International Journal of Research in Advent Technology (IJRAT) (E-ISSN: 2321-9637)
Special Issue
National Conference "CONVERGENCE 2016", 06th-07th April 2016

Data Aggregation Techniques in Wireless Sensor Networks: A Review

Ms.Laxmi S.Waghmare[1]
First year ME Student,Computer Science and Engineering Department,Pankaj Laddhad Institute of Technology and Management Studies,Buldana

Dr.P.M.Jawandhiya[2]
Principal,Pankaj Laddhad Institute of Technology and Management Studies,Buldana
Email: Waghmaresangita9@gmail.com, principal@plit.ac.in

Abstract: Wireless sensor networks consist of sensor nodes with sensing and communication capabilities.We focus on data aggregation problems in energy constrained sensor networks. The main goal of data aggregation algorithms is to gather and aggregate data in an energy efficient manner so that network lifetime is enhanced. In this paper, we present a survey of data aggregation algorithms in wireless sensor networks. We compare and contrast different algorithms on the basis of performance measures such as lifetime, latency and data accuracy. We conclude with possible future research directions.

Keywords: Routing Protocols**,**performance measures such as lifetime, latency and data accuracy

1. INTRODUCTION

Wireless sensor networks (WSNS): used for numerous applications including military surveillance, facility monitoring and environmental monitoring. Typically WSNs have a large number of sensor nodes with the ability to communicate among themselves and also to an external sink or a base-station [1,2].The sensors could be scattered randomly in harsh environments such as a battlefield or deterministically placed at specified locations.

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 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. A comprehensive survey on wireless sensor networks is presented in [3]. Data gathering is defined as the systematic collection of sensed data from multiple sensors to be eventually transmitted to the base station for processing. Since sensor nodes are energy constrained, it is inefficient for all the sensors to transmit the data directly to the base station. Data generated from neighboring sensors is often redundant and highly correlated. In addition, the amount of data generated in large sensor networks is usually enormous for the base station to process.

Hence, we need methods for combining data into high quality information at the sensors or intermediate nodes which can reduce the number of packets transmitted to the base station resulting in conservation of energy and bandwidth. This can be accomplished by data aggregation. *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). In the rest of the paper, we use the term data aggregation to denote the process of data gathering with aggregation. We also use the term sink to represent the base station.

Data aggregation attempts to collect the most critical data from the sensors and make it available to the sink in an energy efficient manner with minimum data latency.Data latency is important in many applications such as environment monitoring where the freshness of data is also an important factor. It is critical to develop energy efficient data aggregation algorithms so that network lifetime is enhanced. There are several factors which determine the energy efficiency of a sensor network such as network architecture,he data aggregation mechanism and the underlying routing protocol.

2. ROUTING PROTOCOLS FOR MOBILE WSN

Recent advances in micro-electro-mechanical systems and low power and highly integrated digital electronics have led to the development of micro-sensors [1,5]. Such sensors are generally equipped with data processing and communication capabilities. The sensing circuitry measures ambient conditions related to the environment surrounding the sensor and transforms them into an electric signal. Processing such a signal reveals some properties about objects located and/or events happening in the vicinity of the sensor. The sensor sends such collected data, usually via radio transmitter, to a command center (sink) either directly or through a data concentration center (a gateway). The decrease in the size and cost of sensors, resulting from such technological advances, has fueled interest in the possible use of large set of disposable unattended sensors.

2.1Network dynamics

There are three main components in a sensor network. These are the sensor nodes, sink and monitored events. Aside from the very few setups that utilize mobile sensors [1], most of the network architectures assume that sensor nodes are stationary. On the other hand, supporting the mobility of sinks or cluster-heads (gateways) is sometimes deemed necessary [2]. Routing messages from or to moving nodes is more challenging since route stability becomes an important optimization factor, in addition to energy, bandwidth etc. The sensed event can be either dynamic or static depending on the application [3]. For instance, in a target detection/tracking application, the event (phenomenon) is dynamic where as forest monitoring for early fire prevention is an example of static events. Monitoring static events allows the network to work in a reactive mode, simply generating traffic when reporting. Dynamic events in most applications require periodic reporting and consequently generate significant traffic to be routed to the sink.

2.2 Node deployment

Another consideration is the topological deployment of nodes. This is application dependent and affects the performance of the routing protocol. The deployment is either deterministic or self-organizing. In deterministic situations, the sensors are manually placed and data is routed through pre-determined paths. However in self organizing systems, the sensor nodes are scattered randomly creating an infrastructure in an ad hoc manner [2]. In that infrastructure, the position of the sink or the cluster-head is also crucial in terms of energy efficiency and performance. When the distribution of nodes is not uniform, optimal clustering becomes a pressing issue to enable energy efficient network operation.

2.3 Energy considerations

During the creation of an infrastructure, the process of setting up the routes is greatly influenced by energy considerations. Since the transmission power of a wireless radio is proportional to distance squared or even higher order in the presence of obstacles, multi-hop routing will consume less energy than direct communication. However, multi-hop routing introduces significant overhead for topology management and medium access control. Direct routing would perform well enough if all the nodes were very close to the sink [4]. Most of the time sensors are scattered randomly over an area of interest and multi-hop routing becomes unavoidable.

2.4 Data delivery models

Depending on the application of the sensor network, the data delivery model to the sink can be continuous, event-driven, query-driven and hybrid [13]. In the continuous delivery model, each sensor sends data periodically. In event-driven and query driven models, the transmission of data is triggered when an event occurs or a query is generated by the sink. Some networks apply a hybrid model using a combination of continuous, event-driven and query-driven data delivery. The routing protocol is highly influenced by the data delivery model, especially with regard to the minimization of energy consumption and route stability. For instance, it has been concluded in [7] that for a habitat monitoring application where data is continuously transmitted to the sink, a hierarchical routing protocol is the most efficient alternative. This is due to the fact that such an application generates significant redundant data that can be aggregated on route to the sink, thus reducing traffic and saving energy.

2.5 Node capabilities

In a sensor network, different functionalities can be associated with the sensor nodes. In earlier K. Akkaya, M. Younis / Ad Hoc Networks 3 (2005) 325–349 327 works [5], all sensor nodes are assumed to be homogenous, having equal capacity in terms of computation, communication and power. However, depending on the application a node can be dedicated to a particular special function such as relaying, sensing and aggregation since engaging the three functionalities at the same time on a node might quickly drain the energy of that node. Some of the hierarchical protocols proposed in the literature designate a cluster-head different from the normal sensors. While some networks have picked cluster-heads from the deployed sensors [4], in other applications a cluster-head is more powerful than the sensor nodes in terms of energy, bandwidth and memory [5]. In such cases, the burden of transmission to the sink and aggregation is handled by the cluster-head. Inclusion of heterogeneous set of

sensors raises multiple technical issues related to data routing[2]. For instance, some applications might require a diverse mixture of sensors for monitoring temperature, pressure and humidity of the surrounding environment, detecting motion via acoustic signatures and capturing the image or video tracking of moving objects. These special sensors either deployed independently or the functionality can be included on the normal sensors to be used on demand. Reading generated from these sensors can be at different rates, subject to diverse quality of service constraints and following multiple data delivery models, as explained earlier. Therefore, such a heterogeneous environment makes data routing more challenging.

2.6 Data aggregation/fusion

Since sensor nodes might generate significant redundant data, similar packets from multiple nodes can be aggregated so that the number of transmissions would be reduced. Data aggregation is the combination of data from different sources by using functions such as suppression (eliminating duplicates), min, max and average [4]. Some of these functions can be performed either partially or fully in each sensor node, by allowing sensor nodes to conduct in-network data reduction[8]. Recognizing that computation would be less energy consuming than communication [4], substantial energy savings can be obtained through data aggregation. This technique has been used to achieve energy efficiency and traffic optimization in a number of routing protocols [8]. In some network architectures, all aggregation functions are assigned to more powerful and specialized nodes [6]. Data aggregation is also feasible through signal processing techniques. In that case, it is referred as data fusion where a node is capable of producing a more accurate signal by reducing the noise and using some techniques such as beam forming to combine the signals [4].

3. CLASSIFICATION OF ROUTING PROTOCOLS BASED ON THE STATE OF THE INFORMATION

Because of multiple and diverse ad hoc protocols there is an obvious need for a general taxonomy to classify protocols considered. Traditional classification is to divide protocols to table-driven and to source-initiated on-demand driven protocols [1]. Table-driven routing protocols try to maintain consistent, up-to-date routing information from each node to every other node. Network nodes maintain one or many tables for routing information. Nodes respond to network topology changes by propagating route updates throughout the network to maintain a consistent network view. Source-initiated on-demand protocols create routes only when these routes are needed. The need is initiated by the source, as the name suggests. When a node requires a route to a destination, it initiates a route discovery process within the network. This process is completed once a route is found or all possible route permutations have been examined. After that there is a route maintenance procedure to keep up the valid routes and to remove the invalid routes. This classification has though some drawbacks because of its rough granularity. To that classification it is possible to make some modifications (e.g. in [2]). These modifications can make some assumption about if the routing is flat or hierarchical and if any means to obtain global positioning information is in use. One very attractive taxonomy has been introduced by Feeney [3].

This taxonomy is based on to divide protocols according to following criteria, reflecting fundamental design and implementation choices:

- **Communication model.** What is the wireless communication model? Multi- or single channel?
- **Structure.** Are all nodes treated uniformly? How are distinguished nodes selected? Is the addressing hierarchical or flat?
- **State Information.** Is network-scale topology information obtained at each node?
- **Scheduling.** Is route information continually maintained for each destination? This model does not take an account for if a protocol is unicast, multicast, geo-cast or broadcast. Also the taxonomy doesn't deal with the question how the link or node related costs are measured. These properties are however worth to be considered in classification and evaluating applicability of protocols. Based on that lack the taxonomy has been slightly modified by adding such features as **type of cast** and **cost function**. Type of cast feature is an upper level classification and so the protocols to be classified must firstly divide by type of cast and after that the more accurate taxonomy can be applied. The above mentioned taxonomy is applied to unicast protocols, while in the context of multicast and geo-cast protocols a specified taxonomy has been introduced. The overall taxonomy and specially the unicast protocol classification can be seen in figure 1. The cost function is a classification to be concatenated after presented taxonomy. It is like a remark to be noticed when considering the applicability of the protocol to be chosen.

3.1 Communication Model

Protocols can be divided according to communications model to protocols that are designed for **multi-channel** or **single-channel** communications. Multi-channel protocols are routing protocols generally used in TDMA or CDMA-based networks. They combine channel assignment and routing functionality. That kind of protocol is e.g. Cluster head Gateway Switched Routing (CGSR) [4]. Single -channel protocols presume one shared media to be used. They are generally CSMA/CA-oriented, but they have a

wide diversity in which extend they rely on specific link-layer behaviors.

3.2 Structure

Structure of a network can be classified according to node uniformity. Some protocols treat all the nodes uniformly, other make distinctions between different nodes. In **uniform protocols** there is no hierarchy in network, all nodes send and respond to routing control messages at the same manner. In **non-uniform protocols** there is an effort to reduce the control traffic burden by separating nodes in dealing with routing information. Non-uniform protocols fall into two categories: protocols in which each node focuses routing activity on a subset of its neighbors and protocols in which the network is topologically partitioned. These two different methods for non uniformity are called **neighbor selection** and **partitioning** respectively. With neighbor selection mechanism, every node has its own criteria to classify network nodes to near or to remote nodes. In partitioning protocols that differentiation is to use hierarchical node separation. Hierarchical protocols have some upper-level and lower level nodes and certain information difference between them.

3.3 State Information

Protocols may be described in terms of the state information obtained at each node and / or exchanged among nodes.

Topology-based protocols use the principle that every node in a network maintains large scale topology information. This principle is just the same as link-state protocols use.

Destination-based protocols do not maintain large-scale topology information. They only may maintain topology information needed to know the nearest neighbors. The best known such protocols are distance-vector protocols, which maintain a distance and a vector to a destination (hop count or other metric and next hop).

3.4 Scheduling

The way to obtain route information can be a continuous or a regular procedure or it can be trigged only by on demand. On that basis the protocols can be classified to proactive and on-demand protocols. **Proactive protocols**, which are also know as table-driven protocols, maintain all the time routing information for all known destinations at every source. In these protocols nodes exchange route information periodically and / or in response to topology change. In on-demand i.e. in **reactive protocols** the route is only calculated on demand basis. That means that there is no unnecessary routing information maintained. The route calculation process is divided to a route discovery and a route intenance phase. The route discovery process is initiated when a source needs a route to a destination. The route maintenance process deletes failed routes and re-initiates route discovery in the case of topology change.

3.5 Type of Cast

Protocols can be assumed to operate at unicast, multicast, geocast or broadcast situations. In **unicast protocols** one source transmits messages or data packets to one destination. That is the most normal operation in any network. The unicast protocols are also the most common in ad hoc environment to be developed and they are the basis on which it is a possibility to construct other type of protocols. Unicast protocols have thought some lacks when there is a need to send same message or stream of data to multiple destinations. So there is an evitable need for multicast protocols.

Multicast routing protocols try to construct a desirable routing tree or a mesh from one source to several destinations. These protocols have also to keep up with information of joins and leave ups to a multicast group. The purpose of **geocast protocols** are to deliver data packets for a group of nodes which are situated on at specified geographical area. That kind of protocol can also help to alleviate the routing procedure by providing location information for route acquisition. Broadcast is a basic mode of operation in wireless medium. Broadcast utility is implemented in protocols as a supported feature. Protocol only to implement broadcast function is not a sensible solution. That is the reason not to classify protocols to broadcast protocols. But it is worth to mention if a protocol is not supporting that method.

3.6 Cost Function

When making routing decisions in ad hoc environments, it is normally not enough to take only considerations to hop count. In ad hoc networks there is a wide variety of issues to consider such as link capacity, which can vary in large scale, latency, link utilization percentage and terminal energy issues to mention a few most relevant. That is why there is a need to adapt cost functions to route calculations. Rough classification of protocols according to cost function can be based on **hop count** approach (no special cost function applied) and to **bandwidth** or **energy** based cost functions. Also quite a different approach to routing metrics is used by Associativity Based Routing (ABR) protocol, which uses **degree of association stability** for a metric to decide for a route That means that presumably more permanent routes are preferred. [5].

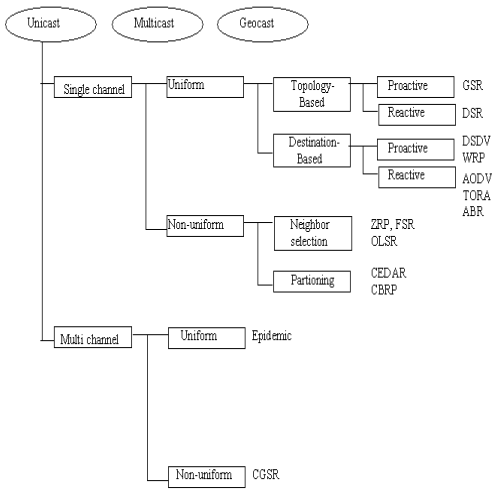

Figure 1: Taxonomy of Protocols.

4. DATA AGGREGATION PROTOCOLS BASED ON NETWORK ARCHITECTURE

The architecture of the sensor network plays a vital role in the performance of different data aggregation protocols. In this section, we survey several data aggregation protocols which have specifically been designed for different network architectures.

4.1 Flat networks

In flat networks, each sensor node plays the same role and is equipped with approximately the same battery power. In such networks, data aggregation is accomplished by data centric routing where the sink usually transmits a query message to the sensors, e.g, via flooding and sensors which have data matching the query send response messages back to the sink. The choice of a particular communication protocol depends on the specific application at hand. In the rest of this subsection, we describe these protocols and highlight their advantages and limitations.

5. SECURITY ISSUES IN DATA AGGREGATION

Security in data transmission and aggregation is an important issue to be considered while designing sensor networks. In many applications, sensors are deployed in open environments and are susceptible to physical attacks which might compromise the sensor's cryptographic keys. Secure aggregation of information is a challenging task if the data aggregators and sensors are malicious. In this subsection, we describe some recent work which solve the secure data aggregation problem and also discuss some of the main issues involved in implementing security in sensor networks.

It is analyzed the two main practical issues involved in implementing data encryption at the sensors viz., the size of the encrypted message and the execution time for encryption at the sensors. Privacy homo morphisms (PH) are encryption functions which allow a set of operations to be performed on encrypted data without the knowledge of decryption functions. In [8], PH has been used to analyze the feasibility of security implementation in sensors. PH uses a positive integer for computing the secret key. The size of the encrypted data increases by a factor of *d* compared to the original data. Hence in the light of minimizing packet overhead, *d* should be chosen in the range of 2-4 as suggested in [8]. Execution times for encryption operation at the sensors increase with *d*. For instance when *d*=2, the execution time for encryption of one byte of data is 3481 clock cycles on a MICA2 mote which increases to 4277 clock cycles when *d*=4 as reported in [2]. MICA2 motes cannot handle the computation for *d* Hence, the tradeoff between security and computation complexity should be considered when implementing data encryption schemes on sensors. The other main aspect of security in sensor networks is the establishment of secret keys between the sensor and the base station. [6] have proposed security protocols for sensor networks which address the key establishment problem. In the approach proposed in [5], all nodes trust the base station at the network creation time and each node is given a master key which is shared with the base station. To achieve authentication between a sensor and base station, a message authentication code (MAC) is used. The keys for encrypting the data and computing the MAC are derived from the master key using a pseudo random function. All keys derived using this procedure are computationally independent. Hence, if an attacker hacks the key, it would not help in determining the master key or any other key. In scenarios where a key is compromised, a new key can be derived without transmitting confidential information. Przy datek et al. [3] have proposed a framework for secure data aggregation in large sensor networks. They have presented secure protocols for the computation of median, maximum, minimum and average of sensor measurements and estimation of network size. The following issues have been addressed for secure data aggregation. a) Some sensor nodes may be compromised and transmit wrong data values to the aggregator that corrupts the aggregation result. b) The aggregator may be compromised and report malicious aggregate values to the home server or sink. c) Estimation errors introduced by the

sampling techniques used by the aggregator to compute the result.

6. CONCLUSIONS

We have presented a comprehensive survey of data aggregation algorithms in wireless sensor networks. All of them focus on optimizing important performance measures such as network lifetime, data latency, data accuracy and energy consumption. Efficient organization, routing and data aggregation tree construction are the three main focus areas of data aggregation algorithms. We have described the main features, the advantages and disadvantages of each data aggregation algorithm. We have also discussed special features of data aggregation such as security and source coding. The trade-offs between energy efficiency, data accuracy and latency have been highlighted. Security is another important issue in data aggregation applications and has been largely unexplored. Integrating security as an essential component of data aggregation protocols is an interesting problem for future research. Data aggregation in dynamic environments presents several challenges and is worth exploring in the future.

REFERENCES

[1] H. O. Tan and I. Korpeoglu, " Power efficient data gathering and aggregation in wireless sensor networks," *SIGMOD Record*, vol. 32, no. 4, December 2003, pp 66-71.

[2] K. Vaidhyanathan, S. Sur, S. Narravula, P. Sinha, "Data aggregation techniques sensor networks," Technical Report, OSU-CISRC-11/04-TR60, Ohio State University, 2004.

[3] K. Kalpakis, K. Dasgupta and P. Namjoshi, "Efficient algorithms for maximum lifetime data gathering and aggregation in wireless sensor networks," *Computer Networks*, vol. 42, no. 6, August 2003, pp.697-716.

[4] Y. Xue, Y. Cui and K. Nahrstedt, "Maximizing lifetime for data aggregation in wireless sensor networks," *ACM/Kluwer Mobile Networks and Applications (MONET) Special Issue on Energy Constraints and Lifetime Performance in Wireless Sensor Networks*, Dec. 2005, pp. 853- 64.

[5] B. Hong, V.K. Prasanna, "Optimizing system lifetime for data gathering in networked sensor systems," *Workshop on Algorithms for Wireless and Ad-hoc Networks* (*A-SWAN*), August 2004, Boston.

[6] R. Cristescu, B. Beferull-Lozano and M.Vetterli, "On network correlated data gathering," *IEEE INFOCOM*, vol. 4, no. 4, March 2004, pp 2571- 82.

[7] N. Sadagopan, and B. Krishnamachari, "Maximizing data extraction in energy-limited sensor networks," *INFOCOM 2004*, vol.3, March 2004, pp. 1717-1727.

[8] F. Ordonez, and B. Krishnamachari, "Optimal information extraction in energy-limited wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, August 2004, pp. 1121-1129.

AUTHOR'S BIOGRAPHY:

Ms. Laxmi S. Waghmare

Pursuing M.E. (CSE Engg) 1st Year from PLITMS, Buldana. SGBAU Amravati, (MS), India.

Dr. P.M.Jawandhiya

Principal, PLITMS, Buldana. SGBAU Amravati, (MS), India.