A Stochastic Analysis on Message, Time and Energy Complexity on Wireless Sensor Networks

A.Praveena, D.Shyam Prakash, S.V.Mani Sekaran

Abstract— Recently considerable contributions have been made in the area of wireless sensor networks (WSN) because they can monitor the environment in a more efficient and convenient way. These are increasingly deployed for applications such as wildlife habitat monitoring, forest fire prevention, battlefield, emergency relief, environment monitoring and military surveillance. Due to its wide-range potential applications sensor networks has recently emerged as a premier research topic. In many sensor network applications, often the ultimate goal is to collect sensing data from all sensors to certain sink nodes and the sensed data is aggregated and transmitted to the sinks for analysis. Typically, an aggregate (or summarized) value is computed at the data sink by applying the corresponding aggregate function, e.g., MAX, COUNT, AVERAGE or MEDIAN to the collected data. It is envisioned that the sink node issues queries regarding the data collected by some target sensors, and the sensors collaboratively generate an accurate or approximate response. Researchers have designed several energy-efficient algorithms for computing aggregates using the tree-based approach. Thus, in-network data aggregation [1] becomes an important technique in WSN and has been well studied in recent years. In this paper, we study three different data processing, operations, namely data collection, data aggregation, and data selection. For each problem, we will study its complexity and present efficient algorithms to solve it. The main contributions of this paper are as follows: We design algorithms for data collection, data aggregation and data selection whose time complexity and message complexity are within constant factors of the optimum. Thus, the method achieves the best trade-offs among the time complexity, message complexity, and energy complexity. The analysis and simulation results show that the proposed algorithm out performs other aggregation scheduling algorithms.

Index Terms— Complexity analysis, Data aggregation, Energy Complexity, Wireless Sensor Networks.

1 INTRODUCTION

DEVELOPMENTS in micro electro mechanical systems (MEMS) and wireless networks are opening a new domain in networking history. Recent technological advances in wireless networking, IC fabrication and sensor technology have led to the emergence of millimeter scale devices that collectively form a WSN and are radically changing the way in which we sense, process and transport signals of interest.

WSN's are given by a large number of small, low cost sensor nodes that are densely deployed either inside or close to a phenomenon of interest with computational capabilities connected through wireless links and collect and disseminate environmental data. Each sensor node is an independent, low-power, smart device with sensing, processing and wireless communication capabilities. WSN's are delivering near-real-time information to scientists worldwide. Extracting this information to gain knowledge and understanding is one of the greatest challenges faced today. These networks are quickly gaining popularity due to the fact that they are potentially low cost solutions to a variety of real world challenges and are expected to play an essential role in the upcoming age of pervasive computing.

- A.Praveena is currently working as Assistant Professor in Dept of IT, Dr.N.G.P Institute of Technology, Coimbatore, Tamil Nadu, India PH-9965545508. E-mail:praveensngp@gmail.com
- D.Shyam Prokash is currently working as Assistant Professor in Dept of IT,Dr.N.G.P Institute of Technology, Coimbatore, Tamil Nadu, India PH-9043698876. E-mail:shyam15287@gmail.com
- S.V.Mani Sekaran is currently working as Assitant Professor in Dept of IT, Anna University of Technology, Regional Center, Coimbatore, Tamil Nadu, India.

These networks are an important ingredient of "anywhere and anytime" ubiquitous wireless next generation communication infrastructure. In this diversified yet integrated future network environments, WSN has a role of reliable monitoring and control of variety of applications based on environmental sensing with better accuracy. They have applications in a variety of fields such as environment monitoring (air, soil and water, condition based maintenance), habitat monitoring (plant, animal species population and behavior), seismic detection, military surveillance, inventory tracking, smart spaces, gathering sensing information in inhospitable locations, medical and home security to machine diagnosis, chemical/biological detection etc. In spite of diverse applications, WSN pose a number of unique technical challenges due to Adhoc deployment, unattended operation, unethered, dynamic changes. All of these sensor applications depend on the ability to extract data from the network. Often, this data consists of summaries (or aggregations) rather than raw sensor readings. Other researchers have noted the importance of data aggregation in sensor networks [13, 10, 12]. This previous work has tended to view aggregation as an application-specific mechanism that would be programmed into the devices on an as-nee ded basis, typically in error-prone, low-level languages like C.

With advance in technology, sensor networks composed of small and cost effective sensing devices equipped with wireless radio transceiver for environment monitoring have become feasible. The key advantage of using these small devices to monitor the environment is that it does not require infrastructure such as electric mains for power supply and wired lines for Internet connections to collect data, nor need human interaction while deploying. These sensor nodes can monitor

ISSN 2229-5518

the environment by collecting information from their surroundings, and work cooperatively to send the data to a base station, or sink, for analysis. The main goal of data aggregation algorithms is to gather and aggregate data in an energy efficient manner so that network lifetime is enhanced. WSN offer an increasingly attractive method of data gathering in distributed system architectures and dynamic access via wireless connectivity. They have limited computational power and limited memory and battery power, this leads to increased complexity for application developers and often results in applications that are closely coupled with network protocols.

1.1 Contributions of the Paper

This paper is intended to be an introduction to WSN with an emphasis on structural and environmental monitoring applications. In this paper we concentrate on the complexity of distributed data collection, data aggregation, and data selection in WSNs. The rest of the paper is organized as follows: In Section 2, we first present the necessary background information. The overview of data aggregation is discussed in Section 3. Section 4 discusses the related work. Section 5 discusses the limitations with previous work. Section 6 discusses the proposed scheme and paper concludes with future work in Section 6.

2 Sensor Network Architecture

A typical architecture of WSN is shown in the figure 1. The sensor nodes are usually scattered in a sensor field and has the capabilities to collect data and perform partial or no processing on the data. Each sensor node has the required infrastructure to communicate with other nodes. Data are routed back to sink/base station by a multihop infrastructure less architecture through the sink. A special type of node called a gateway nodes are connected to components outside of the network through long range communication cables or satellite links), and all communication with users of the sensor network goes through the gateway node.

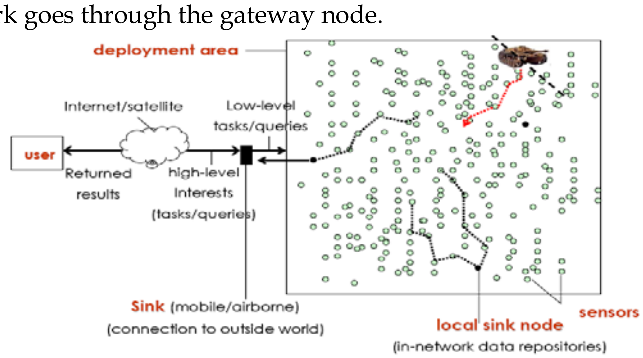

Fig. 1. Typical Sensor Network

The sink node communicates with the task manager via core network which can be Internet or Satellite. Since sensors are low cost, low power, and small in size, the transmission power of a sensor is limited. The data transmitted by a node in the field may pass through multiple hops before reaching the sink. Many route discovery protocols (mostly inherited from Ad hoc networks) have been suggested for maintaining routes from field sensors to the sink(s). Due to low memory, scarcity of available bandwidth and low power of the sensors, many researchers considered these separate route discovery mechanisms undesirable. Once sensors are deployed they remain unattended, hence all operations e.g. topology management, data management etc. should be automatic and should not require external assistance. In order to increase the network life time, the communication protocols need to be optimized for energy consumption. It means a node must be presented lowest possible data traffic to process.

A sensor node is made up of four (fig 2) basic components: sensing unit, processing unit, transceiver unit and power unit. The additional application-dependent components are location finding system, power generator and mobilizer. Sensing units are usually composed of two subunits: sensors and analog to digital converter. The analog signals produced by the sensors based on the observed phenomenon are converted to digital signals by the ADC, and then fed to the processing unit. It is generally associated with a small range a small storage unit, manages the procedures that make the sensor node collaborate with the other nodes to carry out the assigned sensing tasks. A transceiver unit connects the node to the network. The most important component is the power unit and is supported by power scavenging units such as solar cells. There are also other subunits that are application dependent.

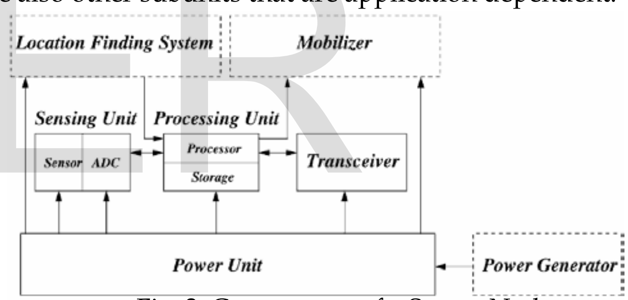

Fig. 2. Components of a Sensor Node

The emergence of WSN as one of the dominant technology trends in the coming decades has posed numerous unique challenges to researchers. Sensor nodes are functioning autonomously without access to renewable energy resources. Cost constraints and the need for ubiquitous, invisible deployments will result in small sized sensor nodes. Such nodes have resource constraints such as communication, power consumption, computation and uncertainty in sensor readings. While the set of challenges in WSN are diverse, we focus on fundamental security challenges in this paper.

3 Overview of Data Aggregation in WSN

Our main motivating applications in this survey arise in the field of sensor networks [9]. The authors in [10{12] report the deployment of such networks in a wide range of scientific, security, industrial and business applications. Examples include climatologically and environmental monitoring, traffic monitoring, smart homes, re detection, seismic measurements, structural integrity, animal control and habitat monitoring. Apart from sensor networks, other motivating applications include IP routing and network traffic monitoring. Generally nodes in WSNs periodically collect data from physical world and transmit them to monitor nodes. However these data usually have some redundancy or semantically similarity, if we can combine duplicate data and reduce redundancy, we will significantly reduce the amounts of data transmitted in the networks so as to save energy. This is the original idea of Data Aggregation [2] which now also becomes indispensable in wireless networks.

3.1 Aggregation

In the scenarios outlined above, single individual values are usually not of great relevance. In fact, users are more interested in the quick extraction of succinct and useful synopses about a large portion of the underlying observation set trying to collect all data monitored by the sensors would be unrealistic in terms of bandwidth, power consumption and communication intensity. So, the canonical approach is to compute statistical aggregates, such as max, min, average, quantiles, heavy hitters, etc., that can compactly summarise the distribution of the underlying data.

3.2 Overview

Data aggregation is a process of aggregating the sensor data using aggregation approaches. The general data aggregation algorithm works as shown in the below figure. The algorithm uses the sensor data from the sensor node and then aggregates the data by using some aggregation algorithms such as centralized approach, LEACH(low energy adaptive clustering hierarchy),TAG(Tiny Aggregation) etc. This aggregated data is transfer to the sink node by selecting the efficient path.

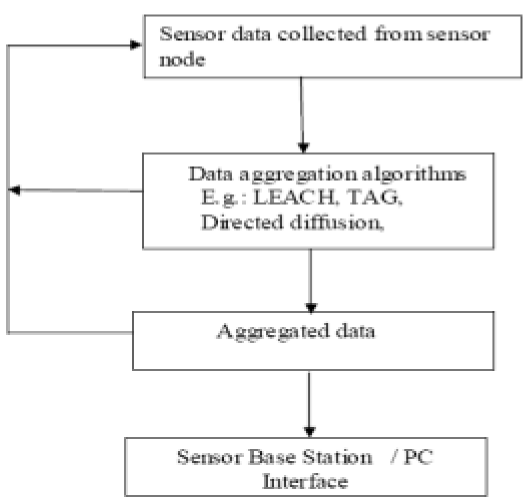

Fig. 3. General Architecture of Data Aggregation

There are many types of aggregation techniques available and some of them are given below.

Centralized Approach: This is an address centric approach where each node sends data to a central node via the shortest possible route using a multihop wireless protocol. The sensor nodes simply send the data packets to a leader, which is the powerful node. The leader aggregates the data which can be queried. Each intermediate node has to send the data packets addressed to leader from the child nodes. So a large number of messages have to be transmitted for a query in the best case equal to the sum of external path lengths for each node.

In-Network Aggregation[7]: In-network aggregation is the global process of gathering and routing information through a multi-hop network, processing data at intermediate nodes with the objective of reducing resource consumption (in particular energy), thereby increasing network lifetime. There are two approaches for in-network aggregation: with size reduction and without size reduction. In-network aggregation with size reduction refers to the process of combining & compressing the data packets received by a node from its neighbors in order to reduce the packet length to be transmitted or forwarded towards sink. In-network aggregation without size reduction refers to the process merging data packets received from different neighbors in to a single data packet but without processing the value of data.

Tree-Based Approach [8]: In the tree-based approach perform aggregation by constructing an aggregation tree, which could be a minimum spanning tree, rooted at sink and source nodes are considered as leaves. Each node has a parent node to forward its data. Flow of data starts from leaves nodes up to the sink and therein the aggregation done by parent nodes.

Cluster-Based Approach [6]: In cluster-based approach, whole network is divided in to several clusters. Each cluster has a cluster-head which is selected among cluster members. Cluster-heads do the role of aggregator which aggregate data received from cluster members locally and then transmit the result to sink.

4 Other Related Work on Data Aggregation

Data aggregation in sensor networks has been well studied in recent years [4]–[7]. In-network aggregation means computing and transmitting partially aggregated data rather than transmitting raw data in networks to reduce the energy consumption [1]. There are vast amounts of extant work on in-network aggregation in the literature [8], [9]. Suppression scheme and model-driven approach were proposed in [10], [11] towards reducing communication cost. The tradeoff between energy consumption and time latency was considered in [12]. A heuristic algorithm for both broadcast and data aggregation was designed in [13]. Another heuristic algorithm for data aggregation was proposed in [14] aiming to reduce time latency and energy consumption. [15] proposed a randomized and distributed algorithm for aggregation in a n-node sensor network with an expected latency of O(log n).

In their model, there are two assumptions. One is that each sensor node has the capability of detecting whether a collision occurs after transmitting data. Another one is that sensor nodes can adjust their transmission range without any limitation. These assumptions pose some challenging issues for hardware design and the latter assumption is almost impossible when the network scale is very large. A collision-free scheduling method for data collection is proposed in [16] aiming at optimizing energy consumption and reliability. All these work did not discuss the minimal-time aggregation scheduling problem.

The most related work to aggregation scheduling is as follows. The minimum data aggregation time problem was proved NP-hard and a $(\Delta-1)$-approximation algorithm was proposed in [3], where $-1)$-approximation algorithm was proposed in [3], where $\Delta$is the maximum degree of the network graph. Another aggregation scheduling algorithm was proposed in [2], which has a latency bound of$23R+\Delta-18$, where$R$is the network radius and $\Delta$is the maximum degree.

Unfortunately, there are some mistakes in their algorithm and the generated schedules are not collision-free in many cases. We discuss their algorithm in Section V-C. All these algorithms mentioned above are centralized. In many cases centralized algorithms are not practical, especially when the network topology often changes in a large sensor network.

As we know data aggregation does make sense to WSN, but it is an NP-hard problem to find an optimized data aggregation algorithm within a random wireless sensor networks. To acquire an approximate optimized data aggregation algorithm in actual applications, reference [3] tries to optimize this problem according to the following three network topologies: CNS (Center at Nearest Source), SPT (Shortest Paths Tree), GIT (Greedy Incremental Tree). Though data aggregation saves energy by reducing the number of data packets in the network, the benefit of data aggregation depends on the positions of the sources and the network topology. To investigate these factors, [3] also proposed two models of source placement: the event-radius(ER) model, and the random sources (RS) model. In the ER model all sources are located within a distance S of a randomly chosen "event" location, while in the RS model k random nodes are chosen to be sources. We can conclude that data aggregation is very energy-efficient by mathematical analysis and experiment result, and this conclusion is a foundation of our research on data aggregation.

TAG [2] has a deep impact on the basic idea of data aggregation in WSNs, and is the foundation of data aggregation schemes. In TAG system, the whole query-processing is divided into two phases: a distribution phase in which queries are pushed down into the network, a collection phase in which the aggregated results are continually routed up from children to parents. The TAG data aggregation scheme can obtain excellent energy-efficiency and query-processing efficiency but TAG only aggregates data packets belonging to the same query, and does not do the inter-query data aggregation. We propose an algorithm for inter-query data aggregation, and analyze the problems existing in inter-query data aggregation and try hard to solve them.

It is worth to note that the only distributed algorithms for converge cast scheduling were proposed in [17], [18]. However, this work focused on the scheduling problem for data collection in sensor networks, but not data aggregation. In data collection, the sink must receive N packets from all the nodes since data cannot be merged, where $N$is the number of sensor nodes in the network. Thus the lower bound of latency is N. The upper bound of the time latency of this algorithm is $\max(3nk-1,N)$, where nk is the number of nodes in the largest one-hop sub tree. This result has much higher latency than our algorithm because it solves the collection scheduling but not aggregation scheduling.

There are several approaches which use tree structure for collecting and aggregating data. The presented approach in [6], with combining Clustering and Directed Diffusion Protocol [7], could process, collect, and aggregate data of sensor nodes without any dependency to the related environment. This paper, with presenting a dynamic clustering structure, could enable the nodes to join to the nearest head cluster while sending data to the gateway node.

Most of data gathering algorithms focus network life- time and saving energy [4,8-11]. In the TAG (Tiny Aggregation) approach [4], each epoch divides to some time slots and these time slots specify to different levels of routing tree in reversal form. In this manner, each node depends on its situation in the tree, and in its related time slot will send its data. The node synchronization of this approach for sending and receiving data could effectively reduce the average energy consumption.

In Directed Diffusion Approach [12,13] receivers and resources using some attributes for recognizing the produced or required information and the goal of this approach is finding an efficient multi way route between senders and receivers. In this approach, each task is represented as an interest and each interest is a set of attribute-value pairs.

The LEACH (Low-Energy Adaptive Clustering Hierarchy) protocol [14] uses a random approach for distributing energy consumption among the nodes. In this approach, the nodes organize themselves as local clusters and one node roles as a local base station or a cluster head. If the cluster heads can be selected base on a priority permanently and they also can be permanent in the whole life time of system, it is obvious that the bad luck nodes which are selected as the cluster heads will be died soon and the life of all the nodes in their cluster will be finished. Thus, LEACH chooses the cluster head among the nodes which have enough energy randomly.

This can prevent the discharging of the battery of a special node. In addition, LEACH uses local data fusion for compressing the data which should be sent from cluster heads to the base station. FTEP [15] is a dynamic and distributed CH election algorithm based upon two level clustering schemes. If energy level of current CH falls below a threshold value or any CH fails to communicate with cluster members then election process is started which is based on residual energy of sensor nodes.

In EEMC (An Energy Efficient Multi Level Clustering) [16], CHs at each level are elected on the basis of probability function which takes into consideration the residual energy as well as distance factor very efficiently. In this scheme whole information is sent and received by sink node for cluster formation. Steiner Points Grid Routing was proposed by, Chiu- Kuo Liang, et al.[17] In order to reduce the total energy consumption for data transmission between the source node and the sink node, a different virtual grid structure instead of virtual grid in GGR is constructed. The idea is to construct the virtual grid structure based on the square Steiner trees [18].

The paper in [19] presents a new version of LEACH protocol called VLEACH which aims to reduce energy consumption

ISSN 2229-5518

within the wireless network. In this approach, by selecting a Vice-CH, cluster nodes data will always reach the BS; no need to elect a new CH each time the CH dies. This will extend the overall network life time.

In summary, there have been lots of work on in-network aggregation and some work on centralized aggregation scheduling, but no work on distributed aggregation scheduling and the existing aggregation scheduling algorithms still have high latencies.

5 Proposed Work

5.1 Motivation

Data aggregation has been put forward as an essential paradigm for wireless routing in sensor networks [3, 6]. The idea is to combine the data coming from different sources en route – eliminating redundancy, minimizing the number of transmissions and thus saving energy. This paradigm shifts the focus from the traditional address-centric approaches for networking (finding short routes between pairs of addressable end-nodes) to a more data-centric approach (finding routes from multiple sources to a single destination that allows in-network consolidation of redundant data).

In this paper we study the energy savings and the delay tradeoffs involved in data aggregation and how they are affected by factors such as source-sink placements and the density of the network. We also investigate the computational complexity of optimal data aggregation in sensor net-works and show that although it is generally NP-hard, there exist polynomial special cases.

5.2 Network Model

In this paper, we mainly focus on studying the complexities of various data operations in wireless sensor networks. For simplicity, we assume a simple and yet general enough model that is widely used in the community. We assume that n+1 wireless sensor nodes V={v0,v1,v2,….vn} are deployed in a certain geographic region, wherev0is the sink node. Each wireless sensor node corresponds to a vertex in a graph G and two vertices are connected in G iff their corresponding sensor nodes can communicate directly. The graph G is called the communication graph of this sensor network. We assume that links are "reliable": when a node vi sends some data to a neighboring node vj, the total message cost is only 1. We assume that all sensor nodes have a communication range r and a fixed interference range R=⊙(r). Each wireless node has the ability to monitor the environment, and collect some data (such as temperature). Assume that A={a1,a2,…aN} is a totally ordered multiset of N elements collected by all n nodes. Here, N is the cardinality of set A.

For data queries in WSNs, we often need build a spanning tree T of the communication graph G first for pushing down queries and propagating back the intermediate results. Given a tree T, let H(T)denote the height of the tree, i.e., the number of links of the longest path from the root to all leave nodes. The depth of a node vi in T, denoted by hT(vi), is the hop number of the path from the root to vi ,i.e., hTðviÞ¼hTðvi;v0Þ. The subtree of T rooted at a node vi, the parent node ofvi, and the set of children nodes of vi are denoted byT(vi), pT(vi), and Child(vi), respectively.

5.3 Complexity Measures

We will study the time complexity, message complexity, and energy complexity of three different data operations, namely data collection, data aggregation, and data selection. The complexity measures we use to evaluate the performance of a given protocol are worst-case measures. The message complexity (and the energy complexity, respectively) of a protocol is defined as the maximum number of total messages (the total energy used, respectively) by all nodes, over all inputs, i.e., over all possible wireless networks G of n nodes (and possibly with additional requirement of having diameter D and/or maximum nodal degree□) and all possible data distributions of A over V.

5.4 Data Collection

Message, Energy, and Time Complexity: Obviously, the data collection can be done with minimum number of messages using a BFS tree with root v0. We now study the data collection with the minimum energy cost. Apparently, for any element, it should follow the minimum energy cost path from its origin to the sink node v0 in order to minimize the energy consumption, where the weight of each link is the energy needed to support a successful transmission using this link. So minimizing the energy is equivalent to the problem of finding the shortest paths from the sink to all nodes, which can be done distributively in time O(m+nlogn) for a communication graph of n nodes and m links [9]. We call the tree formed by minimum energy paths from the root to all nodes as the minimum energy path tree (MEPT). Then we study the time complexity of data collection. Algorithm 1 presents our efficient data collection method based on a good CDSC. The constructed CDS has the maximum nodal degree at most a constant d, all nodes in CDS can be scheduled to transmit once in constant □¼□ðdÞ time slots without causing interferences to other nodes in CDS. We take□time slots as one round.

Algorithm 1.Efficient Data Collection Using CDS
Input: A CDS C with a bounded degreed, tree TC.
1: Every node vi sends its data to its dominator node ð(vi).
2: for t=1to N do
3: for each node vi 2VCdo
4: If node vi has data not forwarded to its parent,
vi sends a new data to its parent in TC in round t.

Algorithm 1 is a constant approximation for both time complexity and message complexity. However, it is not a constant approximation for energy complexity.

5.5 Data Aggregation

We consider the case when, given any node v and its set of children nodes in a data aggregation tree, the aggregation data produced by node v has size same as the maximum size of data from all children nodes. Typical examples of such aggregation are min, max, average, or variance. In data aggregation, if one node sends information twice, it can always save the first transmission. Hence, the data aggregation should be done using a tree.

In emerging pervasive scenarios, data is collected by sensing devices in streams that occur at several, distributed points of observation. The size of data typically far exceeds the storage and computational capabilities of the tiny devices that have to collect and process them. A general and challenging task is to allow (some of) the nodes of a pervasive network to collectively perform monitoring of a neighborhood of interest by issuing continuous aggregate queries on the streams observed in its vicinity. This class of algorithms is fully decentralized and diffusive in nature: collecting all data at few central nodes of the network is unfeasible in networks of low capability devices r in the presence of massive data sets. Two main problems arise in this scenario: i) the intrinsic complexity of maintaining statistics over a data stream whose size greatly exceeds the capabilities of the device that performs the computation; ii) composing the partial outcomes computed at different points of observation into an accurate, global statistic over a neighborhood of interest, which entails coping with several problems, last but not least the receipt of duplicate information along multiple paths of diffusion. Streaming techniques have emerged as powerful tools to achieve the general goals described above, in the first place because they assume a computational model in which computational and storage resources are assumed to be far exceeded by the amount of data on which computation occurs. In this contribution, we review the main streaming techniques and provide a classification of the computational problems and the applications they effectively address, with an emphasis on decentralized scenarios, which are of particular interest in pervasive networks.

5.6 Traditional Vs. Sensor Network Streaming

It is evident that there are two levels of computation and aggregation in distributed settings. At a low level, each sensor observes a stream of data and needs to efficiently extract and maintain information about it. This is essentially the problem of traditional, centralized streaming which has been extensively studied during the last two decades [15{17]. Aggregation is considered with respect to the individual values comprising the data stream, into a concise summary. This is the subject matter of section 2. At a higher level, all remote sites should coordinate to combine these partial information computed from each device. Here, aggregation is considered with respect to this merging process of creating summaries that describe the entire infrastructure. Obviously, new challenges are imposed in such distributed settings, which we address in section 3. It should be clear that this in-network aggregation model generalizes traditional streaming, in the way that a single data stream can be seen as values distributed along a linear-chain topology [18, section 1.3]. Efficient algorithms for distributed computation, that do not make stringent assumptions about the infrastructure topology, can be readily used for classical streaming problems. In analyzing message, energy, and time Complexity, the total message complexity for data aggregation using any tree T is n, where n is the number of nodes of the network. This is because every node v needs send at least once. We obviously can do data aggregation using any spanning tree and every node only needs to send once.

5.7 Data Selection

In this section, we consider the scenario when we want to find the k th smallest data (or median when k¼N=2) among all N data items stored in n wireless sensor nodes. Here, we assume that each wireless sensor node will store at least one data item, and may store multiple data items. All data items are assumed to have a complete order. In most results here, we use the selection of median as an example to study the complexity.

6 Conclusions And Future Directions

In this paper, we study the time complexity, message complexity, and energy complexity of data collection, algebraic data aggregation, and data selection in WSNs. We first study lower bounds of the complexities for these problems and then present efficient algorithms that achieve asymptotically optimal time complexity, and message complexity. We proposed efficient algorithms for data aggregation when each node will produce a data stream. One problem that needs to be resolved is what the best algorithm is when we do not require that the found data item to be precise, i.e., we allow certain relative errors or additive errors on the found answer. We also need to derive better lower bounds on energy cost and design efficient algorithms for holistic data operations. Another question is to study the time complexity and message complexity for other holistic queries such as most frequent items, number of distinctive asymptotically optimal time complexity, and message complexity. We proposed efficient algorithms for data aggregation when each node will produce a data stream. One problem that needs to be resolved is what the best algorithm is when we do not require that the found data item to be precise, i.e., we allow certain relative errors or additive errors on the found answer. We also need to derive better lower bounds on energy cost and design efficient algorithms for holistic data operations. Another question is to study the time complexity and message complexity for other holistic queries such as most frequent items, number of distinctive items. The last but not the least important is to study lower bounds on complexities, and to design efficient algorithms to address these questions when the communication links are not reliable.

References

[1] M. A. Sharaf, J. Beaver, A. Labrinidis, and P. K. Chrysanthis, “Tina: A scheme for temporal coherency-aware in-network aggregation,” in Proc. of MobiDE Workshop, 2003.

[2] S. C.-H. Huang, P.-J. Wan, C. T. Vu, Y. Li, and F. Yao, “Nearly con stant approximation for data aggregation scheduling in wireless sensor networks,” inINFOCOM, 2007, pp. 366–372.

[3] X. Chen, X. Hu, and J. Zhu, “Minimum data aggregation time prob lem in wireless sensor networks,” in 1st Int’l Conference on Mobile Ad-hoc and Sensor Network–MSN, 2005, pp. 133–142.

[4] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, “Tag: a tiny aggregation service for ad-hoc sensor networks,” inOSDI, 2002.

[5] J. Beaver, M. A. Sharaf, A. Labrinidis, and P. K. Chrysanthis, "Loca tion aware routing for data aggregation for sensor networks," in Post Proc. of Geo Sensor Networks Workshop, 2003.

[6] C. Intanagonwiwat, D. Estrin, R. Govindan, and J. Heidemann, "Im pact of network density on data aggregation in wireless sensor net works," in 22nd Int'l Conference on Distributed Computing Sys tems–ICDCS, 2002.

[7] Considine, F. Li, G. Kollios, and J. Byers, "Approximate aggregation techniques for sensor databases," in ICDE, 2004, pp. 449–460.

[10] D. Chu, A. Deshpande, J. Hellerstein, and W. Hong, "Approximate data collection in sensor networks using probabilistic models," in ICDE, 2006

[11] S. Lindsey and C. Raghavendra, "PEGASIS: Power-efficient gather ing in sensor information systems," inProceedings of IEEE Aero space Conference, vol. 3, Mar. 2002, pp. 1125–1130.

[12] M. Lee, and S. Lee, "Data Dissemination for Wireless Sensor Net works", in Proceedings of the 10 th IEEE International Symposium on Object and Component- Oriented Real-Time Distributed Compu ting (ISORC'07).

[13] C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann, and F. Silva, "Directed Diffusion for Wireless Sensor Networking", IEEE/ACM Transactions on Networking, Vol. 11, no. 1, Feb 2003.

[14] H. Cam, S. Ozdemir, P. Nair, and D.Muthuavina shiappan, "ESP DA: Energy-Efficient and Secure Pattern-based Data Aggregation for Wireless Sensor Networks", in Proceedings of IEEE Sensor- The Second IEEE Conference on Sensors, Toronto, Canada, Oct. 22-24, 2003, pp. 732-736.

15] E. Fasolo, M. Rossi, J. Widmer, and M. Zorzi, "In-Network Aggrega tion Techniques for Wireless Sensor Networks: A Survey", IEEE Wireless communication 2007.

[16] M. Lee and V.W.S. Wong, "An Energy-aware Spanning Tree Algo rithm for Data Aggregation in Wireless Sensor Networks," IEEE PacRrim 2005, Victoria, BC, Canada, Aug. 2005.

[17] M. Ding, X. Cheng, and G. Xue, "Aggregation tree construction in sensor networks," in Proc. of IEEE VTC'03, Vol. 4, Orlando, FL, Oct. 2003.

[18] "The Design Space of Wireless Sensor Networks" by Kay R¨omer and Friedemann Mattern http://www.vs.inf.ethz. ch/publ/papers /wsn-designspace.pdf

[19] http://en.wikipedia.org/wiki/Wireless_Sensor_Network 12. Wendi B. Heinzelman, Anantha Chandrakasan, and Hari Balakrishnan, "Energy Efficient Communication Protocol for Wireless Microsen sor Networks", (33rd Hawaii International Conference on System Sciences, 2000).

[20] I. Akyildiz, W. Su, Y. Sanakarasubramaniam, and E. Cayirci. Wire less sensor networks: A survey. Computer Networks Journal, 38(4):393–422, 2002.