The Impact of Data Aggregation on the Performance of Wireless Sensor Networks: A Survey

Kemal Akkaya
Dept. of Computer Science
Southern Illinois University Carbondale
Carbondale, IL 62901
kemal@cs.siu.edu

Murat Demirbas
Dept. of Computer Science and Engineering
University at Buffalo
Buffalo, NY 14260
demirbas@cse.buffalo.edu

R. Savas Aygun
Computer Science Department
University of Alabama in Huntsville
Huntsville, AL 35899
raygun@cs.uah.edu

Abstract. With the increasing need for different energy saving mechanisms in Wireless Sensor Networks (WSNs), data aggregation techniques for reducing the number of data transmissions by eliminating redundant information have been studied as a significant research problem. These studies have shown that data aggregation in WSNs may produce various trade-offs among some network related performance metrics such as energy, latency, accuracy, fault-tolerance and security. In this paper, we investigate the impact of data aggregation on these networking metrics by surveying the existing data aggregation protocols in WSNs. Our aim is twofold: First, providing a comprehensive summary and comparison of the existing data aggregation techniques with respect to different networking metrics. Second, pointing out both the possible future research issues and the need for collaboration between data management and networking research communities working on data aggregation in WSNs.

1. Introduction

Recent years have witnessed a tremendous growth and interest in the usage of tiny sensors in data gathering by forming large-scale ad hoc wireless sensor networks (WSNs). Several WSNs have been deployed for environmental monitoring (such as monitoring nesting behavior of endangered birds in a remote island [1]), precision agriculture (monitoring of temperature and humidity in vineyards [2]), and military and surveillance purposes (classification and tracking of trespassers [3][4]). It is envisioned that WSNs will be part of the future Internet where real-time information will be queried through the physical sensors deployed almost everywhere in our living environments. This direction suggests that retrieving and processing of large volume of data from WSNs will continue to be one of the most important problems for the researchers in coming years.

However, since sensors have severe resource constraints in terms of power, processing capability, memory and storage, it is a challenging task to provide efficient solutions to data gathering problem. Especially energy limitation has been a pressing issue which affects the design of WSNs at all layers of protocol stack [5][6]. Many researchers have investigated various mechanisms such as shutting down the radio, eliminating control packets, and usage of topology control algorithms etc. [7][8][9] in order to reduce energy consumption in WSNs. Data aggregation is also among those mechanisms which is utilized in order to save energy.

Data aggregation in WSNs is the process of combining multiple data packets into one by looking at their contents and is one of the mechanisms often employed for achieving energy-efficiency. For instance, WSNs may have a lot of redundant data since multiple sensors can sense similar information when they are close to each other. Therefore, there is no need to send the same information to the base-station more than once. Instead, a summary of the readings from those sensors can be sent. Thus, using data aggregation will decrease the number of transmissions in the network reducing the bandwidth usage and eliminating the unnecessary energy consumption in both transmissions and receptions.

Data aggregation in WSNs suggests that the intermediate nodes perform the aggregation incrementally when the data is enroute to the base-station. Therefore, it is sometimes referred as *in-network* data aggregation. Typically, an aggregation tree rooted at the base-station is created where the source sensors are leaf nodes. Each intermediate node may have multiple child nodes to receive data and a parent node to transmit data. Therefore, an intermediate node can combine multiple packets, suppress them and help to calculate part of some aggregation functions such as average, count, maximum and minimum.

While the described data aggregation may help reduce the number of transmissions and hence energy consumption [10][11][12][13], it may also affect other performance metrics such as delay, accuracy, fault-tolerance and security. For instance, data aggregation may cause the nodes to wait for their children to send their data, increasing the delay that the packets face [14][15]. In addition, it can decrease the accuracy of the result received at the base-station since many readings from different sources are eliminated through data aggregation during the packets are enroute [16]. Another issue is providing the fault tolerant transmission of the aggregated packets, given the unreliable nature of the wireless environments. This may require multi-path routing [17] which contradicts with the motivation of aggregation. Finally, security is yet another concern since authentication and encryption are required before packets are safely combined [18][19]. Nevertheless, this necessitates aggregation of encrypted data which is not a conventional problem [20][21].

In this paper, by considering each of these metrics separately we survey the current data aggregation research for each metric under a different subsection. While data aggregation has been studied as a means of saving energy in WSNs, none of previous work considered a comprehensive relation of data aggregation with all of the aforementioned networking metrics. To the best of our knowledge, there is no work which categorizes and summarizes the data aggregation approaches in WSNs with respect to their effect on various networking metrics.

Beside these metrics, data aggregation also affects the processing of multimedia data [22][23][24]. Given the hardware improvements in image and video sensor technologies and their possible usage in many applications, we believe that aggregation of multimedia data will require special network architectures and hence it is very important to identify the challenges in this new research area.

Finally, we also briefly summarize and discuss the database centric approaches to data aggregation in order to demonstrate that there is a strong relation between networking and data management aspects of data aggregation in WSNs. However, this relation did not build the desired collaboration between the networking and data management communities. Therefore, as a final contribution, we emphasize the need for collaboration and state the research challenges in order to fill the gap between the two research communities in the future. We believe that this is extremely important since WSNs are envisioned to be used as database systems where large amount of data will be processed.

It is important to note that we only focus on the networking issues of data aggregation. When other issues such as data processing and information extraction are considered, it is usually referred as data fusion [25][26]. Although data fusion and data aggregation are used interchangeably in the research community, data fusion is a more broad area which includes aggregation as a sub-process and focuses on information rather than data with the use of several interdisciplinary techniques such as signal processing, statistical analysis, machine learning and probability. Therefore the research efforts on data fusion are beyond the scope of this paper. We only provide a brief summary on data fusion research.

This paper is organized as follows. In the next section, we define data fusion and state the differences between data aggregation and fusion. In Section 3 we cover the discussion of data aggregation protocols with respect to various metrics and present possible open research problems. In section 4, we discuss the database centric approaches to aggregation and the gap between the works of networking and data management communities on data aggregation. Finally in section 5, we conclude the paper with a summary and a table of the included protocols.

2. Data Fusion vs Data Aggregation

Data fusion and data aggregation are very interrelated and often used interchangeably in WSN research. While data aggregation is introduced with the need of reducing data redundancy and number of transmissions in WSNs, data fusion has already been used in the past extensively for different systems including wired multi-sensor systems [26]-[31].

The definition of data fusion is given in [26] as follows: "Data fusion is a process that combines data and knowledge from different sources with the aim of maximizing the useful information content, for improved reliability or discriminant capability, while minimizing the quantity of data ultimately retained." A data fusion process is characterized by its ability to combine, possibly uncertain, incomplete, and contradictory data. Hence, data fusion takes multiple sources and forms of data and uses all of this data in a way such that a better picture of the observed phenomena is formed. In this way, the phenomena can be better predicted and understood and in turn controlled. Note that, as a consequence of the data fusion, the resulting information is often abstract, generalized or summarized, and, hence, the amount of data is reduced. The popular and widely accepted process model for data fusion identifies five levels in a data fusion process [32]: Level 0 is the sub-object assessment, level 1 is object assessment, level 2 situation assessment, level 3 impact assessment, and finally level 4 is process refinement. In this hierarchy, data aggregation is employed in level 0 where some signal processing is performed [32].

Applications of data fusion includes disparate fields [26], avionics [27][28], command control [29], remote sensing and identification of weather patterns [30], control of complex machinery and assembly robots and financial analysis [31]. In WSNs the usage of data fusion is numerous. Current data fusion research in WSNs is focused on applying data fusion techniques to WSNs by considering their special constraints. Data fusion in WSNs is used to do filtering, eliminating redundancy, eliminating noise and cleaning data, making predictions based on spatio-temporal characteristics etc. [33][34][35][36]. The aim in most of these approaches is to model the sensor data by using statistical techniques such as Bayesian-based approximation and Gaussian distribution in order to handle the imprecision. Handling imprecision and modeling sensor data enables future estimations, eliminate outliers and hence reduce the number of transmissions.

While data fusion includes a wide range of different techniques that can be studied by different research communities [32], data aggregation on the other hand can be seen level 0 of data fusion where only the unneeded redundancy is reduced in the system and a summary of the data is produced. In WSNs, this is achieved through some in-network processing which is mainly the focus of the networking research community. In this paper, we will only focus on data aggregation in WSNs and investigate its networking aspects. Given that data aggregation may affect the design of WSNs at most layers of the networking protocol stack, we will look at various networking challenges that arise with data aggregation.

3. Effects of Data Aggregation on Network Performance

Extensive use of WSNs in many real-life applications introduced several new network and node level performance metrics. Since sensor nodes have significant battery constraints, energy consumption, lifetime of the network and lifetime of a node are among the most important metrics to measure the effectiveness of a given algorithm in WSNs. Accuracy is another metric which has become important due to uncertain number of data sources that can contribute to a given query. In addition to these contemporary metrics, latency of the received data, security and fault tolerance level of the communication have always been among the priorities for real-time and robust network operation.

We argue that when in-network data aggregation is performed, it will have significant impact on these metrics. In fact, most of the time a combination of the mentioned metrics is simultaneously related and creates certain performance trade-offs in WSNs. While data aggregation has been studied as a means of saving energy in WSNs, none of these works considered a comprehensive relation of data aggregation with all of the aforementioned networking metrics. To the best of our knowledge, there is no work which categorizes and summarizes the data aggregation approaches in WSNs with respect to their effect on various networking metrics. This section provides a comprehensive survey of data aggregation protocols in WSNs as well as suggests some open research problems for future studies.

3.1 Data Aggregation and Energy Efficiency

The main motivation of data aggregation in WSNs is to provide energy efficiency through reducing the number of transmissions by exploiting the redundancy in the sensor data. This has raised the question of how to build efficient data structures to promote in-network data aggregation in WSNs. Specifically, how to create the data gathering tree and select the routes and where to do the aggregation in this tree are the main problems to consider. In this section, we will summarize the protocols which considered data aggregation problem from the energy-efficiency perspective. Note that most of the protocols that will be described here are the first works that attempted to model the data aggregation problem using different approaches. This section also includes some application specific examples of data aggregation.

3.1.1 Data Aggregation for Traffic Reduction

Impact of Data Aggregation in WSNs: Finding the optimal aggregation tree in WSNs was first reduced to a Minimum Steiner Tree problem in [10][13]. This problem is known to be NP-Complete which can be defined as follows: Given a complete graph $G=(V,E)$ and a subset $S < V$ of required vertices then a Steiner tree is a subtree of G that includes all the vertices in S and has the minimum sum of weights. Since optimum solution to this problem is NP-Complete, the authors propose three sub-optimal schemes as a solution in the paper. Basic schemes for the aggregation of data include the Center at the Nearest Source (CNS), here data is aggregated at the source nearest to the destination; Shortest Path Trees (SPT), where data is sent along the shortest path from source to base-station and aggregated at common intermediate hops along the way; and Greedy Incremental Trees (GIT), which builds an aggregation tree sequentially to merge paths and provide more aggregation. These protocols are implemented in the paper and a comparison in terms of energy gain is made by trying different number of sources, transmission range and distance to base-station. The results show that the energy saving through SPT and GIT are more significant since they provide aggregation at all levels of the tree. However, this type of aggregation will cause to increase the latency of packets which will be discussed in section 3.2.

Data Aggregation with Low Level Naming: In [11], application specific in-network data aggregation is suggested to be done as close to the data sources as possible which is similar to the CNS heuristic proposed in [10]. This work utilizes the ideas of Directed Diffusion [37] (a routing approach for WSNs) for naming the data and sensors based on the application. This naming is then exploited in order to perform in-network aggregation. Once the data is known through low level naming, intermediate nodes receiving the data can cache, filter and suppress the data before transmitting to the base-station.

Note that this is similar to what Directed Diffusion does and completely application dependent which is the major difference from the approaches mentioned in [10]. The approach can not only provide 42% traffic reduction with respect to non-aggregation but also handle nested queries with the same mechanisms. This handling is however mostly related to query processing discussed in section 4 and hence will be elaborated in that section.

Data Aggregation with Path Sharing: Directed Diffusion [37] and path sharing can be combined to promote data aggregation in WSNs [38]. This is especially useful when there are multiple base-stations in the network. The authors in [38] argue that directing different base-stations to share as much of a network path as possible toward a common source will reduce network traffic and save energy for the system while maintaining the same throughput. Maximum path sharing will help the network aggregate data easily and efficiently. This not only reduces network traffic but also provides energy savings as well.

The protocol uses a greedy algorithm to construct and maintain dynamic network paths between sources and base-stations. The authors map the path maintenance algorithm to the weighted set-covering problem which is NP-Complete, and use its greedy approximation algorithm as their approach. The approach is compared to opportunistic aggregation where data aggregation is performed in a data tree whenever possible. While the two approaches perform similarly at low density networks, greedy approach performs significantly better at high density networks achieving up to 45% energy savings with respect to opportunistic approach. This is due to the fact that with large number of nodes, the probability that the paths to the base-station will be diverse is very high which reduces the chance for aggregation.

However, this work differs from [10] and [11] in the sense that it considers multiple base-stations rather than one. In addition, the main problem with this approach is that a lossless aggregation method in addition to path sharing may increase network congestion in particular segments of the path. This will not only increase the latency of the packets but also depleting the battery of nodes on the shared paths. Therefore, network traffic level should also be used as a parameter to reinforce the paths.

Energy-efficient Data Aggregation Hierarchy: An interesting approach to figure out the optimal number of aggregators for maximized energy efficiency in WSNs is to employ a randomized algorithm. [39] presents a distributed and randomized algorithm where sensors select optimal number of aggregators after deployment. This paper assumes a general compression function which makes the approach usable for any aggregation function in different applications. This function basically computes the size of output data for a given input data. Thus, it can be a simple function such as sum as well as a complex function such as wavelet compression. The authors also extend their approach to a hierarchical model for a more general framework. The hierarchical aggregation model assumes aggregators at multiple levels up to the base-station which is the last level. The optimal number of aggregators in hierarchical model is also calculated.

The approach in this work only focuses on the number of optimal aggregators and does not handle how the aggregation will be performed. However, the usage of a general compression function is a good metric which can be applied to other data aggregation approaches for performance assessments.

Optimal Data Aggregation: In addition to finding the optimal number of aggregators as discussed in [39], it is important to figure out the optimal data aggregation tree. Although this problem can be NP-Complete, a randomized algorithm can provide a constant time approximation of the optimal aggregation tree [40]. The authors claim that the approach will eliminate the need for special data routing structures for some aggregation functions. Although it is not quite clear how to employ this algorithm in very large scale WSNs, this is an interesting and promising direction which can provide significant energy savings without introducing any computational complexity to both sensors and the base-station.

Some of the mentioned approaches for data aggregation have been exploited in some applications or systems which benefit from data aggregation for reduced energy consumption. For instance, *EScan* is one of such systems that utilizes in-network data aggregation [41]. This work aims to provide a residual energy map of the sensor network in order to help users to take necessary actions depending on the nature of the application. This information is provided through in-network data aggregation of individual sensor reports. Each sensor generates a portion of the energy map (i.e. its own information about the energy levels of the neighbor nodes) and those are aggregated near the sources in order to reduce the number of transmissions to the base-station. The base station will eventually have less accurate but still useful information about the energy levels of the sensor nodes at certain areas of the event region.

Similar ideas are explored in [42] in order to offer energy efficiency in an inventory control application. The authors present an aggregation mechanism which basically sums the number of objects and transmits this information along with the region information to the base-station. The aim is to perform the counting near the sources and send an aggregated report. They present a population estimation algorithm by considering the spatial location of the sensors. The simulation results confirm the effectiveness of their approach by providing %65 energy saving with respect to the baseline.

3.1.2 Open Problems

The creation of data aggregation trees according to different application requirements is still a challenge especially with the increasing number of WSNs applications. Therefore, the approaches such as [40] which determine the optimal aggregation tree can be an interesting direction to pursue for ultimate and scalable solutions.

In addition, employment of heterogeneous nodes for aggregation purposes and determining the best layouts for such nodes in the event areas are possible research areas for future investigation. In fact this type of usage has created a new type of network which is called heterogeneous wireless sensor networks. In those networks, aggregation and computationally expensive operations are planned to be performed in special nodes which are assumed to be more powerful than sensors. In this case, an additional research consideration is how to perform aggregation in a multiple sources-multiple base-stations sensor network which has been mentioned in [38]. Such problem can also be studied under the newly emerged Wireless Mesh Networks [43] which is able to integrate WSNs with other types of networks such as Internet.

3.2 Data Aggregation and Latency

While data aggregation is confirmed to save significant energy in WSNs, there is a price for it: an inherent trade-off between energy and latency. Data aggregation may increase the end-to-end latency for the data packets and hence may not be applicable for the applications where end-to-end latency is critical. This increase is due to increased wait time for the intermediate nodes when performing data aggregation. Each intermediate node may have multiple children and the data gathering process from these children may not always be synchronized due to the unbalanced structure of data aggregation trees, node failures, congestion and packet losses. Therefore, the increased latency for the aggregated packets becomes an issue when data aggregation trees are in use. Increased latency is also closely related to another metric which is called data freshness. It refers to the time between generating a data packet and receiving it at the base-station. The less that time is, the fresher the data will be. Data freshness can be used in some of the applications but we consider it as another version of latency and will not discuss it separately.

In this section, we summarize the aggregation approaches in WSNs for minimizing the latency while at the same time performing the desired level of aggregation. We also make a comparison of the approaches and provide some future research issues.

3.2.1 Delay-constrained Data Aggregation

AIDA: The first work that considers latency minimization along with data aggregation is proposed in [44]. The proposed protocol is an application independent data aggregation (AIDA) which does not depend on the type of application as the previous research does.

The protocol can work between MAC and network layer with any kind of routing approach. The main idea is to concatenate "DOA" number of packets at the MAC layer in order to minimize the number of transmissions and packet overhead for accessing the channel. DOA is the degree of aggregation which is basically the number of packets to be concatenated at once. The concatenated packets are decomposed when reach to the destination. This is an adjustable parameter and the authors propose a feedback based mechanism to adjust it based on the changing traffic conditions. Queuing delay is the factor used in the paper as feedback. They claim that by eliminating the need for accessing the channel and transmission for packets, the energy is saved and end-to-end delay is not affected and even decreases in high loads.

However, their approach may be too complex for resource constrained sensor nodes. In addition, they just consider aggregation at the MAC layer and there is no optimization at the network layer. Therefore the proposed approach is not suitable for implementing typical aggregation functions such as sum, average etc. Furthermore, one of the main disadvantage of their approach is that when the concatenated packet is dropped or lost due to any reason then the recovery process will be very expensive, reducing the energy and latency gains.

WFQ-based Delay-constrained Data Aggregation: Providing delay-constrained delivery of data when data aggregation is in use can be achieved through utilizing some special packet scheduling techniques such as Weighted Fair Queuing (WFQ) [14]. This proposed approach in [14] initially forms an aggregation tree that suits contemporary best-effort traffic and utilizes Weighted Fair Queuing (WFQ) in order to support on-time delivery of delay-constrained (real-time) data. The idea is to identify the longest path in terms of hop counts on the aggregation tree for which the end-to-end delay is acceptable. A work around mechanism is presented to ensure timeliness of packets on unfeasible paths by adjusting the tree so that the packets are aggregated at another relay node that is closer to the gateway (base-station) node.

The authors analytically proved that when a feasible path is found for the longest path in terms of hop counts among the real-time sources, the other sources connecting to this longest path can meet the end-to-end delay bounds. Simulation results show that the approach provides significant increase in timeliness at the price of a slight increase in energy consumption when compared to non-QoS-aware aggregation. The approach maintains the same level of timeliness for low traffic rates and slightly increases deadline misses for reasonably higher rate.

Although the presented approach in [14] is complex to implement, it can provide delivery guarantees in terms of delay when used with constantly generated source data and is suitable for calculating aggregate functions mentioned above. It can work at the network layer as opposed to the approach in [44]. It is among a few works that claims to provide delay guarantees along with data aggregation.

A variance of the approach in [14] is pursued in [45] without utilizing WFQ scheduling. The protocol uses the same queuing model as [14] and aggregates the non-real-time packets based on the application up to allowable MAC Maximum Transfer Unit (MTU) using the similar idea in [44]. The real-time packets however are immediately transmitted from the queue in order to meet the timing constraints. The approach can be seen as a combination of [44] and [14]. However, without any scheduling like WFQ there is no guarantee that the real-time packets will make their deadlines. The approach is good for providing priority to real-time packets and hence helps to reduce their end-to-end delay. In addition, the aggregation up to MTU number of non-real-time packets provides energy efficiency.

QAM-based Data Aggregation: QAM-based aggregation proposes packet scheduling algorithms for data gathering in real-time monitoring applications using a technique called modulation scaling [15]. The idea is to model the transmission energy using the Quadrature Amplitude Modulation (QAM) scheme, which represents the energy consumption as a function of the modulation frequency. By adjusting the bit rate at each node, significant energy savings can be achieved while keeping the latency in check. Fig. 1 shows how energy can be controlled with the adjustment of delay performance for 1 bit transmission. While this solution seems to be promising, it is not clear whether modulation scaling can always be applicable and how it affects the design of WSNs.

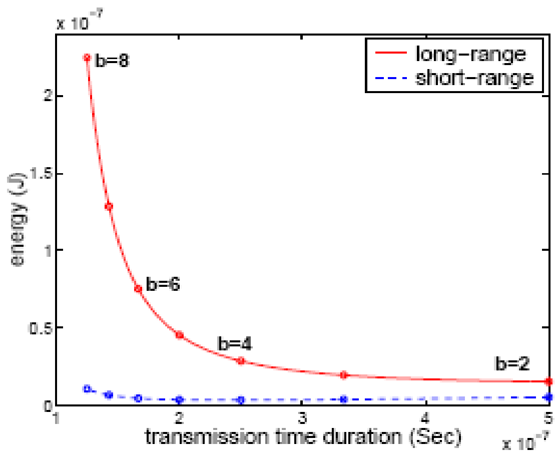


Fig. 1: Energy vs delay in transmission [15]

Determination of Aggregation Points in WSNs: In [46] a mechanism to determine where to aggregate data and how long to wait for achieving an end-to-end latency not exceeding a certain time period is presented. The total deadline is divided into the number of sensors on a path so that each sensor node will know how long it is allowed to wait before sending the packet. This approach is simple however it does not discuss how the accuracy of the received data is affected at the base-station. Depending on the latency constraint, the performance will differ. For instance, if the latency is very small then the waiting time for a sensor will be very short and hence the accuracy of the data at the base-station will not be good. Therefore, this factor should also be included when formulating the solution.

3.2.2 Delay Performance and Network Topology

Apart from the mentioned approaches, another interesting study related to the delay performance of data aggregation is done under the influence of topology control [12]. In this work, the authors study the possible effects of the network topology on the performance of in-network data aggregation in terms of energy, delay and fidelity. This work is a nice example of the inherent trade-offs between maintaining network topology or connectivity with less number of active nodes and performing in-network data aggregation.

The simulation results show that topology control can have a detrimental effect on the network in terms of increased delays. This is because, since connectivity strives to maintain minimum possible number of active nodes for energy considerations, the aggregation process is delayed until some of the nodes wake up. The paper concludes that shorter and fatter aggregation trees should be employed for best energy-delay tradeoff results.

3.2.3 Open Problems

Delay-constrained data aggregation will be an important problem for large scale sensor databases, where users submit frequent aggregation queries and would like to get the results within a certain amount of time. When this problem is modeled as a constrained Steiner-tree problem, it will be an NP-Complete problem [47]. Therefore, efficient heuristics for solving this problem are needed.

In addition, the approaches mentioned in this section mostly consider simple aggregation functions. For complex aggregation functions such as median, histogram etc. the approaches will not work. Hence, delay minimization with complex aggregation is still an open research issue.

Finally, the topology control problem mentioned in [12] is an interesting and novel challenge which involves connectivity, latency and aggregation at the same time. This problem has not been given enough consideration by the research community.

3.3 Data Aggregation and Accuracy

In WSNs, it is very important to decide how much a node will wait for its children to forward the aggregated data to the upper nodes in a data aggregation tree in order to receive accurate data at the base-station. Accuracy is measured by the number of nodes contributing to the result received at the base-station. The more nodes contribute, the more accurate the data will be.

Deciding the number of sensors nodes contributing the result is a difficult problem since such decision can also affect energy and latency performance. While an increased number of readings mean increased accuracy of the result, collecting those readings from the sensors may increase the waiting times at the intermediate nodes. In addition, the increase in the number of contributions will mean more number of transmissions which eventually boosts the energy consumption of the sensor nodes.

This section covers the description of protocols that explore the trade-off among accuracy and other metrics such as energy and latency when performing in-network data aggregation in WSNs.

3.3.1 Determination of Waiting Times

Impact of Timing in Data Aggregation: The effects of data aggregation on accuracy and its inherent trade-off with energy consumption have been studied in [16]. This paper describes three different mechanisms in order to set the time duration for a particular node to wait before aggregating the messages from its children and forwarding it to the upstream node. These timing models namely *periodic simple, periodic per-hop* and *periodic per-hop adjusted* have a strong effect on the accuracy of the received data at the base-station.

In the periodic simple approach, each node is assigned a predefined and fixed amount of waiting time. Periodic per-hop approach on the other hand requires each node to wait for all of its children so that all of them send their data and aggregation can be performed. The third approach is similar to TAG's [48] epoch division approach which will be discussed in detail in section 4. In this approach, the time durations are assigned based on the position of the nodes in the aggregation tree. This is called *cascading timeout* approach where a node's timeout happens right before its parents. Simulation results have shown that cascading timeout approach can provide six times more energy saving with respect to the other two approaches without negatively affecting the accuracy of the data received at the base-station.

Synchronization of Multiple Levels of Data Fusion: There is always a trade-off between latency and accuracy in WSNs. For instance, an increased accuracy will come with an increased latency since the intermediate nodes will have to wait for all their children to contribute to the result [49]. One solution to eliminate this problem is to have a perfectly balanced tree so that two nodes at the same level will face same delay since they will have same number of nodes under their sub-tree. However, it is clear that creating a perfectly balanced tree is not always possible due to large number of sensors, random deployment of sensors and complexity of the tree creation operation. Thus, the authors present a mechanism which is independent from the structure of the tree and assigns waiting times to intermediate nodes based on

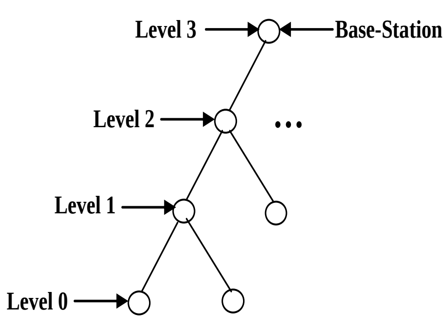


Fig. 2: Multiple levels in an aggregation tree [49].

their proximity to the base-station (See Fig. 2, redrawn from [49]). For instance, a node very close to the base-station should wait more than the node which is far away from the base-station.

The presented mechanism broadcasts to all sensors the level of desired data accuracy and their hop distances to the base-station. The sensors using this information compute their waiting times for data aggregation. The simulation results have confirmed that the proposed synchronization protocol can achieve the desired accuracy irrespective of the aggregation tree structure.

This approach is very similar to the approaches presented in [16] and [48]. The only difference from [16] and [48] is that in this case, the base-station decides a level of accuracy depending on the application and sends a parameter to all the nodes accordingly. The nodes receiving this parameter compute their waiting times for data aggregation. Therefore the level of accuracy is adjustable.

A similar approach to [49] is pursued in [50] without considering any accuracy-latency trade-off. The authors present a biological pulse-coupled oscillator based synchronization of the sensor nodes. This is the system used in pacemaker cells and fireflies. Once this global synchronization is achieved, the sensors can adjust their waiting times for data aggregation as performed in [49]. Different from the approach in [49], they suggest turning the radio off for the nodes which are neither transmitting nor receiving any data. This information will be available to them through global synchronization system. While this can provide significant energy savings, there is no consideration of accuracy in this case.

3.3.2 Energy Consumption vs Accuracy Level

Accuracy-energy tradeoff: Accuracy will be improved with less data aggregation since a whole picture of the data will be in hand when all the data is received. However, this means increased number of transmissions leading to more energy consumption. Therefore, there is always a trade-off between accuracy and energy consumption in WSNs. The work in [67] investigates this trade-off in periodic aggregation applications. These are the applications which require the result of a certain aggregation function at certain periods. The authors present a distributed estimation based aggregation approach. This approach has the ability to adjust the level of data accuracy by changing a threshold value which is very similar to the adjustment mechanism used in [49].

The approach is a nice example of works which both exploits some data fusion algorithms and considers networking aspects when performing data aggregation. On one hand the authors use data fusion methods to compute an estimation of the information. On the other hand, they also explore the trade-off between accuracy and energy in the whole network. The fusion approach presented in the paper considers the spatio-temporal correlation of sensors' data in order to make an estimate of the global result. Each sensor receives an estimation of the global result and compares its own estimate with the received estimate. If the contribution of the sensor node makes a significant difference to the global estimate, a new estimate is made and transmitted. Otherwise, no transmission is made.

Link layer recovery for increased accuracy: An interesting solution to improve the accuracy level can be pursued at the link layer. Since this layer is responsible for improving reliability of the packet delivery, it can be exploited to improve the transmission reliability of the aggregated packets [51]. This is very important since aggregated packets have a lot of information and have been put significant effort to create. If they are not recovered well enough, this may cause a significant degradation in the accuracy of the received information at the base-station.

The authors in [51], therefore, try to make an assessment of how much effort to put and decide on different options of error correction at the link layer depending on the capacity of information carried out by the data packet. For instance, if it is an aggregated packet, more sophisticated techniques such as forward error correction or ARQ can be employed before the transmission of packet at the link layer is performed. While these sophisticated mechanisms can increase data accuracy at the base-station, they can also increase energy consumption of the sensor nodes.

Therefore, depending on the application and the amount of information an aggregated packet carries decision of how to reliably transmit the packet to the base-station is an interesting research challenge which creates a trade-off between energy and accuracy. Note that, this problem is independent from the problem of adjusting waiting times which also affects the accuracy. The solution to this problem will be a complementary to that problem which is studied in [16][49][50].

3.3.3 Open Problems

Since data accuracy is closely related to data content, protocols which consider data processing and networking issues together are needed. Current protocols mostly try to decide the appropriate waiting times for the sensors before performing the data aggregation. However, employing sophisticated data and signal processing techniques can improve the accuracy of the data at the base-station. These approaches should also investigate the amount of resources (i.e. CPU, bandwidth, energy) they can allocate for such techniques. This necessitates novel and comprehensive approaches which can combine network and application layer information.

Another problem is to deal with scalability when considering accuracy performance. For instance, in a very densely populated event region, the number of children for each sensor will probably be very high. In this case, the waiting time for the nodes can be significantly affected depending on the selected approach. Therefore, how the proposed solutions can be adapted to large scale WSNs and how to deal with multi-clustered WSNs will also be important research issues in the future.

Finally, the problem of link layer recovery for aggregated packets, mentioned in [51] can be another interesting future challenge which may promote the research on error control and correction with energy trade-offs.

3.4 Data Aggregation and Fault Tolerance

WSNs are prone to message losses due to the hidden terminal problem inherent in wireless ad hoc communication. Moreover, due to the low-cost, low-power, low-bandwidth nature of WSN radios, several anomalies occur, such as asymmetric links, time-varying quality of links, highly non-deterministic communication at the gray-area band [52][53]. Therefore, fault-tolerance of the data aggregation is crucial for dealing with the unreliable nature of communication channels in WSNs.

In this section, we survey fault-tolerance techniques used by the aggregation protocols under two parts: In the first part, we consider fault-tolerance of various routing structures used for data aggregation which will be referred as *convergecast* structures. In the second part, the orthogonal fault-tolerance mechanisms used for improving the reliability of data transmission will be discussed.

3.4.1 Fault-tolerance of Convergecast Structures

Tree-based structures: A spanning tree, rooted at the base-station, is perhaps the most popular routing structure for aggregation [48][54][55]. The tree construction is often performed by flooding initiated at the base-station, and during the flooding a node selects the first node that it receives a flood message as its parent in the tree. Incidentally, there are several works on the non-uniform and malformed shape of the resulting tree structure in the literature [52][55]. However, the biggest problem with the tree structure is its susceptibility to node and link failures, which results in loss of an entire sub-tree of readings. Therefore, tree-based structures cannot be regarded as fault-tolerant.

Hierarchical/Cluster-based Structures: Disregarding its hierarchical construction, this case can also be viewed, in essence, as a tree-based structure with which it shares the same drawbacks. Slepian-Wolf coding [56], which assumes knowledge of correlation of data and encodes data at the nodes without explicit communication among the nodes, performs very badly due to message-losses when aggregating data over cluster-based structures. This is because; the collector node may not be able to reconstruct several sensor values if the encoded bits from one node are lost. Data aggregation based on explicit communication does not assume knowledge of the correlation structure and exploit data correlation only by receiving explicit communication (information) from other nodes, and hence is more tolerant to message losses. For energy efficient aggregation, a large cluster size is optimal for highly correlated spatial data [57]. However, large clusters are more intolerant to link failures in aggregation, as node or link failure of a cluster-head results in more data to be lost.

Grid-based structures: A grid-based structure can also viewed as a tree-based structure at any snapshot in time; however, convergecast over a grid-based structure can be designed to be very robust and to recover from link and node failures quickly in a local manner [4]. The idea in [4] is to impose a logical grid structure on the network, the base-station being at coordinate (0,0). Each node has four immediate neighbors on the grid, two of which are closer to the base-station than itself. Ideally, a node chooses one of these two neighbors as its parent. In case link or node failures make these two neighbors inaccessible, the node can fall-back on any of the other two non-optimal parents. This case is noted as inserting a back-link, and only up to a certain number of back-links are allowed along any path before declaring the base-station as unreachable and giving up.

Since grid structure allows choosing parent based only on local coordinate information and recovery is also local (not affecting the down neighbors in the grid), the grid structure provides very nice fault-tolerance properties. The grid-based convergecast structure has been successfully used in large-scale real-world field deployments in “Line In The Sand (Lites)” [3] over a 60 node WSN and in “ExScal” [4] over a 1000 node WSN, and proven its scalability and robustness. The tradeoff here with the tree-based approach is the initial time and work required for the set up of the logical grid structure, potentially by using a localization service.

Multipath convergecast structures: For all of the convergecast structures above, at any given time there is at most one path between any node in the network and the base-station. Hence, one link failure can disrupt the aggregation significantly for all downlink readings. Note that the grid structure is also subject to this shortcoming, although its advantage is to be able to recover fast from these failures. The reason most data aggregation protocols insist on the single path rule is to avoid double counting of some data in the aggregation. For example, if a temperature reading transmitted by a node is double counted by two different parents, results of an “average” or “sum” query would be skewed.

Recently there has been some works which try to overcome the double counting problem and provide multi-path support for aggregation to improve fault-tolerance. One suggestion is to use a Directed Acyclic Graph (DAG) [48], and let each node with accumulated value v send v/k to each of its k parents. This way, the effects of a link or node failure on accuracy of aggregated value would be limited as parts of the results in the downlink nodes would still find its way to the base-station through other paths.

A more fundamental solution, though, is to divorce the aggregation from the convergecast structure completely. To this end, approximate “order and duplicate insensitive (ODI) synopsis” is introduced in [17][59]. Approximate (based on randomization) ODI synopses for aggregates such as sum, count, avg, median, and uniform sample, are provided alongside with the error bounds for these approximate answers. The net effect of decoupling aggregation from the convergecast structures is the freedom to use any of the available multi-paths to the base-station at the time, as the effects of double counting are voided through the use of ODI synopses. By routing the message through many of the available paths improve fault-tolerance of aggregation drastically [60].

3.4.2 Mechanisms for Reliable Convergecast

Explicit and Implicit acknowledgments: To deal with transient communication failures due to collisions, acknowledgments are an effective mechanism. The sender, upon failing to receive an acknowledgment from its parent can retransmit its packet to ensure reliable delivery. The acknowledgments do not have to be always explicit. Due to the broadcast nature of the wireless channels, the sender node can snoop on the packets its parent is forwarding to its grandparent to detect if the packet it sent is effectively included, and retransmit the packet if needed.

*TDMA or epoch-based transmission**:*** To avoid hidden terminal problem, time synchronization is a viable option. The idea here is to propagate the results level-by-level up the tree in distinct epochs, where each node waits messages from its children before sending its results in the epoch allocated to itself [48][61][62].

*Gossiping**:*** A gossip-based algorithm where each node gets a copy of the result of aggregation from another node can help overcome the problems associated with a single leader or parent in hierarchical/cluster-based aggregation structures. A hierarchical gossiping algorithm is suggested for fault-tolerance of global function evaluation problem in [63]. Here every node in the network participates in gossiping with a node in other clusters at increasingly higher levels.

Unfortunately, the protocol is not energy-efficient, and does not address the problem of routing messages between far away nodes when nodes are gossiping at higher levels. By counting a communication between any two nodes as one message and spanning one unit time, the message and time complexity of the protocol for an N node WSN is O(Nlog2N) and O(log2N) respectively. However, for two nodes far apart, for example at different cluster partitions, a communication would involve many messages since it has to be conveyed over multi-hops using relay nodes, and would amount to number of messages and time units proportional to the distance between the two nodes. That being said, more lightweight versions of gossiping may still prove to be useful for increasing the fault-tolerance of convergecast.

*Reliable bursty convergecast**:*** Even though a typical WSNs radio can transmit 40Kbits/sec, a node should regulate its data transmission rate as it also needs to relay data from other nodes. Since multiple nodes are originating traffic, rate control and contention management are major issues that needs to be addressed for effective data aggregation. In response to an event such as a detection of a trigger for a query, multiple nodes may start sending several packets in a bursty manner, which may lead to more than %50 data loss [64]. Using implicit acknowledgements, arbitrated retransmission, and prioritization of traffic, a protocol, Reliable Bursty Convergecast (RBC), that achieves reliable transmission is presented in [64]. The RBC protocol has been used in both Lites [3] and ExScal [4] field deployments successfully.

3.4.3 Open Problems

Tradeoffs between fault-tolerance and energy-efficiency require more research. Protocols that optimize both or that can be tunable to (adaptively) substitute one for another would be very useful. In addition, lightweight TDMA or epoch based solutions that can tolerate changes in the topology (due to changes in physical link quality) as well as self-stabilize starting from any arbitrary state are of interest for preventing hidden-terminal problem and for boosting reliable message delivery. Finally, finding more accurate ODI synopses is an active area of research.

3.5 Data Aggregation and Security

Most of the research on data-aggregation in WSNs has assumed that all the participating sensors and base-stations are honest and trustable. However, sensors and other powerful nodes within

WSNs can easily be compromised by physical tampering and other mechanisms. In that case, the aggregation mechanisms should be resilient enough to handle the attacks initiated by those compromise nodes. While the type of these attacks may differ, a typical attack that is related to data aggregation is to send false reports to the base-station so that it will infer wrong results. Most of the protocols that will be described in this section deal with this problem. We discuss and compare the different solutions.

We also surveyed two other types of secure aggregation protocols: 1) The protocols that propose aggregation mechanisms for data which is encrypted or signed. 2) The protocols which are able to adjust encryption level depending on the information carried in the aggregated packet.

3.5.1 Data Aggregation under Compromised Sensors

Secure Information Aggregation (SIA): In order to achieve secure and reliable data aggregation, security mechanisms which prevent the user from making false decisions at the base-station is needed. The work in [65] is known to be the first work which considers security in data aggregation in WSNs. The work provides a security framework in order to develop algorithms for computation of several aggregation functions such as, max, min, median, sum etc. even though the aggregator and/or a fraction of sensors node are corrupted. The framework which is named as *aggregate-commit-prove* has three steps. First, an aggregator node (other than the base-station) collects the data from the sensors which are authenticated. Second, the aggregator computes the result based on the received authenticated data. And finally, it transmits the result along with a correctness proof to the base-station.

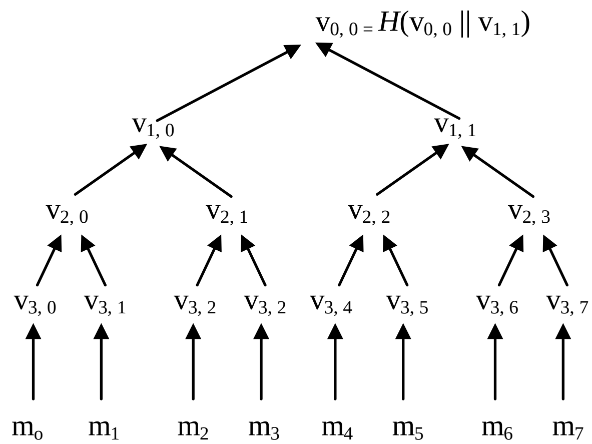


Fig. 3: Merkle hash tree creation [65]

The proofs are created through *Merkle hash trees* in which a hash is computed by concatenation of two children of a node as seen in Fig. 3 which is redrawn from [65]. The base-station by checking the correctness of the proof can conclude whether the aggregator is cheating or not. Note that the paper provides an algorithm for the computation of each aggregation function including median, average, count, maximum and minimum.

*Witness-based aggregation (WBA)***:** The same problem has also been studied in [66] which we will refer as *WBA* throughout this section. The author presents an aggregation approach where some selected witness nodes monitor the aggregation process. In addition to witness nodes, they assume that there is also a data aggregator node which transmits the result to the base-station. Each witness node is responsible for computing the aggregated result, getting a MAC (Message Authentication Code) of the result and forwarding it to the aggregator node. The aggregator node along with the result it has received from the sensors should transmit the proofs (i.e. MACs from witnesses) to the base-station. The base-station can check the proofs and determine whether the aggregated results are correct or not as seen in Fig. 4, redrawn from [66]. In the case of incorrect results, the base-station can pull one of the witness nodes to receive the correct result.

Note that the approach summarized here is very similar to *SIA* discussed in [65]. The only difference is that *WBA* additionally employs some witness nodes to double perform the result computation. In fact both approaches utilize the same idea of using some one way hash functions in sensor nodes for relaying the results to the base-station. While MAC computation is used for creating proofs in *WBA*, Merkle-based hash trees are used in *SIA* in order to verify whether the aggregators are cheating or not. Given that the two approaches follow the same idea, an experimental study would be an interesting research in order to compare their performance in terms of complexity and energy consumption.

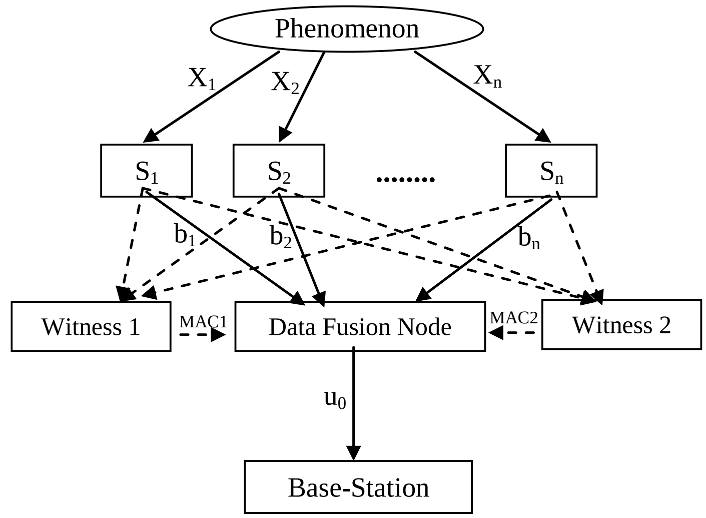


Fig. 4: Verification by two witnesses and an aggregator [66]

*Secure Aggregation (SA)****:*** Yet, another approach to the same problem is discussed in [19]. In this case however, there are neither aggregators nor witness nodes as opposed to *SIA* and *WBA*. Rather, the proposed solution in *SA* involves something called *delayed aggregation and authentication* in order to prevent intruders or compromised nodes to change the result at the base-station.

Delayed aggregation suggests that the aggregation of data at a parent node will not be performed but rather forwarded to the grandparent as is. The grandparent will however perform the aggregation and authentication of the data. The main idea of aggregation at the grandparent is to be able to detect the bogus data coming from the possibly compromised nodes that are close neighbors (i.e. child and parent). For example, in Fig. 5, nodes A and B send their IDs, data R_A, R_B to node E with a MAC of the aggregation. E then forwards this information to its parent G. G finally performs aggregation of R_A and R_B and hence can verify its correctness by computing its MAC and comparing it with the one it has.

In fact *SA* uses MAC computation as done in *WBA* for performing authentication, but it is not performed at each level of the tree. Therefore, it is not clear whether it is as secure as *WBA* unless an experimental evaluation is made. Although this is lacking in the paper, the authors present an analytical cost of the proposed protocol in order to show that the protocol significantly reduces energy and introduces very little overhead when compared to insecure aggregation.

In addition to the three approaches *SIA*, *WBA* and *SA*, [68] proposes another approach called *resilient aggregation*. However, in this approach there is no in-network data aggregation. Aggregation is only performed by a trusted base-station. The paper's approach for preventing the base-station from making false conclusions is based on robust statistics. Robust statistics is the ability of performing aggregation and computing the result even under the noisy and error prone data. Some solutions robust statistics provide are truncating (placing upper and lower bounds on sensor readings) and trimming (eliminating highest and lowest 5% of the readings).

3.5.2 Aggregation of Encrypted Data

A very interesting security problem is the aggregation of encrypted data without decrypting it [20][21]. In fact, the research on this problem is quite necessary since most of the communication among the sensor nodes is envisioned to be encrypted and authenticated in future applications. In such deployments, decrypting the data, performing the aggregation and re-encrypting it for transmission will not only be very inefficient in terms of energy consumption but will also generate unnecessary traffic, decreasing the available bandwidth.

If the aggregation function is known a priori, one solution to this problem is to be able to perform the aggregation with encrypted data which is called *homomorphic encryption*. Using this approach, a different scheme is needed for each aggregation function. For instance performing encrypted addition and average require different algorithms. [20] is a study on these kind of algorithms which include schemes for addition, average and variance. The main issue in this research is to come up with lightweight schemes which will not bring a computationally expensive load to the sensor nodes. However, this type of aggregation may not work for all functions. For example it is not possible to eliminate data redundancy by just looking at the encrypted data.

3.5.3 Encryption Level in Data Aggregation

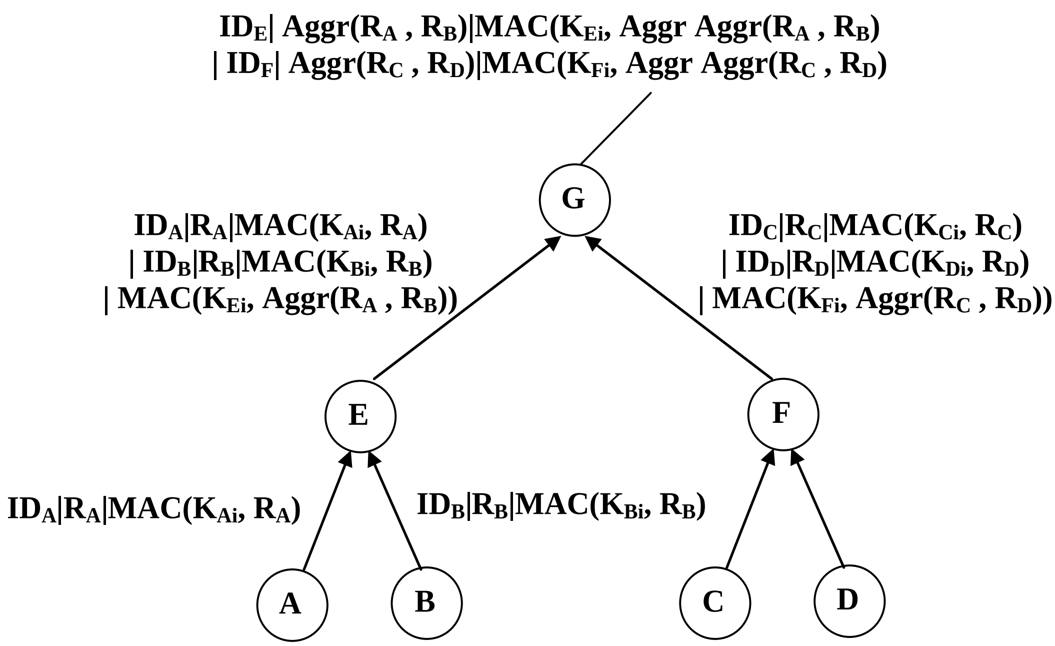


Fig. 5: Delayed aggregation in a sample tree [19].

A slightly different problem for aggregation is providing the privacy of the data. Specifically, this is related to the confidentiality of the relayed aggregated information. The work in [18], Secure and Reliable Data Aggregation (SRDA), claims that the closer the data flows to the base-station, the more information it will contain due to contributions from the intermediate nodes. This means that as data packet travels, more secure algorithms will be needed in order to preserve privacy and confidentiality of the data. SRDA presents a mechanism in which encryption level of the data is strengthened while coming closer to the base-station. This is done by increasing the number of rounds in the underlying RC6 encryption algorithm at each level of the aggregation tree. Note that RC6 is an encryption algorithm which is widely used in wireless environments.

3.5.4 Open Problems

Most of the problems discussed in this section about security in data aggregation are fairly new and very novel. Particularly, aggregation of encrypted data is an interesting study to pursue even though some initial efforts have already been made. The works in [20] and [21] for instance can be extended to be more energy and resource efficient and to cover more complex functions such as histogram, median etc.

In addition, some performance comparison studies are needed in order to assess the best approach (such as *SIA*, *WBA* and *SA*) in terms of energy, delay and accuracy when compromised nodes and/or aggregators falsely contribute to the aggregated results.

It is interesting to note that false contribution is a typical security problem which may also be handled by some fusion algorithms. Data fusion algorithms such as cleaning, outlier detection, and distributed estimation can be utilized to solve this problem since their goal is very similar. This can then be cast as a reliable data aggregation problem.

3.6 Aggregation of Multimedia Data

While data aggregation has been explored in the context of simple functions such as min, max, sum, average etc., very little research has been performed for aggregation and compression of multimedia data that exists in wireless image and video sensor network applications. This is mainly due to current hardware limitations of sensor nodes in the market, however it is expected and envisioned that in the future with the growing capabilities of current MEMS devices, some sensor nodes may have similar computational resources such as a laptop.

Given the broad range of video and image sensor network applications, it is necessary to investigate issues such as image/video compression/aggregation under limited energy resources. In this section, we will first describe the data reduction techniques for the multimedia data and then survey the proposed aggregation techniques for such data in WSNs.

3.6.1 Data Reduction

In general, the reduction in the data size can be investigated in four parts:

Data compression: The compression is one of the most effective methods to reduce the data size. However, there is a tradeoff between the compression and the size of the data to be transmitted. The larger the data, the more energy will be consumed in performing the compression. Thus, the main issue is to identify the ratio of the energy that needs to be utilized for both compression and transmission. Obviously, the advantage of compression is the possibility of data generation close to the original. Most research considers JPEG (Joint Photographic Experts Group) for image and MPEG (Moving Picture Experts Group) for video. The motion-compensated encoding may be employed to reduce the bit-rate [22]. The Discrete Cosine Transform (DCT), which is widely used in JPEG and MPEG, is computationally intensive and it is doubtful whether energy-limited sensors can handle it. The alternate approach is to use Discrete Wavelet Transform (DWT), which can be used in JPEG-2000 and MPEG-4 standards.

Data Elimination: The easiest way to reduce the data size is to reduce the frame rate. Depending on the available power resources, the frame-rate may be reduced. In this case, some of the events might be missed if they happen during frame dropping [22]. For example, the regular video transmission is based on 30 frames per second. If the power resources are limited, the frames may be sent as 1 frame per second (or one frame per 10 seconds).

Data Filtering: The captured data at the sensor may be filtered. Some systems may identify region of interest (ROI). The ROI may be sent with high quality [22]. For example, in surveillance systems, the moving objects are more important than the background. Instead of sending the whole frame, only the blocks that contain the moving objects can be transmitted. The sensors may detect events and transmit data as long as the event continues.

Data Summarization: During aggregation the data may be summarized by extracting high-level information like events and objects in the area. For example, the number of moving objects can be extracted by investigating the blocks in consecutive frames. The neighboring moving blocks are usually assumed to refer a single object. During aggregation, the aggregator node may just keep the number of objects in an area (e.g., shopping center).

3.6.2 Multimedia Data Aggregation

In order to handle multimedia data aggregation, the aggregator nodes need to be more powerful than the sensor nodes since they continuously process to aggregate the incoming data. Multimedia data aggregation can be investigated under low-level and high-level data aggregation.

Low-level data aggregation is similar to image registration or mosaic generation methods. The sensors need to collaborate by maintaining information on cameras that have the same field of view (FOV), the calibration of these cameras, and temporal synchronization of these cameras [22]. For proper and fast aggregation of data sharing the same FOV, the view of cameras can be aggregated on a single view. However, this requires the knowledge of the respective position of the sensors as well as specific information about sensors like focal length. The temporal synchronization ensures that the data to be aggregated does not belong to different time intervals.

In high-level data aggregation on the other hand, the sensors extract useful information and provide it to the aggregator nodes. This type of information may include the number of objects. The aggregator node provides a summary of the number of objects based on the input received from the sensors.

Using the techniques described above, in recent years many protocols have been proposed for handling multimedia data aggregation. These protocols provided both the necessary network architecture and the algorithms for general solutions. Below, we summarize these works.

DFuse: DFuse [69] provides a general aggregation architecture that can also be used in streaming media applications. The authors present a power-aware role assignment framework called DFuse based on a heuristic. DFuse basically creates an overlay network for assigning aggregation functions to certain nodes in the network and monitors the resources of such nodes continuously for possible relocation of aggregation process to other available nodes (See Fig. 6). Hence, the aggregation assignment is not static and can change depending on the network conditions. The framework also provides the programming API for performing different simple aggregation functions as well as complex aggregation functions such as video sequence analyzing.

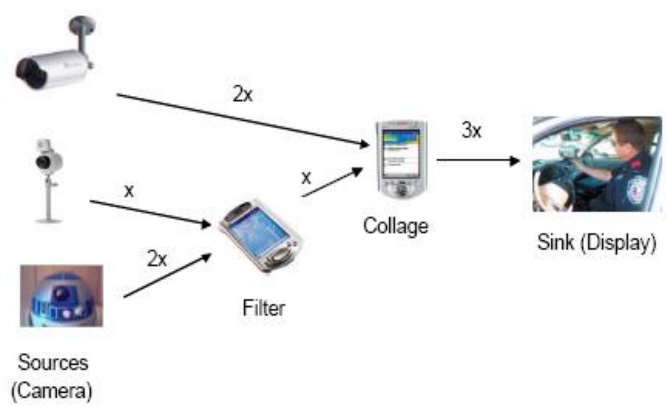


Fig. 6: Aggregation points (filter and collage) for video data coming from cameras [69].

In [70], a system based on DFuse is presented. This system provides aggregation functions like image concatenation, outputting the brightest image, motion detection, face detection, and face recognition. In this system, if a node has significantly more energy than the current aggregation point, then the aggregation point is migrated to that node with more energy.

Camera Sensor Network: In [71], the authors consider an experimental setup of camera sensors where JPEG compression is used. The sensors may trade the quality to the size of the image. The nodes can save energy due to the decrease in the size of data to be transmitted. In their experiments, they show that if the nodes transmit JPEG 1 images (high compression-low quality), the base station is able to cope with high number of requests and images. However, when the JPEG 8 (low compression-high quality), the number of requests and images that can be handled reduce significantly. Therefore, the fidelity of the application can be sacrificed to obtain meaningful readings over a longer time period.

Video-based Sensor Networks: In this network system, the small video sensor devices which have limited capacity to forward the data are directly connected to the base-stations for the aggregation of data [23]. The aggregation is performed only at the base-stations as seen in Fig. 7, since they provide a central storage of the data from the sensors, have more computing power, and can do long-term monitoring. The authors mention that compression is a necessity but computation intensive compression methods like DCT-based compression standards like MPEG is not feasible. Since the sensors may be static, the tracking of objects needs to be accomplished by using multiple cameras. In that case, the data may need to be filtered to avoid information implosion and high priority may be given to important objects.

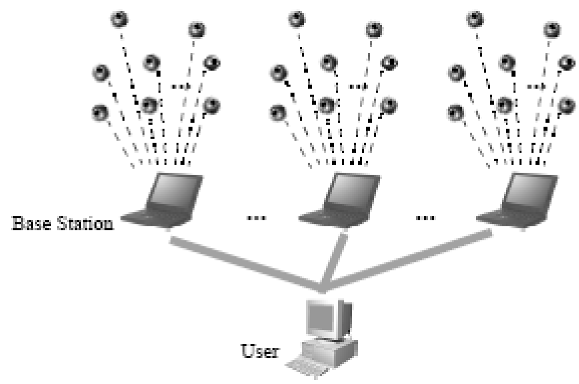


Fig. 7: Wireless Video Sensor Network [23]

Panoptes: Panoptes is a scalable low-power video sensor networking technology that aggregates video data [24]. Panoptes handles the hardware and aggregation issues in multimedia data gathering in WSN. The video sensor which has been used in [24] is called Panoptes (See Fig. 8). There are two types of Panoptes video sensors: the applied data bitsy platform and the crossbow stargate platform. The stargate platform usually yields better performance and aggregation results since it was originally meant to be used in data aggregator nodes. Both platforms are built on top of Linux operating system.

There are three types of compression used: JPEG, differential JPEG, and conditional replenishment. The video sensor applies filtering by comparing the luminance components of the frames pixel-by-pixel. The video sensors use IEEE 802.11 protocol to transmit the data to the aggregating node. The video is only recorded if the changes in macroblocks of an image are higher than a threshold. For querying purposes, an image bitmap is maintained to keep track of the changing macroblocks based on the first frame in the event.

The video aggregation node has 3 components: the camera manager, the query manager, and the stream manager. When video sensors are activated, they register to the camera manager. Video aggregating node may have more than one camera manager to increase the scalability. The Union maps create a single bitmap for an event by employing bitwise OR operation on the bitmaps of the event. The query manager answers the queries by investigating the generated maps as an outcome of the union operation. The stream manager only streams events of interest to the clients.

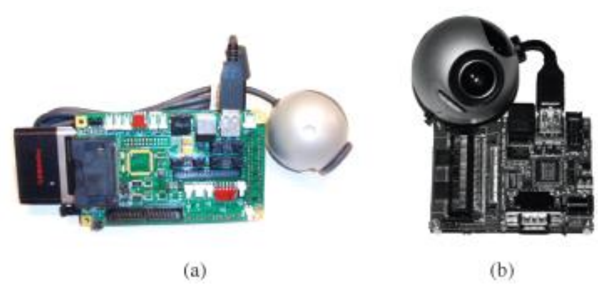


Fig. 8: The types of video sensors [24]

3.6.3 Open Problems

The data aggregation for multimedia data in wireless image and video sensors is challenging due to the limited-power the sensors have. The compression methods are usually image based methods. To the best of our knowledge, no WSN that requires aggregation is utilizing new video standards like MPEG-4. The video data aggregation has not been accomplished. Currently, the aggregation is performed on the images (like JPEG). MPEG-4 may employ Discrete Wavelet Transform (which is considered to be faster than Discrete Cosine Transform) and remove the temporal redundancy. However, the dependency among frames and the necessity of the regeneration of frames by buffering additional frames looks like the handicaps of video-based aggregation although it may provide higher compression.

4. Database Centric Approaches to Data Aggregation

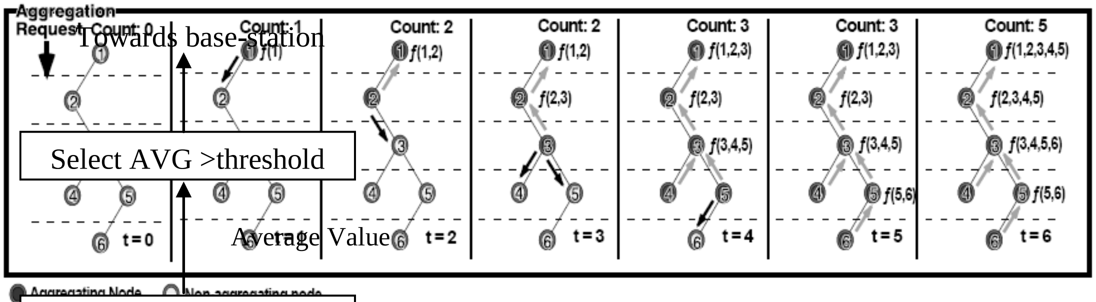
A large group of researchers treat WSNs as a special database where sensors are regarded as tables (data sources) and the base-station is the query generator. In this manner, some of the techniques in traditional databases can be employed when querying the sensor nodes for certain information. Query processing, storage of data, data calibration, data aggregation and sensor data mining are the issues which are yet to be more investigated in sensor databases [35].

Since data aggregation can not be detached from underlying network protocols, it is closely related to the issues we discussed in this paper. Several people by considering sensor network as a database, proposed data gathering and aggregation architectures based on the traditional SQL like models in relational databases. In this section, we summarize and compare these approaches. We believe that the issues in this section require an interdisciplinary research where people from networking and data management can collaborate. We reiterate that one of the goals of this paper is to fill the gap between these two communities when studying data aggregation in WSNs.

4.1 Sensor Database Query and Aggregation Architectures

COUGAR: A data-centric aggregation protocol that views the network as a huge distributed database system is proposed in [72]. The main idea is to use declarative queries in order to abstract query processing from the network layer functions such as selection of relevant sensors etc. and utilize in-network data aggregation to save energy. The abstraction is ensured through a new query layer which is between network and application layer. COUGAR proposes an architecture for the sensor database system where sensor nodes select a leader node to perform the aggregation and transmit the data to the base-station (gateway). The base-station is responsible for generating a query plan which specifies the necessary information about the data flow and in-network computation for the incoming query and sending it to the relevant nodes. The query plan also describes how to select a leader for the query. The architecture provides in-network computation ability for all the sensor nodes as seen in Fig. 9 which is redrawn from [72]. Such ability ensures energy-efficiency especially when the number of sensors generating and sending data to the leader is huge.

Although COUGAR provides a network-layer independent solution for querying the sensors, it has some drawbacks: First of all, introducing another layer (query layer) on each sensor node will bring extra overhead to sensor nodes in terms of energy consumption and storage. Second, in-network data computation from several nodes will require synchronization (i.e. a relaying node should wait every packet from each incoming source) before sending the data to the leader node.



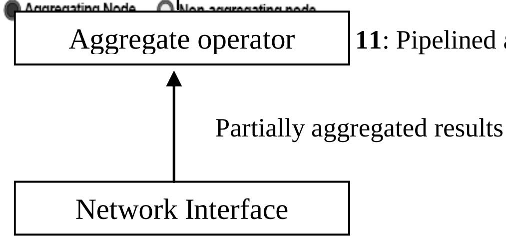


Fig. 11: Pipelined aggregation in Fjords [73]

Fig. 9: Query handling in COUGAR: The leader node gets all the readings, calculates the average and if it is greater than a threshold sends it to the base-station [72].

Third, the leader nodes should be dynamically maintained to prevent them from failure. Fourth, the relaying of data from sources to the leader node and from leader to base-station is another problem. The state of the nodes should be known by the base-station which is another problem. And finally, there is no energy aware algorithms proposed as part of COUGAR which is very crucial for sensor nodes.

Nevertheless, COUGAR's idea of having an independent layer for query handling may increase the efficiency of query processing in large-scale WSNs. This can be a desirable feature in future WSNs. Therefore a joint effort of data management and networking community to solve the above mentioned problems is required.

Similar approach is also pursued in [11]. The main difference from COUGAR is that the authors present an application specific handling which assumes the knowledge of data and geographic locations of sensors. Different than COUGAR, they consider caching and in-network processing at intermediate sensor nodes for energy awareness. In this approach, nested queries are subdivided to be processed by initial and triggered sensors as seen in Fig. 10. While initial sensors are normal sensors, triggered sensors require more resource such as light, image etc. The query is first sent to an initial sensor which then subtasks query to triggered sensors by waking them up. The initial sensor then collects and aggregates data in order to reduce network traffic.

Fjords: Similar to database point of view of querying sensor nodes, the authors in [73] presents data aggregation mechanisms inspired from the solutions provided in database community such as SQL (structured query language). In order to speed up the result of aggregated queries, they suggest using a pipelined aggregation approach which continuously updates the aggregation result. The approach creates a data aggregation tree initially by using broadcasting. The tree consists of multiple levels and each level has some sensors.

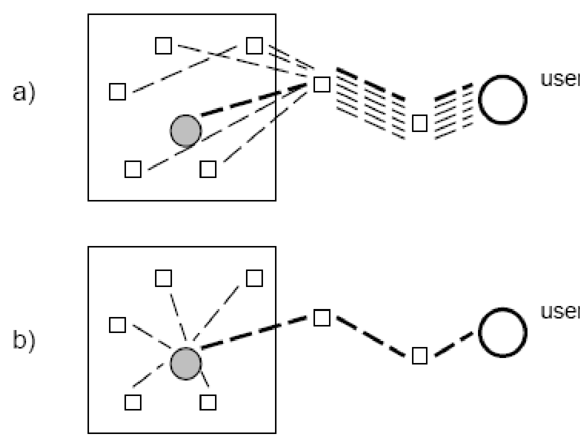


Fig. 10: Nested query handling a) traditional way b) aggregation approach [11]

When an aggregate query for periodic information retrieval is submitted, the nodes on the first level reply to the base-station in a predefined time slot. In the next slot, the nodes on the second level and also on the first level reply. This continues until reaching the leaf nodes (See Fig. 11, redrawn from [73]. When a leaf node is reached, the base-station will have replies from all the nodes at all levels leading to a robust and accurate result. The authors also present mechanisms to handle group by queries similarly.

This work is similar to COUGAR in which WSNs is considered as a database. The database in this work is called Fjord [74].While COUGAR only considers moving selection operators on sensors, Fjord also considers group by queries. In addition, the pipelined approach in Fjord provides multi-level accuracy which is not the case in COUGAR. The authors of Fjord claim that this initial work and others will eventually produce an SQL like query language for WSNs which will be independent from the application.

TAG: Tiny Aggregation (TAG) [48] is another database centric approach to data aggregation which defines an SQL-like declarative query language for expressing aggregation queries over streaming sensor data (See Fig. 12). This query language is applied as an abstraction between the user and the network protocol. Such abstraction gives chance to further apply some data optimizations. They express five SQL aggregate functions (max, min, sum, count, average) and some other functions such as GROUP by using their query language. TAG assumes that the sensors form a single database table and each different sensing capability can be regarded as an attribute in the table as opposed to COUGAR and Fjord.

```
SELECT {agg(expr), attrs} FROM sensors
  WHERE {selPreds}
  GROUP BY {attrs}
  HAVING {havingPreds}
  EPOCH DURATION i
```

Fig. 12: A typical TAG query [48]

TAG has two phases for data collection: *distribution* phase and *collection* phase. It introduces EPOCH clause which represents the time to wait for getting the next answer for the query. The synchronization of the nodes in the tree to receive and transmit data is maintained through the subdivision of EPOCH value into shorter intervals which are called slots. At each slot there is one level of nodes (in the aggregation tree) which are transmitting and another level of nodes which are receiving. This continues similarly up to the level of the base-station as can be observed in Fig. 13.

Note that, this waiting model is different than COUGAR's where a parent node waits for all its children once it builds the list of its children. In TAG, when the child does not have data that conforms to the defined predicate, it does not send the data but rather notifies the parent node that it will not be sending data so that the parent will not wait for that node for performing aggregation. The idea in TAG is to do aggregation whenever possible in the data tree through the base-station. TAG separates data aggregation and routing functions as done in COUGAR, emphasizing that data aggregation should consider what data will be collected rather than how it is collected. Note that this is a different interpretation when compared to the data aggregation approaches we discussed in section 3.3. However, those approaches also consider the network topology when determining the waiting times for different accuracy considerations [16][49]. In fact, some of those approaches utilize the TAG's and COUGAR's approaches in determining the waiting times.

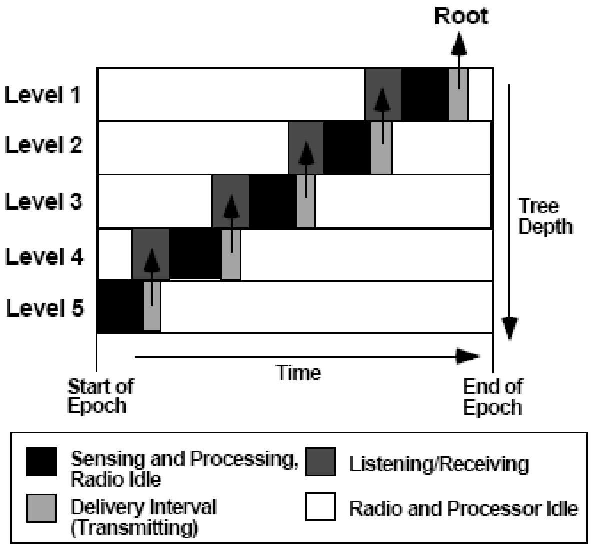


Fig. 13: EPOCH based Aggregation in TAG [48].

```
SELECT {attributes, aggregates}
FROM sensors
GROUP BY {attributes}
EPOCH DURATION i
VALUES WITHIN tct
```

Fig. 14: Query syntax in TiNA [7].

Although TAG's approach provides a very simple interface for efficiently gathering data from WSNs, the determination of slot values and knowledge about the whole network may not be possible in every application. Even if it is possible, it may introduce some overhead to the sensors.

TINA: TiNA [75] (Temporal Coherency Aware in-network Data Aggregation) is a variant of TAG which uses the idea of temporal coherency tolerance. TiNA sends a packet only if the reading of that packet differs from the last sent packet reading by more than a stated tolerance called temporal coherency tolerance (*tct*). This not only eliminates the number of packets transmitted but also helps to reduce the size of aggregated messages at intermediate nodes hence providing significant energy savings. It uses same syntax for expressing the queries and same mechanism for synchronization of the nodes as in TAG. The only extension is a restriction at the end of each query specifying that values within *tct* will not be sent to the parent (See Fig. 14).

Another work which utilizes TiNA's approach is presented in [76]. This is basically a novel aggregation tree construction mechanism that employs TiNA's in-network data aggregation specifically for GroupBy Queries.

A GroupBy query requires information from different groups based on some criteria. The idea in this work is by considering the semantics of the queries to adapt the aggregation tree accordingly so that more energy can be saved. The paper presents two tree construction mechanisms namely GaNC and GaNCi. In this way, the sensors belonging to same group is clustered together, which eventually reduces the number of transmissions by performing the aggregation within the group. The two mechanisms provide significant energy savings when compared to conventional aggregation mechanisms. They are even more effective when used along with TiNA.

4.2 Open Problems

WSNs when viewed as a database will be requiring most of the services of traditional relational databases. However, given that sensors have significant resource problems, a lot of research is needed in order to reach that stage. In particular, more collaborative research is needed as we discussed. For instance, while designing query languages for WSNs, a number of networking aspects such as routing protocol, aggregation points, base-station location, recovery mechanisms, fault tolerance and security should also be taken into account. While a separate layer for query processing and handling is desirable in terms of performance perspective, such separation should not affect the design of low level networking mechanisms of WSNs. It is important to note that this category is expected to produce a vast amount of research in the future.

5. Conclusion

In-network data aggregation in WSNs attracted a lot of attention from the research community due to its potential for reducing the energy consumption of severely constrained sensor nodes.

In this paper, we surveyed the protocols which either employs data aggregation for energy saving purposes or study the computation of some aggregate queries in WSNs. After distinguishing between data fusion and data aggregation, we summarized how data aggregation impacts networking performance in terms of latency, accuracy, energy, fault-tolerance and security by surveying several protocols. We then looked at how aggregation can be implemented when WSNs is considered as a huge database. In addition, we investigated the approaches for performing aggregation of multimedia data in WSNs. Below we provide a table (Table 1) which summarizes the approaches we have covered in this paper.

Although, the summarized approaches solved many interesting problems and contribute to the development of WSNs, there are still many open issues to be investigated as we discussed under each section. We envision that design of query languages for WSNs in order to handle aggregate queries will attract more research in the future. Given that sensors are expected to be employed physically almost everywhere in the daily life, robust information retrieval in sensor databases is very crucial. Therefore, providing acceptable latency, accuracy, fault-tolerance and security along with energy efficiency should have to be included in such design. This definitely necessitates collaboration between networking and data management communities. Similar situation holds for the aggregation of multimedia data where signal and image processing issues should be revisited in the context of WSNs.

In addition to above, security and fault-tolerance of data aggregation arise several issues when designing query languages. This is an issue which cannot be neglected due to weak nature of wireless environments towards several security attacks and high rate of packet loss in wireless transmission. Hence, secure and fault tolerant aggregation is another major future area which requires collaborative efforts.

Table 1: Aggregation protocols and their performance effects on different metrics

	Considered Metrics					
Aggregation Approach	Energy	Latency	Accuracy	Security	Fault-tolerance	Multimedia Aggregation
Data Centric Routing [13]	✓	✓				
Data Aggregation with Low-level Naming [11]	✓					
Data Aggregation with Path Sharing [38]	✓					
Data Aggregation Hierarchy [39]	✓					
EScan [41]	✓					
Optimal Data Aggregation [40]	✓					
AIDA [44]	✓	✓				
WFQ-based Aggregation [14]	✓	✓				
QAM-based aggregation [15]		✓				
Aggregation and Topology Control [12]	✓	✓				
Determination of Aggregation Points [46]		✓				
Impact of Timing on Aggregation [16]		✓				
Multi-level fusion [49]		✓	✓			
Energy-accuracy tradeoff [67]	✓		✓			

Accuracy and error-control [51]	✓		✓			
SIA [65]				✓		
WBA [66]				✓		
SA [19]				✓		
Resilient Aggregation [68]				✓		
Aggregation of encrypted Data [20][21]				✓		
SRDA [18]				✓		
Lites [3]					✓	
ExScal [4]					✓	
Synopsis Diffusion [17][59]					✓	
RBC [64]					✓	
COUGAR [72]		✓	✓			
TAG [48]		✓	✓		✓	
TiNA [75]		✓	✓			
Fjords [74]		✓	✓			
DFuse [69]						✓
Camera Sensor Network [71]						✓
Video Sensor Network [23]						✓
Panaptes [24]						✓

References

[1] Alan Mainwaring, Joseph Polastre, Robert Szewczyk, David Culler, and John Anderson, "Wireless sensor networks for habitat monitoring" in the *Proceedings* of *ACM International Workshop on Wireless Sensor Networks and Applications (WSNA'02)*, Atlanta, GA, September 2002.

[2] J. Burrel, T. Brooke and R. Beckwith, "Vineyard Computing: Sensor Networks in Agriculture Production", *Pervasive Computing, IEEE* Volume 3, Issue 1, Jan.-March 2004 Page(s):38 - 45

[3] A. Arora et al., "A line in the sand: A wireless sensor network for target detection, classification, and tracking", *Computer Networks (Elsevier),* 46(5):605-634, 2004.

[4] A. Arora et al., "Exscal: Elements of an extreme scale wireless sensor network", in the *Proceedings 11th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications*, 2005.

[5] I. F. Akyildiz, et al., "Wireless sensor networks: a survey", in *Computer Networks*, Vol. 38, pp. 393-422, March 2002

[6] K. Akkaya and M. Younis, "A Survey of Routing Protocols in Wireless Sensor Networks, " in *Elsevier Ad Hoc Network Journal*, Vol. 3/3 pp. 325-349, 2005

[7] Y. Xu et al., "Geography-informed Energy Conservation for Ad Hoc Routing," in the *Proceedings of the 7th ACM Mobile Computing and Communication (MobiCom 2001),* Rome, Italy, July 2001.

[8] A. Cerpa and D. Estrin, "ASCENT: Adaptive Self-Configuring Sensor Networks Topologies," in the *Proceedings of the 21st International Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2002),* New York, NY, June 2002.

[9] K. Arisha et al., "Energy-Aware TDMA-Based MAC for Sensor Networks," in the *Proceedings of the IEEE Workshop on Integrated Management of Power Aware Communications, Computing and Networking (IMPACCT 2002),* New York, May 2002.

[10] B. Krishnamachari, D. Estrin, and S. Wicker. "The Impact of Data Aggregation in Wireless Sensor Networks", in the *Proceedings of International Workshop on Distributed Event-Based Systems*, 2002.

[11] John Heidemann, Fabio Silva, Chalermek Intanagonwiwat, Ramesh Govindan, Deborah Estrin, and Deepak Ganesan. "Building Efficient Wireless Sensor Networks with Low-Level Naming", in the *Proceedings of the Symposium on Operating Systems Principles, pp. 146-159,* Alberta, Canada, October, 2001.

[12] Vijay Erramilli, Ibrahim Matta, and Azer Bestavros. "On the Interaction between Data Aggregation and Topology Control in Wireless Sensor Networks", in *Proceedings of the First IEEE Communications*

Society Conference on Sensor and Ad Hoc Communications and Networks (IEEE SECON 2004), Santa Clara, CA, October 2004.

[13] B. Krishnamachari, D. Estrin, S.Wicker, "Modeling Data-Centric Routing in Wireless Sensor Networks", in the *Proceedings IEEE Infocom 2002*, New York, NY, 2002.

[14] K. Akkaya, M. Younis and M. Youssef, " Efficient Aggregation of Delay-Constrained Data in Wireless Sensor Networks," in the *Proceedings of Internet Compatible QoS in Ad Hoc Wireless Networks* 2005(to appear).

[15] Yang Yu, Bhaskar Krishnamachari and Viktor K. Prasanna, "Energy-Latency Tradeoffs for Data Gathering in Wireless Sensor Networks," in the *Proceedings of IEEE Infocom* 2004.

[16] Ignacio Solis and Katia Obraczka, "The Impact of Timing in Data Aggregation for Sensor Networks", in the *Proceedings of International Conference on Communications (ICC 04)*, 2004.

[17] J. Considine et al., "Approximate aggregation techniques for sensor databases", in the *Proceedings of the 20th International Conference on Data Engineering (ICDE'04),* 2004.

[18] Hasan Cam, Suat Ozdemir, Devasenapathy Muthuavinashiappan, and Prashant Nair, "Energy-efficient security protocol for wireless sensor networks", in the *Proceedings IEEE VTC Fall 2003 Conference*, October 4-9, 2003, Orlando, FL, USA

[19] L. Hu and D. Evans, "Secure Aggregation for Wireless Networks", in the *Proceedings of Workshop on Security and Assurance in Ad hoc Networks*, Jan 28, 2003, Orlando, FL.

[20] C. Castelluccia, E. Mykletun and G. Tsudik, "Efficient Aggregation of Encrypted Data in Wireless Sensor Networks," in the *Proceedings ACM/IEEE Mobiquitous Conference*, July 2005, San Diego, USA.

[21] Joao Girao and Dirk Westho and Markus Schneider, CDA: Concealed Data Aggregation in Wireless Sensor Networks," in the *Proceedings ACM WiSe*, 2004.

[22] K. Obraczka, R. Manduchi, and J.J. Garcia-Luna-Aceves, "Managing the Information Flow in Visual Sensor Networks," in the *Proceedings of WPMC 2002: The Fifth International Symposium on Wireless Personal Multimedia Communication*, October 2002, Honolulu, HI.

[23] Wu-chi Feng, Jonathan Walpole, Wu-chang Feng and Calton Pu, "Moving Towards Massively Scalable Video-Based Sensor Networks" , in *Proceedings of the Workshop on New Visions for Large-Scale Networks: Research and Applications,* March 12-14, 2001.

[24] Wu-Chi Feng, Ed Kaiser, Wu Chang Feng, Mikael Le Baillif, "Panoptes: Scalable Low-Power Video Sensor Networking Technologies", *ACM Transactions on Multimedia Computing, Communications, and Applications*, Vol. 1, No. 2, May 2005, pages 151-167.

[25] A. Starr and M. Desforges, "Strategies in data fusion - sorting through the tool box", In *Bedworth and O'Brien*, pages 85–90.

[26] David L. Hall. "Mathematical Techniques in Multisensor Data Fusion", Chapter 1, pages 1–34, 1992.

[27] Stan Musick and Raj Malhotra "Chasing the elusive sensor manager" in the *Proceedings of the National Conference on Aerospace and Electronics*, pages 606–613. IEEE, 1994.

[28] R.A. Adrian, "Sensor management", in the *Proceedings of AIAA/IEEE Conference on Digital Avionics System*, pages 32–37. IEEE Computer Society Press, 1993.

[29] Paul G. Gonsalves and Gerard J. Rinkus, "Intelligent fusion and asset management processor", in the *Proceedings of the 1998 IEEE Information Technology Conference*, pages 15–18. IEEE, 1998.

[30] R. Mavrantza, C. Gountromichou, C. Vamvoukakis, M. Kouli, I. Karfakis, M. Santamouri, K. Seymour-St., D. Argialas, V. Karathanassi, K. Papadimatou, E. Michailidou, and D. Rokos, "Multi-sensor image and data fusion for investigating possible epithermal au in the Aegean back-arc", in the *Proceedings of the 2002 Conference on Fusion of Earth Data*, January 2002.

[31] David Lowe, "The relevance and application of information fusion to financial analysis", In *Bedworth and O'Brien,* pages 5–12.

[32] D. L. Hall and J. Llinas, "Handbook of Multisensor Data Fusion", *CRC Press*, 2001

[33] A. Faradjian, J. E. Gehrke, and P. Bonnet. Gadt, "A probability space adt for representing and querying the physical world", in the *Proceedings of the 18th International Conference on Data Engineering (ICDE 2002)*, February 2002.

[34] Reynold Cheng and Sunil Prabhakar, "Managing Uncertainty in Sensor Databases", in *Special Section on Sensor Network Technology and Sensor Data Management, SIGMOD Record*, Vol.32, No.4, pp. 41-46, Dec 2003

[35] Eiman Elnahrawy, "Research Directions in Sensor Data Streams: Solutions and Challenges", DCIS Technical Report DCIS-TR-527, Rutgers University, May 2003

[36] Yao-Jung Wen et al. “Fuzzy Validation and Fusion for Wireless Sensor Networks”, Proceedings of IMECE2004 2004 ASME International Mechanical Engineering Congress and RD&D Expo November 13-19, 2004, Anaheim, CA.

[37] C. Intanagonwiwat et al., "Directed diffusion: A scalable and robust communication paradigm for sensor networks", in the *Proceedings of the 6th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom'00),* Boston, MA, August 2000.

[38] C. Intanagonwiwat, D. Estrin, R. Govindan, and J Heidemann. “Impact of Network Density on Data Aggregation in Wireless Sensor Networks”, in the *Proceedings of the 22nd International Conference on Distributed Computing Systems*, Vienna, Austria, IEEE. July, 2002.

[39] Y. P. Chen, A. L. Liestman, and J. Liu, Energy-Efficient Data Aggregation Hierarchy for Wireless Sensor Networks, in the *Proceedings 2nd International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks (QShine'05)*, Orlando, FL, USA, August 22 - 24, 2005

[40] M. Enachescu, A. Goel, R. Govindan, and R. Motwani, “Scale Free Aggregation in Sensor Networks “, in the *Proceedings of the First International Workshop on Algorithmic Aspects of Wireless Sensor Networks (Algosensors)*, 2004

[41] Yonggang Jerry Zhao, Ramesh Govindan, and Deborah Estrin, “Residual Energy Scan for Monitoring Sensor Networks”, in the *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC'02)*, March 2002.

[42] Tsung-Han Lin, Polly Huang, “Sensor Data Aggregation for Resource Inventory Applications,” in the *Proceedings of the IEEE Wireless Communications & Networking Conference 2005 (WCNC),* New Orleans, LA, March 2005.

[43] Akyildiz, IF, Wang, X. and Wang, W., “Wireless Mesh Networks: A Survey," *Computer Networks Journal (Elsevier)*, March 2005.

[44] T. He, B.M.Blum, J. A. Stankovic, T. F. Abdelzaher, "AIDA: Adaptive Application Independent Aggregation in Sensor Networks", in Special issue on dynamically adaptable embedded systems, *ACM Transaction on Embedded Computing System*, 2003

[45] Sangheon Pack, Jaeyoung Choi, Taekyoung Kwon, and Yanghee Choi, "Application Aware Data Aggregation in Wireless Sensor Networks," in the *Proceedings of IEEE International Workshop on Adaptive Wireless Networks (AWiN)*, November 2005

[46] U. Roedig, A. Barroso, C.J. Sreenan, “Determination of Aggregation Points in Wireless Sensor Networks”, *in the Proceedings of 30th Euromicro Conference (EUROMICRO)*, IEEE Computer Society Press, 2004.

[47] V. P. Kompella et al., “Multicast Routing for multimedia communication,” *IEEE/ACM Transactions on Networking*, pp. 286-292, 1993.

[48] S. Madden, M. Franklin, J. Hellerstein, W. Hong, " TAG: A Tiny Aggregation Service for ad hoc Sensor Networks," in the *Proceedings of OSDI Conf.,* Boston, December 2002.

[49] Yuan W., Krishnamurthy S.V., and Tripathi, S.K., “Synchronization of Multiple Levels of Data Fusion in Wireless Sensor Networks,” in the *Proceedings IEEE GLOBECOM,* 2003.

[50] N. Wakamiya and M. Murata, “Scalable and robust scheme for data fusion in sensor networks,” in the *Proceedings of International Workshop on Biologically Inspired Approaches to Advanced Information Technology (Bio-ADIT),* pp. 112–127, Jan. 2004.

[51] H. Karl, M. Löbbers, and T. Nieberg, "A Data Aggregation Framework for Wireless Sensor Networks", in the *Proceedings of Dutch Technology Foundation ProRISC Workshop on Circuits, Systems and Signal Processing*, November 2003.

[52] A. Woo, T. Tong, and D. Culler, “Taming the underlying challenges of reliable multihop routing in sensor networks”, in the *Proceedings of the first international conference on Embedded Networked sensor systems*, pages 14-27, ACM Press, 2003.

[53] J. Zhao and R. Govindan, “Understanding packet delivery performance in dense wireless sensor networks”, in the *Proceedings of the first international conference on Embedded Networked sensor systems,* pages 1-13, ACM Press, 2003.

[54] S. Madden, M. Franklin, J. Hellerstein, and Wei Hong, “Tinydb: an acquisitional query processing system for sensor networks” *ACM Trans. Database Systems*, 30(1):122-173, 2005.

[55] Y. Yao and J. E. Gehrke, “Query processing in sensor networks.”, in the *Proceedings of the First Biennial Conference on Innovative Data Systems Research (CIDR 2003)*, 2003.

[56] D. Ganesan, B. Krishnamachari, A. Woo, D. Culler, D. Estrin, and S. Wicker, "Complex behavior at scale: An experimental study of low-power wireless sensor networks", *Technical Report CSD-TR 02-0013*, UCLA, 2002

[57] S. Coleri and P. Varaiya, "Fault tolerance and energy efficiency of data aggregation schemes for sensor networks", in the *Proceedings of IEEE VTC*, 2004.

[58] S. Pattem, B. Krishnamachari, and R. Govindan, "The impact of spatial correlation on routing with compression in wireless sensor networks", in the *Proceedings of the Third International Symposium on Information processing in sensor networks (IPSN'04*), pages 28-35, 2004.

[59] Y. Choi et al., "Stabilization of grid routing in sensor networks", *AIAA Journal of Aerospace Computing, Information, and Communication,* 2005.

[60] Suman Nath, Phillip B. Gibbons, Srinivasan Seshan, and Zachary Anderson, "Synopsis Diffusion for Robust Aggregation in Sensor Networks", in the *Proceedings ACM SenSys,* 2004

[61] S. S. Kulkarni and M(U). Arumugam, "Ss-tdma: A self-stabilizing MAC for sensor networks", in *IEEE Press*, (To appear), 2005.

[62] L. F. W. van Hoesel and Paul J. M. Havinga, "A TDMA-based MAC protocol for WSNs", in the *Proceedings ACM SenSys*, pages 303-304, 2004.

[63] I. Gupta et al., "Scalable fault-tolerant aggregation in large process groups", in the *Proceedings of the 2001 International Conference on Dependable Systems and Networks (DSN '01),* 2001.

[64] H. Zhang et al., "Reliable bursty convergecast in wireless sensor networks", in the *Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing (MobiHoc '05),* pages 266-276, New York, NY, USA, 2005.

[65] Bartosz Przydatek, Dawn Song, and Adrian Perrig. SIA: Secure Information Aggregation in Sensor Networks", in the *Proceedings of the 2nd ACM Conference on Embedded Network Sensor Systems (SenSys),* Los Angeles, CA, USA, November 2004.

[66] Wenliang Du and Jing Deng and Yunghsiang S. Han and Pramod Varshney, "A Witness-Based Approach For Data Fusion Assurance In Wireless Sensor Networks", in the *Proceedings IEEE 2003 Global Communications Conference (GLOBECOM)*", 2003.

[67] A. Boulis, S. Ganeriwal, M. B. Srivastava, "Aggregation in sensor networks: an energy - accuracy tradeoff," *Elsevier Ad-hoc Networks Journal (special issue on sensor network protocols and applications),* 2003

[68] David Wagner, "Resilient Aggregation in Sensor Networks", in the *Proceedings ACM Workshop on Security of Ad Hoc and Sensor Networks (SASN '04),* October 25, 2004

[69] R. Kumar, M. Wolenetz, B. Agarwalla, J. Shin, P. Hutto, A. Paul, and U. Ramachandran, "DFuse: a framework for distributed data fusion," in Proceedings of the first international conference on Embedded networked sensor systems. ACM Press, 2003, pp. 114—125

[70] M. Wolenetz, R. Kumar, J. Shin, and U. Ramachandran, "Middleware Guidelines for Future Sensor Networks", in the *Proceedings of Workshop on Broadband Advanced Sensor Networks*, 2004.

[71] John Lach, David Evans, Jon McCune, Jasonm Brandon, Lingxuan Hu, "Power-Efficient Adaptable Wireless Sensor Networks", in *Military and Aerospace Programmable Logic Devices (MAPLD) International Conference 2003,* September 9-11, 2003.

[72] A. Demers et al, "The Cougar Project: A Work-in-Progress Report," *ACM SIGMOD Record*, vol. 34, no. 4, Dec. 2003.

[73] S. Madden, R. Szewczyk, M. Franklin, and D. Culler. Supporting aggregate queries over ad-hoc wireless sensor networks", in the *Proceedings of IEEE Workshop on Mobile Computing Systems and Applications*, 2002.

[74] Samuel Madden and M. Franklin, "Fjording the Stream: An Architecture for Queries Over Streaming Sensor Data," in the *Proceedings 18th International Conference on Data Engineering (ICDE'02),* 2002.

[75] Mohamed A. Sharaf, Jonathan Beaver, Alexandros Labrinidis and Panos K. Chrysanthis, TiNA: A Scheme for Temporal Coherency-Aware in-Network Aggregation, in the *Proceedings ACM MobiDE'03*, San Diego, CA,2003.

[76] Beaver J, Sharaf MA, Labrinidis A, Chrysanthis PK, "Location-aware routing for data aggregation for sensor networks", in the *Proceedings of Geo Sensor Networks Workshop,* 2003.