International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Survey on Data Collection Methods for Wireless Sensor Network

Soumya Dath G

Assistant Professor, Department of Information Science & Engineering, GSSSIETW, Mysuru

Abstract: ***The rapid advancement in processor, memory, and wireless technology has enabled the development of distributed networks of small, inexpensive nodes that are capable of sensing, computation, and wireless communication. Sensor networks of the future are envisioned to revolutionize the paradigm of collecting and processing information in diverse environments. Preserving the scarce energy of sensor nodes during data collection is always one of the most crucial problems in wireless sensor networks. This paper presents the overview, characteristics, functionality, benefits and limitations and makes their comparative analysis and analysis of their performance. The objective is to make observations about how the performance of these methods can be improved.***

Keywords: Wireless Sensor Networks (WSN), Data Collection, Data Collection Methods, Energy efficient sensor nodes.

1. Introduction

Wireless Sensor Networks (WSN) consists of many sensor enable nodes which uses batteries as energy resource and distributed in an environment. These tiny sensor nodes, which consist of components for communication, data processing, and sensing data, result in the idea of sensor networks based on collaborative effort of a large number of nodes. Such sensor nodes could be deployed in the area of military, industry, science, and home applications such as health care, disaster recovery, transportation, security, industrial warfare, and building automation, and even space exploration. Among a large variety of applications, current environment monitoring is one of the key areas in wireless sensor networks and in such networks, can query the physical quantities of the environment.

In fact, a typical wireless sensor network is composed of a large number of sensor nodes, which are randomly spread over the interested area, picking up the signals by all kinds of sensors and the data acquiring unit, processing and transmitting them to a node which is called sink node. The sink node requests the information which is sensed by sending a query throughout the sensor field. This query is received at sensor nodes (or sources). When the node finds data matching the query, the data (or response) is routed back to the sink. For example, if the sensors nodes be in a tree like structure, the base station roles as the root of the tree and each node will have a parent. Therefore, the data items can be transmitted hop by hop from the leaf nodes to the root.

In WSNs, to reduce the amount of bytes required to code the different pieces of information the data compression refers to the use of compression techniques and, thus, the traffic load which needs to be processed within the network. As the sensor nodes are small and battery enable devices, they have limited energy which should be used precisely. Thus, the scarce sensor resources (in particular, the battery power) are easily over consumed. Thus, the key challenge in such phenomena monitoring is conserving the sensor energy, so as to maximize their lifetime. Most of the approaches tried to response to this challenge and this will be continue to gain a better solution. Wireless sensor networks enable people to observe details of real-world phenomena in both temporal and spatial dimensions.

2. Data Collection

Data collection is the fundamental function of WSNs, but also a challenging task due to limited resources of those tiny sensor nodes. Among all activities of a sensor node, it is well-known that data communication causes the maximum energy drain. Therefore, data collection methods should avoids abundant communication overhead yet keeps the data quality, becomes the effective method to achieve a longer network lifetime of WSNs for data-driven applications, which require sensor nodes to perform data sampling and transmit data to Sink periodically, such as environmental monitoring.

To conserve the finite resources, such as energy, network bandwidth and CPU usage, extensive research work has been done and various energy-saving protocols and algorithms have been proposed for these data-driven applications. Among of these work, model-driven data acquisition has been proved to be an effective approach to reduce communication without compromising data quality, not only in theory but also in practice. BBQ and Ken approximate the data with user-specified confidence by keeping statistical model local and global in sync.

As typical time series data, sensing readings can be modeled and analyzed with methods of time series analysis. ARIMA model in energy efficient data collection for WSNs, Sink node builds suitable ARIMA model for each sensor node at first. During the adaptive data collection phase, both node and Sink perform forecasting for next sampling value with the same model, and Sink keeps the prediction value as sampling data if it does not receive the real value from sensor node, which sends the actual value only when the prediction error is beyond a pre-defined error-tolerance threshold. With models built by Sink, large amount of data communication are triggered.

Compared to ARIMA, AR model is more lightweight but still offers competitive prediction accuracy. PAQ and SAF both adopt AR model to capture the underlying trend of data distribution. With dual-prediction at both node and Sink, redundant data communications are suppressed and energy is conserved. Furthermore, PAQ has proposed monitoring algorithm to maintain a local dynamic model to adapt to the changing phenomenon. Similar works relying on linear regression to perform data collecting are also presented.

3. Data Collection Methods

a) The Barbie-Q (BBQ) system: Employs multi-variate Gaussian distributions for sensor data acquisition. BBQ maintains a multi-dimensional Gaussian probability distribution over all the sensors. Data is acquired only as much as it is required to maintain such a distribution. Sensor data acquisition queries specify certain confidence that they require in the acquired data. If the confidence requirement cannot be satisfied, then more data is acquired from the sensors, and the Gaussian distribution is updated to satisfy the confidence requirements. Finally the drawback of this system is expensive communication cost to train the data.

b) For reducing the communication cost, the Ken framework employs a similar strategy as PRESTO. Although there is a key difference between Ken and PRESTO. PRESTO uses a SARIMA model; this model only takes into account temporal correlations. On the other hand, Ken uses a dynamic probabilistic model that takes into account spatial and temporal correlations in the data. Since a large quantity of sensor data is correlated spatially, and not only temporally, Ken derives advantage from such spatio-temporal correlation. The Ken framework has two types of entities, sink and source. Their functionalities and capabilities are similar to the PRESTO proxy and the PRESTO sensor respectively. The only difference is that the PRESTO sensor only represents a single sensor, but a source could include more than one sensor or a sensor network. The sink is the base station to which the sensor values are communicated by the source. The fundamental idea behind Ken is that both, source and sink, maintain the same dynamic probabilistic model of data evolution. The source only communicates with the sink when the raw sensor values deviate beyond a certain bound, as compared to the predictions from the dynamic probabilistic model. In the meantime, the sink uses the sensor values predicted by the model. The drawback of this method is the framework is so complicated that related domain knowledge is needed.

c) The ARIMA (AutoRegressive Integrated Moving Average) model: As the data modeling method due to its outstanding performance in modeling and forecasting and its lightweight computational cost on forecasting. Energy efficiency is achieved by suppressing the transmission of samples, whose prediction values based on the ARIMA model are within a pre-defined tolerance value from their actual values. Other benefits can be obtained from the suppression of unnecessary data transmission. For instance, if the traffic load generated by each sensor node is reduced, potential packet collisions will decrease too. The drawback is it sends the actual value only when prediction error is beyond a pre-defined error tolerance threshold.
d) The Probabilistic Adaptable Query (PAQ) system: This is one notable scheme based on time series forecasting. It uses autoregressive models maintained locally per sensor in order to keep from sending data directly to the sink. Instead, nodes communicate model parameters as necessary in order to keep the sink's predictions within some defined error bound. This method raises a drawback it offers accurate prediction of values competitively and produces considerable communication overhead.

e) Similarity-based Adaptive Framework (SAF): Adds robustness to quick changes in data trends as well as a location-independent clustering technique that allows the detection of redundant nodes. It reduces data communication between sensor node and sink yet reduces the accuracy of acquired data.

4. Objectives

To improve these methods:
- It should effectively exploit the ubiquitous temporal – spatial correlation in most natural phenomena for energy efficient data collection of WSNs.
- It should acquire sensing readings without compromising too much data accuracy loss.
- It must avoid abundant communication overhead yet keeping the data quality.
- It is to achieve a longer network lifetime of WSNs for data driven applications.
- It must conserve the finite resources, such as energy, network bandwidth and CPU usage.

5. Conclusion

This paper provides the descriptions of several data collection methods proposed for wireless sensor networks. The performance analysis of data collection methods shows the specific drawbacks of their own.

References

[1] P. Corke, T. Wark, R. Jurdak, H. Wen, P. Valencia, and D. Moore, "Environmental Wireless Sensor Networks," Proceedings of the IEEE, vol. 98, no. 11, pp. 1903-1917,2010.
[2] L. Yunhao, H. Yuan, L. Mo, W. Jiliang, L. Kebin, M. Lufeng, D. Wei, Y. Zheng, X. Min, Z. Jizhong, and L. Xiang-Yang, "Does Wireless Sensor Network Scale? A Measurement Study on GreenOrbs," in IEEE INFO COM, 2011.
[3] G. Anastasi, M. Conti, M. Di Francesco, and A. Passarella, "Energy Conservation in Wireless Sensor Networks: A Survey," Ad Hoc Net- works, vol. 7, no. 3, pp. 537-568, 2009.
[4] A. Deshpande, C. Guestrin, S. Madden, J. Hellerstein, and W. Hong, "Model-Driven Data Acquisition in Sensor Networks," in Proceedings of the 30th International Conference on Very large Data Bases (V LDB), 2004.
[5] D. Chu, A. Deshpande, J. Hellerstein, and W. Hong, "Approximate Data Collection in Sensor Networks using Probabilistic Models," in Proceedings of the 22nd International Conference on Data Engineering (ICDE), 2006.
[6] S. Santini and K. Romer, "An Adaptive Strategy for Quality-based Data Reduction in Wireless Sensor Networks," in Proceedings of the 3rd International

Conference on Networked Sensing Systems (INSS), 2006.

[7] D. Tulone and S. Madden, "PAQ: Time Series Forecasting for Approx- imate Query Answering in Sensor Networks," in European Conference on Wireless Sensor Networks (EWSN), 2006.

[8] L. Chong, W. Kui, and T. Min, "Energy Efficient Information Collection with the ARIMA Model in Wireless Sensor Networks," in IEEE Global Telecommunications Conference (GLOBECOM), 2005.

[9] L. Chong, W. Kui, and P. Jian, "An Energy-Efficient Data Collection Framework for Wireless Sensor Networks by Exploiting Spatiotemporal Correlation," Parallel and Distributed Systems, IEEE Transactions on, vol. 18, no. 7, pp. 1010-1023, 2007. ISBN 978-89-968650-1-8 65.