International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970)
Volume-2 Number-3 Issue-5 September-2012

Secure Query Processing and Routing in Wireless Sensor Networks

R. Jebakumar[1], P. Vivekanandan[2]
Research Scholar, Department of Computer Science and Engineering, Anna University, Chennai, India[1]
Professor and Head, Computer Centre, A.C Tech, Anna University, Chennai, India[2]

Abstract

Data storage has become an important issue in wireless Sensor Networks (WSN) as a huge amount of collected data needed to be archived for future reference. Storage nodes are introduced here to store the data collected from the sensors in their proximities. The storage nodes alleviate the heavy load of transmitting all data to a central place for archiving and reducing the communication cost induced by the network query. This paper proposes a new security mechanism to store the data and the query processing through the secure routing in WSN, and also addresses the storage node placement, aiming to minimize the total energy cost for gathering data to the storage nodes, replying queries and level of encryption for extend the security for data communication. This paper examines the secure routing protocol for query processing and data transmission in it.

Keywords

Wireless Sensor Network, Storage Node, Secure Routing, Query Processing.

1. Introduction

Wireless sensor networks (WSNs) comprise of nodes interacting with the physical environment and collaborate among each other to provide data to the end-users. These motes are small devices that have limited processing, communication and memory. Sensor networks are deployed for computing applications, e.g., sensing environmental conditions and monitoring people's behaviors, generates a large amount of data continuously over a long period of time. This large volume of data has to be stored somewhere for future retrieval and data analysis. One of the biggest challenges in some applications is how to store and search the collected data. The collected data can either be stored in the network sensors, or transmitted to the sink. Several problems arise when data are stored in sensors. First, a sensor is equipped with only limited memory or storage space, which prohibits the storage of a large amount of data accumulated for months or years. Second, since sensors are battery operated, the stored data will be lost after the sensors are depleted of power. Third, searching for the data of interest in a widely scattered network field is a hard problem.

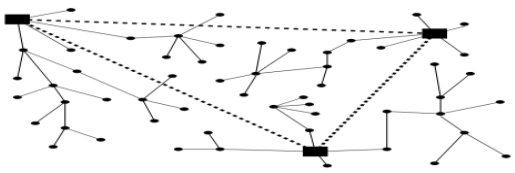
The communication generated in a network-wide search will be prohibitive. Alternatively, data can be transmitted back to the sink and stored there for future retrieval. When huge data stored in sink, delays will be present during query processing. So there is an intermediate node between sink and sensor called storage node. The storage nodes not only provide permanent storage as described previously, but also serve as a buffer between the sink and the sensor nodes. The positioning of storage nodes, however, is extremely important in this communication model. A bad placement strategy may waste the storage resources and have an adverse effect on the performance. Therefore, a good algorithm for placing storage nodes is needed to strike a balance between these two extremes characterizing for both data accumulation and data query. Security is a difficult problem in wireless networks [1]. As WSNs are a classification of wireless networks, therefore, the most of the attacks that are applicable on wireless networks tend to apply on WSNs. This paper's assertion is that sensor network routing protocols must be designed with security in mind, and this is not only effective solution for secure routing in sensor networks; we address the storage node placement and aiming to minimize the total energy cost for gathering data to the storage nodes and replying queries and extended security on storing data and query processing with secure routing.

2. Background

Sensor networks are used to report live weather conditions, monitor traffic on highways, detect disasters, monitor habitat of animals, etc. Tremendous volumes of useful data are generated by these deployments. A large amount of data cannot be transmitted from the sensor network to the sink efficiently. Furthermore, the data communication from the sensors to the sink may take long routes consuming much energy and depleting of the sensor battery power quickly. In particular, the sensors

around the sink are generally highly used and exhausted easily, thus the network may be partitioned rapidly. It is possible that, with marginal increase in cost, some special nodes with much larger permanent storage and more battery power can be deployed in sensor networks. These nodes back up the data for nearby sensors and reply to the queries. The data accumulated on each storage node can be transported periodically to a data warehouse by robots or traversing vehicles using physical mobility as data mule. Since the storage nodes only collect data from the sensors in their proximity and the data are transmitted through physical transportation instead of hop-by-hop relay of other sensor nodes, the problem of limited storage, communication capacity, and battery power is ameliorated. Placing storage nodes is related to the sensor network applications. We believe query is the most important application for sensor networks since in essence sensor networks are about providing information of the environment to the end users.

A sensor network is given with one special sensor identified as the sink (or base station) and many normal sensors, each of which generates (or collects) data from its environment. Users specify the data they need by submitting queries to the sink and they are usually interested in the latest readings generated by the sensors1. To reply to queries, one typical solution is to let the sink have all the data. Then any query can be satisfied directly by the sink. This requires each sensor to send its readings back to the sink immediately every time it generates new data. Generally, transferring all raw data could be very costly and is not always necessary. Alternatively, we allow sensors to hold their data and to be aware of the queries, then raw data can be processed to contain only the readings that users are interested in and the reduced-size reply, instead of the whole raw readings, can be transferred back to the sink. The sink diffuses queries to the storage nodes by broadcasting to the sensor network and these storage sensors reply to the queries by sending the processed data back. Compared to the previous solution, this approach reduces the raw data transfer cost, because some raw data transmissions are replaced by query reply.



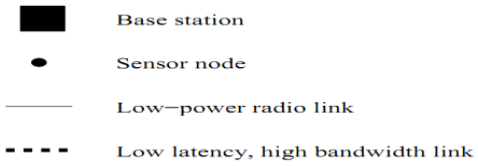


Fig. 1: A sample sensor network architecture

On the other hand, this scheme incurs an extra query diffusion cost. In this paper, we are interested in strategically designing a data access model to minimize energy cost associated with raw data transfers, query diffusion, and query replies. Therefore, our goal is to design a centralized algorithm that can derive the best locations of the storage nodes to guide the deployment of such a hybrid sensor network. In the sensor network, the base station is typically a gateway to another network, a powerful data processing or storage centre, or an access point for human interface. In some previous work on sensor network routing protocols, base stations have also been referred to as sinks. Base stations are typically many orders of magnitude more powerful than sensor nodes.

3. Related Work

There has been a lot of prior research work on data querying models in sensor networks. In early models[2],[3], query is spread to every sensor by flooding messages. Sensors return data back to the sink in the reverse direction of query messages. Those methods, however, do not consider the storage concern in sensor networks. Leach[4] is a clustering based routing protocol, in which cluster heads can use the data collected from its neighbours' to reduce communication cost to the sink. However, leach aims to reduce data transmission by aggregating data; it does not address storage problem in sensor networks.

Data-centric storage schemes, as another category of the related work, store data to different places in sensor networks according to different data types and a data centric storage scheme based on Geographic Hash Table, which inherits ideas from distributed hash table. The home site of data is obtained by applying a hash function on the data type. Thus, queries for the same type of data can be satisfied by contacting a small number of nodes. The scaling behaviour of data-centric queries for both unstructured and structured networks and derive some key scaling conditions. In general, the data-centric storage schemes ([5],[6],[7]) assume some understanding about the collected data and store them

remotely for easy data access. Extra cost is needed to forward data to the corresponding keeper nodes.
Raw data may not be easily categorized into different types in many applications. To transmit the collected data to a remote location is also considered expensive because the total collected data may be in a very large quantity. To facilitate data query[8], a multi resolution data storage system, Dimensions, where data are stored in a degrading loss model, i.e., fresh data are stored completely while long-term data are stored loosely. In comparison, the proposed scheme is more general without any assumption about the data correlation. A proxy tier is introduced between sensor nodes and user terminals and proxy nodes can cache previous query responses. Compared to the storage nodes, proxy nodes in PRESTO[9] have no resource constraints in term of power, computation, storage and communication. It is a more general storage architecture that does not take the characteristics of data generation or query into consideration.

4. Secure Query Processing

A sensor network is given with one special sensor identified as the sink. Users specify the data they need by submitting queries to the sink and they are usually interested in the latest readings generated by the sensors. To reply to queries, one typical solution is to let the sink have the data. Then any query can be satisfied directly by the sink. Otherwise if the data are not in the sink then query is forwarded to the storage node and process it .This requires each sensor to send its readings back to the storage node immediately every time it generates new data.

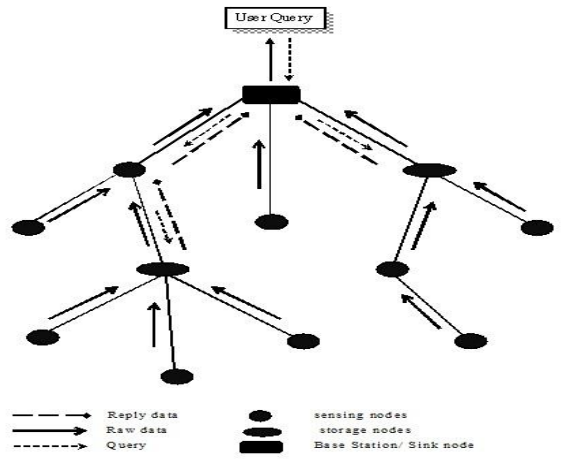


Fig. 2: Query Processing in sensor network.
Storage nodes

These types of nodes have much larger storage capacity than regular sensors. In the data access model, they store all the data received from other nodes or generated by them. They do not send out anything until queries arrive. According to the query description, they obtain the results needed from the raw data they are holding and then return the results back to the sink. Note that except enriched storage capacity, other resources on storage nodes are still constrained as regular sensors. The sink itself is considered as a storage node.

Forwarding nodes:
These types of nodes are regular sensors and they always forward the data received from other nodes or generated by them along a path towards the sink. The outgoing data are kept intact and the forwarding operation continues until the data reach a storage node. The forwarding operation is independent of queries and there is no data processing at forwarding nodes.

4.1. Routing Protocol

Classic routing protocols typically forward data along the shortest path to the destination. If, however, we are interested in processing query to minimize energy expenditure, nodes should route packets based on the packet content and choose the next hop in order to promote in-network aggregation when it required. This type of data forwarding is often referred to data centric routing. According to the data centric paradigm, as a node searches for the relay nodes, it needs to use metrics which take into account the positions of the most suitable processing points for query processing based on availability of data. Altogether, the application scenario, routing scheme, query processing and data aggregation mechanism are closely interrelated.

Moreover, in-network aggregation based on query processing techniques may require some form of synchronization among nodes in the network and the storage nodes may lead to better data accessing opportunities and, in turn, improved performance. Some strategies are required especially in the case of monitoring applications where sensor nodes need to periodically report their readings to the storage node and maintain frequently used data in sink node.

4.2. Secure Data Aggregation

As wireless sensor networks continue to grow in size, so does the amount of data that the sensor networks are capable of sensing. However, due to the computational constraints placed on individual

sensors, a single sensor is typically responsible for only a small part of the overall data. Because of this, a query of the wireless sensor network is likely to return a great deal of raw data, much of which is not of interest to the individual performing the query.

The problem with the standard information aggregation techniques, however, is that they assume that all nodes are trustworthy. Of course, this is not the case and secure data aggregation techniques will be necessary in many wireless sensor networks. The storage node is responsible for committing to the collected data. This commitment ensures that the storage node actually uses the data collected from the sensors and processing the query when it will required.

5. Discussion

In this paper, the performance of each and every node is evaluated. The exact solutions on how to place storage nodes to minimize the total energy cost. Query redirection and complex query processing will lead to more performance in this network done by the sink node, in the routing process that identify the malicious node and maintain a block list for it and it will avoid the future attempt during the data or query transmission.

6. Conclusion and Future Work

This paper considers the storage node placement problem in a sensor network. Introducing storage nodes into the sensor network alleviates the communication burden of sending all the raw data to a central place for data archiving and facilitates the data collection by transporting data from a limited number of storage nodes and examines how to place storage nodes to save energy for data collection, data aggregation and blocking the malicious nodes for the communication to interrelate with our network. We will extend the scope of routing security for secure data communication and query transmission.

References

[1] H. Chan and A. Perrig."Security and privacy in sensor networks," IEEE Computer Magazine, pages 103–105, 2003 2003.

[2] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, "TAG: a tiny aggregation service for ad-hoc sensor networks," SIGOPS Operating System Review, vol. 36, no.SI, pp. 131–146, 2002.

[3] "The design of an inquisitional query processor for sensor networks," in Proceedings of the22nd ACM SIGMOD International Conference on Management of Data, NY, USA, 2003, pp. 491–502.

[4] W. Heinzelman and H. Balakrishnan, "Energy efficient communication protocols for wireless Micro sensor networks,"in Proceedings of International Conference on System Sciences, Jan 2000.

[5] Shenker, S. Ratnasamy, B. Karp, R. Govindan, and D. Estrin,"Datacentric storage in sensornets, "SIGCOMM Computer Communication review, vol. 33, no. 1, pp. 137–142, 2003.

[6] S. Ratnasamy, B. Karp, S. Shenker, D. Estrin, "Data-centric storage in sensornets with GHT, a geographic hashtable," Mobile Networks and Applications, vol. 8, no. 4, pp. 427–442, 2003.

[7] J. Newsome and D. Song, "GEM: graph embedding for routing anddata-centric storage in sensor networks without geographic information".

[8] D. Ganesan, B. Greenstein, D. Estrin, J. Heidemann, and R.Govindan, "Multiresolution storage and search in sensor networks," ACM Trans.Storage, vol. 1, no. 3, pp. 277–315, 2005.

[9] M. Li, D. Ganesan, and P. Shenoy, "PRESTO: Feedback-driven Data management in sensor networks," in Proceedings of the 3rd USENIX Symposium on Networked Systems Design San Jose, CA, USA, May 2006.