Network Structure Based Data aggregation Techniques in Wireless Sensor Networks: A Survey

1Ms. Suriya Praba T and 2Ms. R. Sethukkarasi

1Assistant Professor, Department of Computer Science and Engineering, RMK College of Engineering and Technology, Puduvoyal, Thiruvallur Dt- 601206, Tamilnadu, India.
2Professor, Department of Computer Science and Engineering, RMK Engineering College, Kavaraipettai, Thiruvallur Dt- 601206, Tamilnadu, India.

Received 2 September 2016; Accepted 2 December 2016; Published 31 December 2016

ABSTRACT

Wireless sensor networks (WSNs) consist of sensor nodes. These networks have huge applications in battle field surveillance, Environmental monitoring, habitat monitoring, disaster management, security and military etc. Sensor nodes are very small in size and have limited processing capability and very low battery power. This restriction of low battery power makes the sensor network prone to failure. Data aggregation is very crucial and important technique in wireless sensor networks. With the help of data aggregation we reduce the energy consumption by eliminating redundancy. In this paper, we present a survey of data aggregation algorithms on the basis of performance measures such as Energy Efficiency, latency and data accuracy etc.

KEYWORDS: Data Aggregation, Wireless sensor network, Diffusion, security.

INTRODUCTION

Wireless Sensor Network (WSN):

Nowadays, due to the high popularity of various systems such as laptops and cell phones, The Global Positioning System (GPS) devices and the Radio Frequency Identification (RFID) devices, are become cheaper, mobile and distributed. The emergence of Wireless Sensor Networks (WSN) is the recent trend of Moore’s law towards miniaturization and ubiquity of these devices [1]. WSN is a large network which consists of a group of spatially distributed autonomous sensors interconnected by means of wireless communication channels to monitor physical or environmental conditions, such as temperature, sound, pressure, etc. and to cooperatively pass their data through the network to a source location. Sensor node is a small device, which is used to gather data from its neighboring area, performs simple computations, and communicate with other sensors or with Base Station (BS) [2].

The sensors periodically sense the data, process it and transmit it to the base station. The frequency of data reporting and the number of sensors which report data usually depends on the specific application. Since sensor nodes are energy constrained, it is inefficient for all the sensors to transmit the data directly to the Base station. Hence, the energy conservation techniques are required to provide higher efficiency. One of the ways to achieve energy conservation is Data Aggregation which is elaborately explained in next session.
1.2 Data Aggregation:

Data aggregation is the process of collecting the sensed data by an aggregation mechanism [6]. The main aim of data aggregation algorithms is gathering and aggregating the data in an energy efficient way and improving the network lifetime. It is a technique used to solve the implosion and overlap problems in data centric routing. Data coming from multiple sensor nodes are aggregated as if they are about the same attribute of the phenomenon when they reach the same routing node on the way back to the sink. Robustness and accuracy of information obtained by intact network is improved by data aggregation mechanism. In the data gathered from sensor nodes, assured redundancy may exist. To decrease the redundant information, the data fusion processing is required. The traffic load reduction and energy preservation of the sensors are some of the advantage of data aggregation.

Though there are varieties of challenges in sensor networks, here we focus on communication module which consumes most of the sensor energy. Since communication constitutes aggregation and routing, we are exploring various aggregation strategies. The security issues, data confidentiality and integrity, in data aggregation become vital when the sensor network is deployed in a hostile environment.

This article is structured as follows. In the next section some basic information about the WSNs aggregation and issues related to aggregation are given. Section 2 reports detailed literature survey on Aggregation strategies from the perspective of flat network structure and hierarchical network structure. Section 3 covers security issues and Lastly Section 4 covers the conclusion and possible future directions in this area.

1.3 Issues of Data Aggregation:

A network of energy-constrained sensors deployed over a region is considered; each sensor monitors its surrounding area and periodically generates information.

The systematic gathering and transmission of sensed data to a base station for further processing is the basic operation in such a network. Sensors have the ability to carry out in-network aggregation or fusion of data packets and reroute that to the base station when gathering data. In such sensor system, the lifetime is the time in which the information can be gathered from all the sensors to the base station.

In data gathering, from agreed energy constraints of the sensors expanding the system lifetime is a major threat [8]. The data aggregator node or the cluster head combine the data to the base station and the malicious attacker may attack this cluster node. The base station cannot ensure the accuracy of the aggregate data sent to it, if a cluster head is compromised. Due to the uncompromised nodes, the existing systems may send several copies of aggregate results to the base station and the power consumption at these nodes is increased [7].

2. Literature Survey:

Network Structure Based Data aggregation Techniques:

The architecture of the sensor network plays a vital role in the performance of different data aggregation protocols. In this section, we survey several data aggregation protocols which have specifically been designed for different network architectures. Data aggregation attempts to collect the most critical data from the sensors and make it available to the sink in an energy efficient manner with minimum data latency.

Network lifetime, data accuracy, and latency are some of the important performance measures of data aggregation algorithms. The definitions of these measures are highly dependent on the desired application. We now present a formal definition of these measures.

Energy Efficiency:

The functionality of the sensor network should be extended as long as possible. In an ideal data aggregation scheme, each sensor should have expended the same amount of energy in each data gathering round. A data aggregation scheme is energy efficient if it maximizes the functionality of the network. If we assume that all sensors are equally important, we should minimize the energy consumption of each sensor. This idea is captured by the network lifetime which quantifies the energy efficiency of the network.

Network lifetime:

Network lifetime is defined as the number of data aggregation rounds till α% of sensors dies where α is specified by the system designer. For instance, in applications where the time that all nodes operate together is vital, lifetime is defined as the number of rounds until the first sensor is drained of its energy. The main idea is to perform data aggregation such that there is uniform energy drainage in the network. In addition, energy efficiency and network lifetime are synonymous in that improving energy efficiency enhances the lifetime of the network.
Data accuracy:
The definition of data accuracy depends on the specific application for which the sensor network is designed. For instance, in a target localization problem, the estimate of target location at the sink determines the data accuracy.

Latency:
Latency is defined as the delay involved in data transmission, routing and data aggregation. It can be measured as the time delay between the data packets received at the sink and the data generated at the source nodes.

Control overhead:
is any combination of excess or indirect computation time, Residual energy, memory, bandwidth, or other resources that are required to attain a particular goal.

Mobility of sensor nodes:
The network having a set of sensor nodes moving on their own and communicating with the physical environment is called Mobile WSN.

The design of efficient data aggregation algorithms is an inherently challenging task. There has been intense research in the recent past on data aggregation in WSNs. In this survey paper, we present an extensive overview of several data aggregation algorithms. We first present the basic functionality of the specific algorithm being described and its distinct features. We then discuss the performance of the algorithm and compare it with other similar approaches.

Fig. 1: Network Structure Based Routing Techniques

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

2.1.1. Push diffusion:
In the push diffusion scheme, the sources are active participants and initiate the diffusion while the sinks respond to the sources. The sources flood the data when they detect an event while the sinks subscribe to the sources through enforcements. The sensor protocol for information via negotiation (SPIN) [9] can be classified as a push based diffusion protocol. The two main features of SPIN are negotiation and resource adaptation. For successful data negotiation, sensor nodes need a descriptor to succinctly describe their observed data. These Descriptors are defined in SPIN as metadata. The format of the metadata is application specific. For instance, in area coverage problems, sensors that cover disjoint regions can use their unique ID as metadata. The initiating node which has new data advertises the data to the neighboring nodes in the network using the metadata. A
neighboring node which is interested in this kind of data sends a request to the initiator node for data. The initiator node responds and sends data to the sinks. Each node has a resource manager which keeps track of its energy usage. Each node polls its resources such as battery power before data transmission. This allows sensors to cut back on certain tasks when its energy is low. Simulation results show that SPIN performs almost identical to flooding in terms of the amount of data acquired over time. However, SPIN incurs a factor of 3.5 less energy consumption compared to flooding and is able to distribute 60% more data per unit energy compared to flooding. SPIN is also well suited for environments with mobile sensors since the forwarding decisions are based on local neighborhood information. One of the main advantages of SPIN is that topological changes are localized since each node only requires the knowledge of its single hop neighbors. The main disadvantage of SPIN is its inability to guarantee data delivery. For instance, in intrusion detection applications, if the nodes interested in the data are farther away from the source node, and the intermediate nodes are not interested in the data, then the data is not delivered to the sink nodes.

2.1.2 Two phase pull diffusion:

Intanagonwiwat et al. [10] have developed an energy efficient data aggregation protocol called directed diffusion. Directed diffusion is a representative approach of two phase pull diffusion. It is a data centric routing scheme which is based on the data acquired at the sensors. The attributes of the data are utilized message in the network. The interest propagation in directed diffusion. If the attributes of the data generated by the source match the interest, a gradient is set up to identify the data generated by the sensor nodes. The sink initially broadcasts an interest message in the network. The gradient specifies the data rate and the direction in which to send the data. Intermediate nodes are capable of caching and transforming the data. Each node maintains a data cache which keeps track of recently seen data items. After receiving low data rate events, the sink reinforces one particular neighbor in order to attract higher quality data. Thus, directed diffusion is achieved by using data driven local rules. Average dissipated energy which is the ratio of total energy dissipated per node to the number of distinct events seen by sinks and average delay were used as the performance metrics. Simulation results indicate that Directed diffusion has significantly higher energy efficiency than an omniscient multicast scheme in which each node transmits data along the shortest path multicast tree to all sinks. The average dissipated energy in directed diffusion is only 60% of the omniscient multicast scheme. The average delay of directed diffusion is comparable to omniscient multicast. Directed diffusion is an appropriate choice for applications with many sources and few sinks. In directed diffusion, it is not necessary to maintain global network topology unlike SPIN. However, directed diffusion is not suitable for applications which require continuous data delivery to the sink.

Impact of source-destination location on directed diffusion:

The performance of the data aggregation protocol in directed diffusion is influenced by factors such as the position of source and destination nodes and network topology. Krishnamachari et al. [11] have studied the impact of source-destination placement and communication network density on the energy costs associated with data aggregation. The event radius model (ER) and random source (RS) model are considered for source placement. In the ER model, all sources are assumed to be located within a fixed distance of a randomly chosen “event” location. In the RS model, a fixed number of nodes are randomly chosen to be sources. The analytical bounds on energy costs with data aggregation show that significant energy cost saving is achieved when the sources are close together and far away from the sink. The optimal data aggregation tree can be constructed in polynomial time if the set of source nodes are connected. Simulations were performed on a 100 node network with the number of sources varying from 1 to 15 ensuring that the sources were connected. The energy gains due to data aggregation are predominant in networks with a large number of sources that are several hops away from the sink.

![Fig. 2: Interest Propagation in Direct Diffusion](image-url)
2.1.3 One phase pull diffusion:

Two phase pull diffusion results in large overhead if there are many sources and sinks. Krishnamachari et al. [12] have proposed a one phase pull diffusion scheme which skips the Flooding process of directed diffusion. In one phase pull diffusion, sinks send interest messages that propagate through the network establishing gradients. However, the sources do not transmit exploratory data. The sources transmit data only to the lowest latency gradient pertinent to each sink. Hence, the reverse route (from the source to the sink) has the least latency. Removal of exploratory data transmission results in a decrease in control overhead conserving the energy of the sensors. In [12], simulations have been performed comparing push diffusion with one phase pull diffusion. The simulation results show that one phase pull outperforms push diffusion when the source event rate is very high. However, when the sink interest rate is high push diffusion performs better than one phase pull diffusion. A wrong choice of diffusion mechanism results in excessive control overhead. For instance, when the source rate is high and the sink interest rate is low, employing push diffusion results in 80% increase in control overhead compared to one phase pull diffusion.

2.2. Hierarchical networks:

A flat network can result in excessive communication and computation burden at the sink node resulting in a faster depletion of its battery power. The death of the sink node breaks down the functionality of the network. Hence, in view of scalability and energy efficiency, several hierarchical data aggregation approaches have been proposed. Hierarchical data aggregation involves data fusion at special nodes, which reduces the number of messages transmitted to the sink. This improves the energy efficiency of the network. In the rest of this subsection, we describe the different hierarchical data aggregation protocols and highlight their main advantages and limitations.

2.2.1 Data aggregation in cluster based networks:

In energy constrained sensor networks of large size, it is inefficient for sensors to transmit the data directly to the sink. In such scenarios, sensors can transmit data to a local aggregator or cluster head which aggregates data from all the sensors in its cluster and transmits the concise digest to the sink. This results in significant energy savings for the energy constrained sensors. The cluster heads can communicate with the sink directly via long range transmissions or multi hopping through other cluster heads. Recently, several cluster based network organization and data aggregation protocols have been proposed. In this section we discuss three such protocols viz., Low Energy Adaptive Clustering Hierarchy (LEACH), Hybrid Energy Efficient Distributed Clustering Approach (HEED) and clustered diffusion with dynamic data aggregation (CLUDDA).

Heinzelman [13] et al. were the first to propose an energy conserving cluster formation protocol called LEACH. The LEACH protocol is distributed and sensor nodes organize themselves into clusters for data fusion. A designated node (cluster head) in each cluster transmits the fused data from several sensors in its cluster to the sink. This reduces the amount of information that is transmitted to the sink. The data fusion is performed periodically at the cluster heads. LEACH is suited for applications which involve constant monitoring and periodic data reporting. The two main phases involved in LEACH are: setup phase and steady state phase. The setup phase involves the organization of the network into clusters and the selection of cluster heads.

The steady state phase involves data aggregation at the cluster heads and data transmission to the sink. A predetermined fraction of nodes, f, elect themselves as the cluster head during the set up phase. A sensor node i compares a random number n between 0 and 1 with a threshold \( \eta_i \). If \( n > \eta_i \), the sensor node becomes a cluster head. The threshold \( \eta_i \) is given by \( \eta_i = f/(1-f (n \mod(1/f))) \) where \( \mod \) stands for the modulus operator which returns the remainder after division. All elected cluster heads broadcast a message to all the other sensors in the network informing that they are the new cluster heads. All non cluster head nodes which receive this advertisement decide as to which cluster they belong to based on the signal strength of the message received. LEACH employs randomization to rotate cluster heads and achieves a factor of eight improvement compared to the direct approach in terms of energy consumption. LEACH was compared with minimum transmission energy routing (MTE) in which intermediate nodes are chosen such that the sum of squared distances between adjacent nodes of the route is minimized. The simulation results show that LEACH delivers ten times more data than MTE for the same number of node deaths. Although LEACH improves the system lifetime and data accuracy of the network, the protocol has some limitations. LEACH assumes that all sensors have enough power to reach the sink if needed. In other words, each sensor has the capability to act as a cluster head and perform data fusion. This assumption might not be valid with energy constrained sensors. LEACH also assumes that nodes have data to send periodically. In LEACH, all nodes have the same amount of energy capacity in each election round which is based on the assumption that being a cluster head results in same energy consumption for every node. Hence, LEACH should be extended to account for node heterogeneity. In an improved version of this protocol, called LEACH-C [14], cluster formation is performed in a centralized manner by the sink. LEACHC improves the performance of LEACH by 20 to 40 percent in terms of the number of successful data gathering rounds.
Younis et al. [15] have proposed HEED whose main goal is to form efficient clusters for maximizing network lifetime. The main assumption in HEED is the availability of multiple power levels at sensor nodes. Cluster head selection is based on a combination of node residual energy of each node and a secondary parameter which depends on the node proximity to its neighbors or node degree. The cost of a cluster head is defined as its average minimum reachability power (AMRP). AMRP is the average of the minimum power levels required by all nodes within the cluster range to reach the cluster head. AMRP provides an estimate of the Communication cost. At every iteration of HEED, each node which has not selected a cluster head, sets its probability $PCH$ of becoming the cluster head as $PCH = C \times E_{\text{residual}} / E_{\text{max}}$ where $C$ denotes the initial percentage of cluster heads (specified by the user), $E_{\text{residual}}$ is the estimated current residual energy of the node and $E_{\text{max}}$ is its initial energy corresponding to a fully charged battery. Each node sends a $\text{cluster\_head\_msg}$ where the selection status is set to tentative if $PCH$ is less than 1 or final if $PCH$ is 1. A node selects its cluster head as the node with the lowest cost (AMRP) in the set of tentative cluster heads. Every node then changes its probability to $\min(2PCH, 1)$ in the next iteration. The process repeats until every node is assigned to a cluster head.

Inter cluster communication has not been considered in HEED. The performance of HEED has been compared with generalized LEACH (gen-LEACH) proposed in [10]. In gen-LEACH, the routing protocol propagates the node residual energy throughout the network.

The protocols were simulated for varying network sizes. The simulation results show that HEED improves the network lifetime over gen-LEACH. In gen-LEACH the selection of cluster heads is random which may result in rapid death of certain nodes. However, in HEED the cluster heads are selected such that they are well distributed with minimum communication cost. In addition, the energy dissipated in clustering is less in HEED compared to gen-LEACH. This is due to the fact that gen-LEACH propagates residual energy. To conclude, HEED prolongs network lifetime and achieves a geographically well-distributed set of cluster heads.

Recently a hybrid approach [16] has been proposed which combines clustering with diffusion mechanisms. The new data aggregation scheme proposed in [16] is called clustered diffusion with dynamic data aggregation (CLUDDA). CLUDDA performs data aggregation in unfamiliar environments by including query definitions within interest messages. The interest messages of a new query initiated by the sink contain the query and also a detailed definition of the query. The query definition describes the operations that need to be performed on the data components in order to generate a proper response. This new format of the interest message has some interesting features such as interest transformation and dynamic aggregation. Interest transformation utilizes existing knowledge of queries in order to reduce the overhead in processing. CLUDDA combines directed diffusion with clustering during the initial phase of interest or query propagation. The clustering approach ensures that only cluster heads and gateway nodes which perform inter cluster communication are involved in the transmission of interest messages. This technique conserves energy since the regular nodes remain silent unless they are capable of servicing a request. In CLUDDA, the aggregation points are dynamic. The data aggregation task is not assigned to any specific group of nodes in the network. The nodes performing data aggregation change as the locations of source nodes change. Any cluster head or gateway node which has the knowledge of query definition can perform data aggregation.

An interesting feature of CLUDDA is that a query cache is maintained at the cluster heads and gateway nodes. The query cache lists the different data components that were aggregated to obtain the final data. It also contains the addresses of neighboring nodes from which the data messages originated. These addressers can be used to propagate interest messages directly to specific nodes instead of broadcasting. However, the memory requirements for this technique are yet to be investigated. The technique proposed also needs to be implemented and compared with other existing approaches.

### 2.2.2 Chain based data aggregation:

In cluster-based sensor networks, sensors transmit data to the cluster head where data aggregation is performed. However, if the cluster head is far away from the sensors, they might expend excessive energy in communication. Further improvements in energy efficiency can be obtained if sensors transmit only to close neighbors. The key idea behind chain based data aggregation is that each sensor transmits only to its closest neighbor. Lindsey et al. [17] presented a chain based data aggregation protocol called power efficient data gathering protocol for sensor information systems (PEGASIS). In PEGASIS, nodes are organized into a linear chain for data aggregation. The nodes can form a chain by employing a greedy algorithm or the sink can determine the chain in a centralized manner. Greedy chain formation assumes that all nodes have global knowledge of the network. The farthest node from the sink initiates chain formation and at each step, the closest neighbor of a node is selected as its successor in the chain. In each data gathering round, a node receives data from one of its neighbors, fuses the data with its own and transmits the fused data to its other neighbor along the chain. Eventually the leader node which is similar to cluster head transmits the aggregated data to the sink. In the chain based data aggregation procedure in PEGASIS, Nodes take turns in transmitting to the sink. The
greedy chain formation approach used in [17] may result in some nodes having relatively distant neighbors along the chain. This problem is alleviated by not allowing such nodes to become leaders.

The PEGASIS protocol has considerable energy savings compared to LEACH. The distances that most of the nodes transmit are much less compared to LEACH in which each node transmits to its cluster head. The leader node receives at most two data packets from its two neighbors. In contrast, a cluster head in LEACH has to perform data fusion of several data packets received from its cluster members. The main disadvantage of PEGASIS is the necessity of global knowledge of all node positions to pick suitable neighbors and minimize the maximum neighbor distance. In addition, PEGASIS assumes that all sensors are equipped with identical battery power and results in excessive delay for nodes at the end of the chain which are farther away from the leader node. In [17], two other protocols viz., a binary chain based scheme and a three-level chain based scheme have been proposed. In the binary chain based protocol, each node transmits data to a close neighbor in a given level of the hierarchy. The nodes that receive data at each level form a chain in the next higher level of the hierarchy. At the highest level, the leader node transmits the aggregated data to the sink. In the three level scheme, the protocol starts with the formation of a linear chain among all nodes and then it divides them into \( G \) groups. Each group has \( N/G \) successive nodes of the chain where \( N \) is the total number of nodes. Only one node from each group participates in the second level of the hierarchy. The \( G \) nodes in the second level are further divided into two groups so that only three levels are maintained in the hierarchy. Both energy efficiency and delay are considered while evaluating the performance of the above protocols. The metric is computed by multiplying the average energy cost per data gathering round with the unit delay (transmission time for a 2000 bit message) for the scheme. The performance of the algorithms was compared in terms of the \( \text{Energy} \times \text{Delay} \) metric proposed in [17].

Simulation results show that the chain based binary scheme is eight times better than LEACH and 130 times better than the direct scheme for a 50mx50m network. The chain based three level scheme is 5 times better than PEGASIS and 140 times better than the direct scheme for a 100mx100m network. PEGASIS outperforms LEACH by 100 to 200 percent in terms of the number of data gathering rounds for different network sizes. No conclusions can be drawn about the optimality of a single scheme for optimizing the \( \text{Energy} \times \text{Delay} \) metric. The energy costs of transmissions depend on the spatial distribution of nodes which preclude the optimality of a single scheme for all network sizes. However, experimental results indicate that the binary chain based scheme performs the best for small network sizes.

### 2.2.3 Tree based data aggregation:

In a tree based network, sensor nodes are organized into a tree where data aggregation is performed at intermediate nodes along the tree and a concise representation of the data is transmitted to the root node. Tree based data aggregation is suitable for applications which involve in-network data aggregation. An example application is radiation level monitoring in a nuclear plant where the maximum value provides the most useful information for the safety of the plant. One of the main aspects of tree-based networks is the construction of an energy efficient data aggregation tree. In this subsection, we describe the construction of data aggregation trees.

Ding et al. [18] have proposed an energy aware distributed heuristic (EADAT) to construct and maintain a data aggregation tree in sensor networks. The algorithm is initiated by the sink which broadcasts a control message. The sink assumes the role of the root node in the aggregation tree. The control message has five fields: \( \text{ID} \), \( \text{parent} \), \( \text{power} \), \( \text{status} \) and \( \text{hopcount} \) indicating the sensor ID, its parent, its residual power, the status (leaf, non-leaf node or undefined state) and the number of hops from the sink. After receiving the control message for the first time, a sensor \( v \) sets up its timer to \( T_v \). \( T_v \) counts down when the channel is idle. During this process, the sensor \( v \) chooses the node with the higher residual power and shorter path to the sink as its parent. This information is known to node \( v \) through the control message. When the timer times out, the node \( v \) increases its hop count by one and broadcasts the control message. If a node \( u \) receives a message indicating that its parent node is node \( v \), then \( u \) marks itself as a non leaf node. Otherwise the node marks itself as a leaf node.

The process continues until each node broadcasts once and the result is an aggregation tree rooted at the sink. The main advantage of this algorithm is that sensors with higher residual owner have a higher chance to become a non-leaf tree node. To maintain the tree, a residual power threshold \( P_{th} \) is associated with each sensor. When the residual power of a sensor falls below \( P_{th} \), it periodically broadcasts help messages for \( T_d \) time units and shuts down its radio. A child node upon receiving a help message, switches to a new parent. Otherwise it enters into a danger state. If a danger node receives a hello message from a neighboring node \( v \) with shorter distance to the sink, it invites \( v \) to join the tree.

The protocol proposed in [18] was simulated on a sensor field of 160m × 160m. The results show that EADAT extends network lifetime and conserves more energy in comparison with routing methods without aggregation. The results also indicate that with EADAT, the average residual energy of all alive sensors decreases much more slowly compared to the scenario when no aggregation was used. Another interesting observation was regarding the variation of network lifetime with the network density. The network lifetime increases linearly with the network density. The heuristics proposed in EADAT can thus be used to construct
energy efficient aggregation trees. In applications where each sensor node has data to send to the sink in every round of communication, it is essential to maximize the network lifetime. Tan et al. [19] have proposed a power efficient data gathering and aggregation protocol (PEDAP). The goal of PEDAP is to maximize the lifetime of the network in terms of number of rounds, where each round corresponds to aggregation of data transmitted from different sensor nodes to the sink. PEDAP is a minimum spanning tree based protocol which improves the lifetime of the network even when the sink is inside the field. In contrast, LEACH and PEGASIS perform poorly when the sink is inside the sensor field. PEDAP minimizes the total energy expended in each communication round by computing a minimum spanning tree over the sensor network with link

2.2.4 Grid based data aggregation:

Vaidhyanathan et al. [20] have proposed two data aggregation schemes which are based on dividing the region monitored by a sensor network into several grids. They are: grid based data aggregation and in-network data aggregation. In grid-based data aggregation, a set of sensors is assigned as data aggregators in fixed regions of the sensor network. The sensors in a particular grid transmit the data directly to the data aggregator of that grid. Hence, the sensors within a grid do not communicate with each other. In-network aggregation is similar to grid based data aggregation with two major differences viz.,

a) Each sensor within a grid communicates with its neighboring sensors.

b) Any sensor node within a grid can assume the role of a data aggregator. Terms of rounds until the last node die. The simulation results show that LEACH and direct transmission perform the worst while PEGASIS offers a much improved performance.

We now describe these two data aggregation schemes in greater detail. In grid-based data aggregation, the data aggregator is fixed in each grid and it aggregates the data from all the sensors within the grid. This is similar to cluster based data aggregation in which the cluster heads are fixed. Grid-based data aggregation is suitable for mobile environments such as military surveillance and weather forecasting and adapts to dynamic changes in the network and event mobility.

In in-network aggregation, the sensor with the most critical information aggregates the data packets and sends the fused data to the sink. Each sensor transmits its signal strength to its neighbors. If the neighbor has higher signal strength, the sender stops transmitting packets. After receiving packets from all the neighbors, the node that has the highest signal strength becomes the data aggregator. The in-network aggregation scheme is best suited for environments where events are highly localized. Sensors exchange signal strengths with their neighbors to determine the in-network aggregator which is the node with the highest signal strength. A more efficient approach would choose either the in-network or the grid-based scheme on the fly based on the type of event and its mobility. The authors in [20] have proposed a hybrid scheme which combines the salient features of the in-network and grid-based aggregation schemes. The hybrid scheme accomplishes this goal by combining the best of both the approaches. In the hybrid scheme, sensors are initially configured according to the in-network scheme. When an event occurs, the sensor with the most critical information is identified.

The sensors also maintain a table of past events and the corresponding signal strengths. When a sensor detects an event, it checks its table for the previous event and identifies the nature of the event. The in-network scheme is followed if the sensor identifies the event as a localized event. If the signal strength measurements indicate that the event is mobile, it sends the information to a default aggregator which is a grid based aggregation scheme.

Simulations were performed on a 100 node network with random deployment [20]. In terms of the data acquired (throughput), the hybrid scheme and the in-network scheme perform almost identical to the perfect aggregation scheme in which each sensor is assumed to know the best aggregator. In terms of data latency, the hybrid scheme performs much better than the no-aggregation (classic flooding) and grid based schemes. The schemes have also been compared with respect to the total energy consumption of the sensor network. The simulation results indicate that the energy consumed by the grid based scheme is a factor of 2.2 less than the no-aggregation scheme. The in-network scheme and the hybrid scheme achieve a factor of 2.4 improvement compared to the no-aggregation scheme. The results show the superiority of the aggregation schemes compared to a no-aggregation scheme. However, for a more complete performance evaluation, the schemes need to be simulated under more elaborate scenarios such as multiple event detection.

<table>
<thead>
<tr>
<th>Table 1: Performance analysis of various Data Aggregation Protocols</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protocol</td>
</tr>
<tr>
<td>---------------------------</td>
</tr>
<tr>
<td>SPIN [9]</td>
</tr>
<tr>
<td>Directed Diffusion[10]</td>
</tr>
<tr>
<td>One phase Pull Diffusion[12]</td>
</tr>
</tbody>
</table>
3. Security Issues in Data Aggregation:

Data aggregation in Wireless sensor Network refers to exploit the sensed data from the sensors to the gateway node. Data aggregation plays a significant role in Wireless sensor Networks since the aggregation schemes followed here involve in reducing the amount of power consumed throughout data transmission between the sensor nodes. Within the data aggregation of WSN the most important security requirements, confidentiality and integrity, ought to be consummated. Specifically, the fundamental security issue is data confidentiality that protects the sensitive transmitted data from passive attacks, such as eavesdropping. Data confidentiality is especially very important in a hostile environment, where the wireless channel is at risk of eavesdropping. Though there are many methods provided by cryptography, the difficult encryption and decryption operations, like modular multiplications of large numbers in public key primarily based cryptosystems, will assign the sensor’s power quickly [21]. The other security issue is data integrity that prevents the compromised source nodes or aggregator nodes from considerably altering the final aggregation value [22]. Sensor nodes are easy to be compromised because they lack expensive tampering-resistant hardware, and even that tampering-resistant hardware may not continually be reliable. A compromised node will modify, forge or discard messages.

Conclusion:

In this paper we have studied the data communication in sensor networks i.e data aggregation and realized how communication in sensor networks is different from other wireless networks. Wireless sensor networks are energy constrained. Since most of the sensor energy consumed for transmitting and receiving data, the process of data aggregation becomes an important issue and optimization also needed. Efficient data aggregations is not only to provide energy conservation but also remove redundancy data and hence provide useful data only. When the data from source node is send to sink through neighbors nodes in a multi-hop fashion by reducing transmission and receiving power, the energy consumption is low as compared to that of sending data directly to sink that is aggregation reduces the data transmission then the without aggregation. In this paper, we have proposed energy-efficient techniques for data aggregation in wireless sensor networks. Our scheme integrates energy-efficient and data storage mechanisms. This survey paper shows that these techniques not only reduces power consumption but also prolongs the lifetime of a network.

Future Scope:

The Future work will focuses on developing new aggregation algorithms for routing the data from the source to the sink. Our approach should confront with the difficulties of topology construction, data aggregation, data routing, loss tolerance by including several optimization techniques that further decreases the message costs and improve tolerance to failure and loss. Later, we will simulating our developed technique and compare it with some protocols to prove its efficiency.

REFERENCES


