wireless sensor networks.

B. Chain Construction Algorithms

The effectiveness of chain based data aggregation protocols depends largely on the construction of an energy efficient chain. Du et al. [7] have developed an energy efficient chain construction algorithm which employs insertion operations to add the least amount of energy consumption to the whole chain. The main focus is on energy efficient all to all broadcasting in sensor networks. A multiple chain scheme has been proposed which divides the whole network into four regions centered at the node that is closest to the center of the sensing region. For each region, a linear chain is constructed which ends at the center node. The multiple chain schemes aims to decrease the total transmission distance for all-to-all broadcasting.

In the greedy chain construction algorithm proposed in [10], the process starts with the farthest node from the sink. This node is the head of the chain. At each step, a nonchain node which is closest to the chain head is selected and appended to the chain as the new head. The procedure is repeated until all nodes are in the chain. This process does not necessarily minimize the total transmission energy. The authors in [7] have proposed a minimum total energy algorithm which constructs a chain with minimum Σd^2 where d is the distance between two adjacent nodes in the chain. The chain construction starts with the node farthest from the sink as the leader. At each step, a new node is inserted such that Σd^2 of the current chain with the new node increases to the minimum possible extent compared to the old chain. This new node becomes the leader. The algorithm has a complexity of $O(n3)$ where n is the total number of nodes.

C. Probabilistic Model

This paper [4] proposes a robust approximate technique called “Ken” that uses replicated dynamic probabilistic models to minimize communication from sensor nodes to the network’s PC base station. The basic idea is to maintain a pair of dynamic probabilistic models over the sensor network attributes, with one copy distributed in the sensor network and the other at a PC base station. At every time instance (i.e., with a frequency f), the base station simply computes the expected values of the sensor net attributes according to the model and uses it as the answer to the “SELECT * FREQ f” query. This requires no communication. The sensor nodes always possess the ground truth, and whenever they sense anomalous data – i.e., data that was not predicted by the model within the required error bound – they proactively route the data back toward the base station. As data is routed toward the base station, spatial correlations among the reported data are used to further lower communication. Using these techniques, all user-visible readings are guaranteed to be within a fixed error bound from the measured readings, even though very few readings are communicated to the base station.

An attractive feature of the Ken architecture is that it naturally accommodates applications that are based on event reporting or anomaly detection; these include fire-alert-and response and vehicle tracking. In these scenarios, the sample rate is typically quite high, but the communication rate should remain quite low under most circumstances. The model reflects the expected “normal” state of the environment being monitored; anomalies result in reports being pushed to the base station for urgent handling by infrastructure logic. In addition to naturally supporting these applications, Ken enhances them with additional functionality: the ability to support interactive query results with well-bounded approximate answers. In essence, approximate data collection and event detection become isomorphic. Advantage of this method is that the readings which are visible to the user are within fixed error bound. Disadvantage of using this method is that it is robust to communication failure that is, if the data lost during transmission it cannot be recovered.

D. A Simple Single Cluster Model Using Data Correlation

This paper [1] proposes and evaluates a novel WSN clustering strategy that exploits data correlation. That is, the nodes within each cluster have strong internal correlation, while the inter-cluster data dependence is negligible. To this end,

there is carefully analysing the mutual effect of cluster size and the distance from the sink on reducing total network energy consumption. Although it is computationally difficult to find optimal-sized clusters, this paper proposes a model to obtain a near-optimal solution for forming energy-efficient clusters in the network. In a nutshell, the main contributions of this paper are as follows:

- The paper develop a model to incorporate the effect of spatial data correlation while forming energy efficient clusters in the network.
- Unlike conventional clustering algorithms that result in uniform clusters of almost the same size, this model advocate heterogeneous-sized clusters in the network, where the clusters further from the sink are larger than those located close to the sink.

In this paper, we studied the joint effects of data correlation, distance, and network density on forming optimal sized clusters that require less power than conventional approaches. This paper showed that unlike most of the existing clustering approaches that produce uniform clusters throughout the whole network, heterogeneous-sized clusters are more energy efficient in WSNs with spatial data correlation.

E. Query Model

In this paper [8] we have studied the two most important parts of data communication in sensor networks- query processing, data aggregation and realized how communication in sensor networks is different from other wireless networks. Wireless sensor networks are energy constrained network. Since most of the energy consumed for transmitting and receiving data, the process of data aggregation becomes an important issue and optimization is needed. Efficient data aggregations not only provide energy conservation but also remove redundancy data and hence provide useful data only.

COUGAR approach proposes a query layer to support aggregate queries. With the interface provided, the clients can issue queries without knowing how the results are generated, processed and returned by the sensor network to them. The query layer processes declarative queries and generate a cost effective query plan. They follow a database approach to design a query interface for sensor networks. The view of cost is different for sensor networks. The major factor under consideration is the communication cost, involving the cost of routing the queries and aggregating data over the sensor networks. TAG also proposes a query model for supporting aggregate queries.

TAG and COUGAR are tightly coupled with the underlying aggregation schemes. WSNs Proposes a Query Agent that provides application independent query interface and an API support to map the user specified queries to lower level semantics corresponding to underlying routing and aggregating protocols. It supports different communication models - anycast, unicast, multicast and broadcast. Query agent will support a wide variety of routing and aggregation protocols selecting the best combination based on the type of the query.

F. Derivative Based Prediction (DBP) Model

This paper [11] investigates in practice whether

i) model-driven data acquisition works in a real application;

ii) the energy savings it enables in theory are still worthwhile once the network stack is taken into account. Authors do so in the concrete setting of a WSN-based system for adaptive lighting in road tunnels. The novel modelling technique, Derivative-Based Prediction (DBP), suppresses up to 99% of the data reports, while meeting the error tolerance of application. DBP is considerably simpler than competing techniques, yet performs better in real setting. Experiments in both an indoor testbed and an operational road tunnel show also that, once the network stack is taken into consideration, DBP triples the WSN lifetime—a remarkable result per se, but a far cry from the aforementioned 99% data suppression.

DBP is based on the observation that, in the application, the trends of the sensed values in short and medium time intervals can be accurately approximated using a linear model. Even though this idea has appeared in previous works, there is a key difference to author's approach: while previous studies compute models that aim to reduce the approximation error to the data points in the recent past, DBP aims at producing models that are consistent with the trends in the recently-observed data. DBP initialization consists of a learning phase, gathering enough data to produce the first model. The learning phase involves m data points; the first and the last l we call edge points. The model is linear and is computed as the slope δ of the segment that connects the average values over the l edge points at the beginning and end of the learning phase. This computation resembles the calculation of the derivative, hence the name Derivative-Based Prediction. It is interesting to note that the computation of this prediction is not only very simple, and therefore appealing for implementation on resource-scarce nodes, it also mitigates the problem of noise and outliers.

The first DBP model generated is then sent to the sink, along with its last data point. From that point on, each node buffers a sliding window of the last m data points sampled from its sensor. Upon sampling a point, the "true" value sensed is compared to the "predicted" one computed by DBP according to the current model, i.e., following the slope δ. If the sensor reading is within a value tolerance ε_V w.r.t. the model, no action is required: the sink will automatically generate a new value that is an acceptable approximation of the real one. Otherwise, if the readings continuously deviate from the model for more than ε_T time units, a new model must be recomputed. This is accomplished by using the last m data points in the buffer; the resulting model is transmitted to the sink along with the last data point.

IV. CONCLUSIONS

This paper provides an effective presentation on various data collection techniques that can be applied to wireless sensor networks. By applying these data collection techniques we can significantly reduce the number of messages transmitted.

Thereby redundancy will be reduced. Without redundancy energy consumption will be low and the lifetime of wireless sensor networks increases.

REFERENCES

[1] Ali Dabirmoghaddam, Majid Ghaderi and Carey Williamson,2010.Cluster-based correlated data gathering in wireless sensor networks.*18th Annual IEEE/ACM International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems.*

[2] Anamika Pandey et al. August 2014.Efficient technique for collecting data in wireless sensor network. *International Journal of Science and Engineering Research,*Vol. 5, Issue 8.

[3] C. Intanagonwiwat, R. Govindan, and D. Estrin,2000.Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks.*Proc. MobiCom.*

[4] D.Chu, A. Deshpande, J.M. Hellerstein, and W. Hong,2006.Approximate data collection in sensor networks using probabilistic models.*Proc. 22nd Int'l Conf. Data Eng. (ICDE'06).*

[5] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E.Cayirci,2002. A survey on sensor networks. *IEEE Commun. Mag. 40 (8) (2002)102–114.*

[6] J. Yick, B. Mukherjee, D. Ghosal, 2008.Wireless sensor network survey.*Comput. Networks 52 (12) (2008) 2292–2330*.

[7] K. Du, J. Wu and D. Zhou, April 2003.Chain-based protocols for data broadcasting and gathering in sensor networks.*International Parallel and Distributed Processing Symposium*.

[8] Nandini. S. Patil et al, 2010.Data Aggregation in wireless sensor network. *IEEE International Conference on Computational Intelligence and Computing Research*.

[9] R. Rajagopalan and P.K. Varshney,2006. Data aggregation techniques in sensor networks: a survey" *IEEE Comm. Surveys & Tutorials, vol. 8, no. 4, pp. 48–63.*

[10] S. Lindsey, C. Raghavendra, and K.M. Sivalingam,September 2002.Data gathering algorithms in sensor networks using energy metrics. *IEEE Trans. Parallel and Distributed Systems*, vol. 13, no. 9, pp. 924-935.

[11] Usman Raza, Alessandro Camerray, Amy .L Murphy, Themis Palpanasy and Gian Pietro Picco,2012. "What does Model –driven data acquisition really achieve in wireless sensor network?".

