A Survey on Energy Efficient Data Aggregation Protocols for Wireless Sensor Networks

M. Shanmukhi¹ and O. B. V. Ramanaiah²

¹CSE Department, Malla Reddy College of Engineering, Hyderabad, India.
²CSE Department, JNTUH, Hyderabad, India.

Abstract

Wireless sensor networks (WSN) consist of several sensor nodes which sense different parameters of weather, soil, etc, and send the same to sink/base station. The sensor node has very limited battery power that is used for sensing, computing, and communication. Elimination of redundant sensed data and aggregation of nearest sensor values is necessary in order to minimize the communication overhead. Hence to conserve the battery power and thereby enhance the lifetime of a WSN, we need energy efficient data aggregation mechanisms. Our objective is to differentiate the comprehensive study of data representation model in data aggregation protocols based on energy conservation model using networks based systems. And also intends to study about the various data aggregation network based protocols in the aspects of various factor such as Aggregation Method, Resilience to Link Failure, Setup Overhead, Scalability, Resilience in case of node mobility, Energy saving method and Timing strategy.

Keywords: WSN, Data Aggregation Protocols, Battery Conservation and communication overhead.

INTRODUCTION

A Sensor network is a collection of sensor nodes with sensing, computing and communication capabilities. Sensor nodes have limited battery energy, besides it is not possible either to recharge or replenish the node battery once it is deployed in the field [1]. In a deployed WSN, the data gathering and its radio communication to sink node (base station) consumes more energy. Hence we need energy efficient data gathering and processing mechanisms to enhance the network lifetime.

A sensor node is a tiny device that composed of three portions [2]:

- The data is collected from the physical environment using the sensing subsystem.
- Data manipulation and storage using the processing subsystem.
- Data transmission using wireless communication subsystem.

In Wireless Sensor Networks, the efficient use of energy plays a vital issue. In order to extend the lifetime of the sensor nodes, the energy is a very scrimpy segment. The energy sources may be of “economical” or “non-economical”. The economical energy conservation is designated as sending/receiving data, manipulating the queries and promoting the data and queries to its close-by nodes [3]. The non-economical energy conservation leads to idle listening, collision, packet loss, overhearing, control packet overhead and over-emitting the data. The energy savvy technique is mainly subdivided into a Network subsystem and Sensing subsystem. Here we analyzed the sensing subsystem to increase the lifetime of the sensor nodes using the data driven approach named, Data Aggregation [4].

Data aggregation is defined as gathering and aggregating the sensed data to get the meaningful information. It is a fundamental processing method to save energy and effective way for saving the limited resources [5]. The main aim of a data aggregation protocol is the process of gathering sensed data and its aggregation in an energy efficient way so that the sensor network life time is enhanced. Data aggregation eliminates the redundancy and hence reduces the size of the data to be communicated to the sink node. The Figure 1 illustrates the working procedure of an aggregation algorithm [6]. The data is collected by the sensor node is given to the aggregation algorithm. The aggregated data as an output is communicated to the sink node. Data aggregation eliminates the redundancy and hence reduces the size of the data to be communicated to the sink node. Data representation is an important stage in the data aggregation.

Figure 1: Data Aggregation System [6]
The paper is unionized as follows: Section 2 describes the working of Data Aggregation. The various data aggregation based networks are explored in Section 3. In Section 4, energy efficient based network protocols are discussed. A new thought is concluded in Section 5.

DATA AGGREGATION ARCHITECTURE

The working principle of Data Aggregation is as follows [1]:

i. Select the group of nodes and separate into a cluster.
ii. The cluster should satisfy the following parameters such as RSSI, TTL, MSRC, bandwidth, battery consumption to represent the node in a cluster.
iii. On these above said criteria, the Cluster Head (CH) is selected among the nodes in a cluster.
iv. The CH should be responsible for all nodes in a cluster. It should handle the newly arrived nodes, information exchange, information updation and information transfer.
v. If global cost is maximum then nodes in a cluster is chosen as CH and global cost is reestimated.
vii. In data aggregation process, the queries and its data from user end are processed and exchanged to the query processor.
ix. Then these data are gathered by data cube approach and the base station acquire the aggregated data from data cube approach.

DATA AGGREGATION BASED NETWORKS

There are two types of networks namely, Flat Networks and Hierarchical Networks [7].

FLAT NETWORKS

In Flat Networks, all the sensor nodes are equipped with same battery power, and play the same role [7] [8]. In this type of networks, data aggregation is achieved by data centric routing [9]. In data centric routing the sink node transmits query message to the sensor nodes via flooding or other broadcasting techniques. The sensor nodes which have the data matching with the query message send reply back to the sink node.

HIERARCHICAL NETWORKS

The computation and communication burden is high at sink node in flat networks. Consequently a lot of energy is consumed. In Hierarchical Networks, special nodes in the field perform the data aggregation and the aggregated data is sent to the sink node. All the sensor [8] nodes send their sensed data to the special nodes based on the query received. The special nodes reduce the data communicated to the sink by the use of aggregation technique. This results in the reduction of energy consumption. Hence, energy is utilized efficiently in Hierarchical Networks. As a result network life time is increased [10].

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Flat Networks</th>
<th>Hierarchical Networks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Aggregation</td>
<td>performed by the nodes along the path to sink</td>
<td>performed by Leader nodes or CHs</td>
</tr>
<tr>
<td>Overhead</td>
<td>On the nodes in the Communication path to the Sink</td>
<td>Only on CH</td>
</tr>
<tr>
<td>Fault Tolerance</td>
<td>If the sink node fails it results in the breakdown of the entire network</td>
<td>Even if CH fails the network can still be in operation</td>
</tr>
<tr>
<td>Latency</td>
<td>Latency is high because the</td>
<td>Latency is low</td>
</tr>
</tbody>
</table>
communication to sink via multi hop link because communication to the sink through CH

Routing Optimal routing with guaranteed overhead Routing is simple but may not be optimal

Node Heterogeneity It cannot utilize node heterogeneity for improving energy efficiency Node heterogeneity can be used to assign high energy nodes as CHs

ENERGY EFFICIENT ROUTING PROTOCOLS
Routing protocols are very important phase in supporting the Data Aggregation process. The objective of the data aggregation is to lessen the energy consumption. In accord to promote the network aggregation, the sensor nodes should track the packets based on content of the data packets and also select the next hop [10]. Since there is no specific infrastructure in WSNs, the sensor node should meet the energy saving requirements. Routing protocols developed in WSN for energy efficient is tabulated based on seven categories:

Table 2: Routing Protocols in WSN [10]

<table>
<thead>
<tr>
<th>Category</th>
<th>Representative Protocols</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location based protocols</td>
<td>MECN, SMECN, GAF, GEAR, Span, TBF, BVGF, GeRaF</td>
</tr>
<tr>
<td>Data-centric Protocols</td>
<td>SPIN, Directed Diffusion, Rumor Routing, COUGAR, ACQUIRE, EAD, Information-Directed Routing, Gradient Based Routing, Energy-aware Routing, Information-Directed Routing, Quorum-Based Information Dissemination, Home Agent Based Information Dissemination</td>
</tr>
<tr>
<td>Hierarchical Protocols</td>
<td>LEACH, PEGASIS, HEED, TEEN, APTEEN</td>
</tr>
<tr>
<td>Mobility-based Protocols</td>
<td>SEAD, TTDD, Joint Mobility and Routing, Data MULES, Dynamic Proxy Tree-Base Data Dissemination</td>
</tr>
<tr>
<td>Multipath-based Protocols</td>
<td>Sensor-Disjoint Multipath, Braided Multipath, N-to-1 Multipath Discovery</td>
</tr>
<tr>
<td>QoS-based protocols</td>
<td>SAR, SPEED, Energy-aware routing</td>
</tr>
</tbody>
</table>

LOCATION BASED PROTOCOLS
The sensor nodes are covered by their locations is known as location based protocols. The distance between two nodes in a sensor networks is estimated for its location information. Here, we reviewed some of the protocols designed based on the location-oriented routing protocols.

Li and Joseph Y. Halpern [11] proposed a method, Subnetwork that obtains a minimum energy path between the nodes. They computed at least one minimum energy path between the node pairs to enhance the energy saving process. In grid based architecture, the energy consumption is handled by improving its communication security and cost by Melkang Qlu et al [12]. The Access Points (AP) is also used between the node communications. Each node possesses some energy and this energy is harvested for further use was proposed by Liang Liu et al [13]. In Multihop framework, the data packets are transferred between the nodes from Cluster Header. Using the tree based deployment approach the energy transmission ratio is kept fixed which tries to solve the NP-complete problem [14]. When minimum energy is deployed, the maintenance cost is increased simultaneously [15]. To efficiently utilize the energy at low cost, a segment in data link acquires Automatic Repeat Request (ARQ) protocol is utilized. The concept of virtual Multiple Input Multiple Output (MIMO) in a cross layer approach [16] can also used to reduce overall energy consumption in per packet transmission is modeled. The centralized algorithms are employed to localize the energy consumption [17]. A one-to-many protocol in dense networks is used to transfer the information between nodes using transmission power limitation. The broadcasting process is usually consumes a lot of energy. It can also optimized by reducing the retransmissions rate by maximizing the hop length. Without the neighbor information, optimizing the broadcast networks leads to low communication and memory overhead [18] [20] [21]. A kind of trust management is employed between the nodes and sensing report is maintained. The target is to reduce global False Alarm (FA) and Missed Detection (MD). The probabilities rate should be maintained using data fusion techniques [19] [23][24]. Anis Ouni et al framed an energy efficient protocol using mesh networks. A linear programming model is used to show the relationship between the throughput and energy consumption. Thus the combination of single hop and multi-hop routes are generated for continuous power control [25] [26]. In terms of QOS, the Wireless Sensor Networks in data aggregation process. The node reliability and transmission distance should be minimized [27]. In tree based approach, the data transmitting energy is consumed. Applications such as Cognitive radios, the proper use of energy efficient is important. The resource allocation protocols are enabled to obtain a near-optimal with low-complexity channel assignment was initiated and framed by Kandasamy Illankoff et al [28]. Hamed Yousefi et al suggested the concept of scheduling in the data aggregation techniques [29]. The idea behind their research is to minimize the latency time so as to schedule the process and thus in this way the energy can be saved. He suggested a collision free schedule with least number of time slots [30]. In a cluster approach, the resource efficiency and dependability is vital issue in energy saving system. Xiaoyong Li et al framed a lightweight and dependable trust system to avoid the effect of malicious nodes. The cluster member or head feedback is avoided due to dependable trust system which automatically improves the energy-saving [31]. Topology management protocols should be well defined in utilizing the energy in sleep transitions of the nodes. A static topology management protocol is efficient in real time WSN that reduce the bound delay and route fidelity [32]. On the contrary to static process, dynamic topology management was initiated to balance the nodes and enhance the network.
lifetime by managing event driven data transmissions. The process of synchronization between the nodes has established in [33]. Regression analysis is used to estimate the relationship between two variables. In Multihop architecture, the sources of errors and data transmission rate are estimated to find a new relationship between the nodes. The nodes in a cluster may be local or global. A k level hierarchical structure with low energy localized clustering algorithm established with parameters such as residual energy of nodes, the aggregation degree and lifetime of the nodes [34]. Duty cycling is an important issue in densely connected wireless sensor networks. The local information about the node is gathered in the geographic oriented system. The k covered WSNs are estimated by the Cover Sense Inform (CSI) framework which intends to merge the sensor scheduling and data forwarding. The high k coverage is actualized by k active sensors to the sink node to eliminate the intruder detection and tracking [35]. Data fidelity is also used for energy efficiency in data aggregation schemes. The Compressed Sensing (CS) is employed to achieve the data fidelity. The diffusion wavelet characterizes spatial correlated data. The NP complete and integer programming model is instantiated to solve the minimum energy compressed data aggregation problem [36]. In cluster oriented architecture, the data collection process is efficient in handling the communication between the nodes. The node in sleep/ awake scheduling incurs energy. It is handled by an adaptive scheme cluster to cluster propagation scheme to reduce communication cost and limiting prediction cost [37]. Though there was a lot of energy saving methods that dealt with one to one communication, one to many communications etc. In [38], the many to many communications has been proposed using the MUSTER routing protocol. It has ability to operate in multicast networks using multiple sources to multiple sinks which concurrently increase the network lifetime and balance the nodes.

Table 3: Merits and Demerits of Location based Protocols [38]

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Merits</th>
<th>Demerits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trajectory Based Forwarding (TBF)</td>
<td>Improved Reliability Network management is proper.</td>
<td>Overhead is high.</td>
</tr>
<tr>
<td>Geographic Random Forwarding (GeRaF)</td>
<td>Unsettled nodes. Lack of creating multi-hop overhead.</td>
<td>Necessary to initiate the user interaction in each stage.</td>
</tr>
<tr>
<td>Minimum Energy Communication Network (MECN)</td>
<td>Energy sustainability with low power.</td>
<td>Based on specific application, the fault tolerance varies.</td>
</tr>
<tr>
<td>Small Minimum Energy Communication Network (SMECN)</td>
<td>Incurs less energy than the MECN.</td>
<td>Power usage is high.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Size of message broadcasting is high.</td>
</tr>
</tbody>
</table>

DATA CENTRIC PROTOCOLS

A data centric protocol is varied from the Address centric protocols. A sink node forwards the queries to the specific regions and searches the data in that specific region from the sensors. Each data queries are represented by attribute based naming strategy [39]. The research conducted on the data centric protocols are reviewed as follows:

First of all, there should not be any redundant data transmissions to save the energy. Harshavardhan et al [39] framed location aided flooding scheme to efficiently utilize the energy. He used local information about the node which is used to discards the same packet send by a node is transmitted in virtual grid based architecture. In dynamic networks, the node transmission is done in a random manner. Flooding time is estimated to define the node transmission rate. These information are bounded with high probability with the best node mobility has framed by Andrea Clementi et al [40]. Though the node is transmitted from source to destination, an acknowledgement from the receiver is essential to verify whether the data packets are being forwarded to designated destination. In [41], a link correlation is defined between the source and destination nodes. The collective ACK is calculated in single hop routing with low energy consumption. The node can also reference to other nodes in the time synchronization process. Within this context, the synchronization accuracy and flexibility has been reduced for slow flooding process [42]. Coming to the security issue,
many intruders are there to spoil the systems by introducing a new node in the system. The Open Network Foundation (ONF) ensured to decrease the unwanted node transmission. The Software Defined Networks (SDN) is introduced to have a control over the control messages and incoming packets to classify the nodes [43].

The energy maximization is done by restricting the resources [44]. The smallest distance between the two nodes is a straight line. These straight lines are used for framing an improved protocol in WSN. Without the aid of geographic locations, the straight line routing reduces the energy consumption level. The power constrained WSN acquires two quantities namely, broadcast capacity and information diffusion rate [45]. In Multihop relay, the rate of broadcasting stream is estimated to discover the asymptotic relationship between unified routing structure and MAC schedule. The energy consumption is directly proportional to network performance. A single path can connect a various nodes [46]. The multipath should discover to solve the issue of control plane problem and data plane problem so as to enhance the network performance along with the minimized energy consumption. A key management system was introduced to enhance the information security process. The secret key was generated for edge routers which generate minimal communication [47]. In diffusion-based molecular nanonetworks, the routing functionalities play an important. The gradient based information is obtained between the nodes. The OR function is utilized to serve the best routing protocols [48].

The error propagation approach is used to display the mutual information on the achievable throughput and throughput-delay. Sometimes the transmitting nodes may not cooperate due to interference-limited setting. This delimitation is improved in [49]. The networks may be noisy that leads to high power consumption. To overcome this issue, the data delivery networks should be reliable one. This was solved in [50]. A flexible cross layer design is maintained to enhance the link estimation capabilities and efficient management of neighbor tables that minimizes the power consumption. Another aspect of energy conservation is the energy balance routing protocols. The Forward Factor (FF) supports between the nodes can also establish for the energy balanced mechanism. The selection of the next hop node in accord to the link weight and predecessor nodes information, the FF value is computed [51]. In Tree oriented architecture, a node should be a self organized in case of embedded nodes. The base station assigns a root node and broadcasts the messages to all sensor nodes. Each node randomly chooses its parent and neighbor’s information to make a dynamic protocol [52]. The scalable nodes are formed in the hierarchical structure for efficient use of energy. The cluster head is responsible for network lifetime. When the cluster head fails, the node’s position is not properly organized which leads to overhead. This can be eliminated by selecting the CH based on the node eligibilities [53]. A fuzzy based approach in the CH formation can reduce the risk of node collision issue. Thus the optimization of the routing protocols can also achieve in this [54] [55].

<table>
<thead>
<tr>
<th>Protocols</th>
<th>Merits</th>
<th>Demerits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flooding</td>
<td>Easy to implement</td>
<td>Node collision is high. Resource blindness.</td>
</tr>
<tr>
<td>Gossiping</td>
<td>It avoids node collision issue.</td>
<td>Node intersection is high.</td>
</tr>
<tr>
<td>Directed Diffusion</td>
<td>Data cache prevents loops in data delivery. Saves network bandwidth.</td>
<td>Not suitable in continuous data streams.</td>
</tr>
<tr>
<td>Rumor Routing</td>
<td>A single path is maintained between source and destination. Node failure handling is efficient.</td>
<td>Lack of handling events.</td>
</tr>
<tr>
<td>Energy aware routing</td>
<td>Use sub-optimal paths to increase network lifetime.</td>
<td>Route setup process is complicated.</td>
</tr>
<tr>
<td>Constrained Anisotropic Diffusion Routing (CADR)</td>
<td>Minimize the latency and bandwidth. Dynamically adjusts the data routes.</td>
<td>Node Intersection problem is high.</td>
</tr>
<tr>
<td>COUGAR</td>
<td>Best use of declarative queries. Data abstraction is efficient.</td>
<td>Cluster Head must be maintained in dynamic. Communication overhead is high to sensor nodes. Need of synchronization.</td>
</tr>
<tr>
<td>Active Query forwarding in sensor networks (ACQUIRE)</td>
<td>Query is very comprehensive.</td>
<td>If network size is equal to node size then it behaves like flooding.</td>
</tr>
</tbody>
</table>

**Hierarchical Routing Protocols**

This type of routing protocols is treated as the most energy efficient protocols in WSN because of its higher energy conservation, network versatility and lower data transmission [56] [75] [76].

In Zigbee mesh networks [57], the hierarchical protocols are used to find the shortest path for addressing scheme. This protocol lessens the broadcast overhead and no memory overhead to maintain the routing information. The chain based protocol minimizes the energy consumption level because each node communicates with the next close-by neighbor and then transmits to the base station. The data gathering process is reduced to yield the optimal path [58]. The energy should be dissipated to the BS to improve the network lifetime. The sensor nodes are arranged in the clustering based schemes to maximize the life span of the network [59]. According to the node proximity or node degree, the CH is periodically chosen that achieves fairly uniform cluster formation and low message overhead [60] [61]. Route election and route
distribution is an effective task in the energy efficiency issue. A proper route should be elected for transmitting the nodes and its information [62]. Nodes under heavy traffic might cause energy efficiency problems. This problem exists due to heavy burden to the CH in the data collection scenarios. The CH size should determine effectively to eliminate the higher level of energy consumption. The data sink contains burden of the sensory data in the data collection approach [63] [68]. The hop distance size is estimated for CH in the data sink to dispose the burdened sensor data. The issues such as congestion control, node collision and resource blindness solved in cross layer fashion. The design focused on the initial determination of nodes in the networks. The initial determination such as receiver based contention, initiative based forwarding, local congestion control and distributive cyclic operation to adopt a reliable communication [64]. In [65], the false alarm rate in CH formation can enhance the node scalability. The multiaxis division based approach is introduced to overcome the threshold problem and eliminate the sink dependency. The data aggregation approach handles a variety of queries. The Poisson Arrival Rate is estimated to handle the heavy loads using Time Division Multiple Access (TDMA) [66] [67].

Table 5: Merits and Demerits of Hierarchical Protocols [68]

<table>
<thead>
<tr>
<th>Protocols</th>
<th>Merits</th>
<th>Demerits</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEACH</td>
<td>Load balancing is efficient.</td>
<td>Fit to small regions.</td>
</tr>
<tr>
<td></td>
<td>Collision is prevented.</td>
<td>Energy is not utilized for selection of</td>
</tr>
<tr>
<td></td>
<td></td>
<td>cluster head that leads to overhead.</td>
</tr>
<tr>
<td>Power Efficient Gathering in</td>
<td>Decreased overhead in dynamic cluster</td>
<td>Not suit to time variant topology.</td>
</tr>
<tr>
<td>Sensor Information System</td>
<td>formation.</td>
<td>Data Processing is delayed.</td>
</tr>
<tr>
<td>(PEGASIS)</td>
<td>Data transmission rate is low.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Balanced node formation.</td>
<td></td>
</tr>
<tr>
<td>Hybrid Energy Efficient</td>
<td>Distributed clustering method.</td>
<td>Inappropriate energy consumption.</td>
</tr>
<tr>
<td>Distributed (HEED)</td>
<td>CH is uniform.</td>
<td>Cause overhead in cluster head selection.</td>
</tr>
<tr>
<td></td>
<td>Maximized energy conservation.</td>
<td>Expired soon due to overload.</td>
</tr>
<tr>
<td></td>
<td>Support long range communications</td>
<td></td>
</tr>
<tr>
<td>Threshold sensitive Energy</td>
<td>Energy consumption is reduced by the use of</td>
<td>Periodic reports are not generated.</td>
</tr>
<tr>
<td>Efficient sensor Network</td>
<td>thresholds.</td>
<td></td>
</tr>
<tr>
<td>Protocol (TEEN)</td>
<td>Fit to reactive scenes.</td>
<td></td>
</tr>
<tr>
<td>Adaptive</td>
<td>Applicable to proactive and reactive applications.</td>
<td>Design complexity is large in supporting multi-path.</td>
</tr>
<tr>
<td>Threshold sensitive Energy</td>
<td>Controlled energy consumption</td>
<td></td>
</tr>
<tr>
<td>Efficient sensor Network</td>
<td></td>
<td></td>
</tr>
<tr>
<td>protocol (APTEEN)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MOBILITY BASED PROTOCOLS

Based on the mobility, the cluster head are selected. The node which possesses less mobility is selected as the cluster head which leads to more stable clusters. These are also known energy-unaware routing protocols. There is a chance of packet loss in inter-cluster communication [69] [70]. In data collection process, energy hole problem is a critical issue. The sink mobility is an efficient way to eliminate this issue and extend the lifetime of a network by lessening the communication overhead which is close to the sink [71] [72] [77]. The mobile node should be of high fidelity. A cooperative engagement should be followed in unreliable wireless networks to effectively coordinate the communications. The objects/ events are updated frequently to yield the optimal path [73] [78]. The process of handling multiple mobile sink is also an issue in data aggregation process. The next position of the sink is selected by biased random walk. The rendezvous point acts a threshold value. This significantly reduces the reliability, communication overhead and flexibility [74] [79]. The beaconless routing schemes are well defined in dynamic network topology. Each node transfers the packets without the use of beacons interval and neighbor information. In forward decision process, the routing schemes are robust [80]. By encountering the RTS and CTS message, the packet is unicasted to next hop relay which reduces the energy consumption. The route should be optimized in concept of discover-before-forward. The signaling cost is reduced which directly symbolizes energy consumption [81] [82].

Without scheduling the direction for a mobile sink ahead of time, an information gathering convention utilizing versatile sinks recommends that a versatile sink declare the location data habitually all through the system. On the off chance that sensors can anticipate the portable sink's development, the vitality utilization would be extraordinarily diminished furthermore, information packets handoff would be smoother. In dense oriented networks [83], a convention, called a Scalable Energy-effective Asynchronous Spread (SEAD) is another system for sense the data to versatile sink. The thought is to develop a least Steiner tree for the portable sink and assign some hubs on the tree as access points [84]. The portable sink registers itself with the nearest get to hub. At the point when the sink moves out of scope of the entrance hub, the route is reached out through the incorporation of new get to hub. A stable and high recharging rate power supplies for sensors, and effectively alleviates energy cost on data gathering at the same time, [85] proposed a joint design of energy replenishment and data gathering by exploiting mobility.

Table 6: Merits and Demerits of Mobility based protocols [85]

<table>
<thead>
<tr>
<th>Protocols</th>
<th>Merits</th>
<th>Demerits</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEAD</td>
<td>Sense the data based on the geographic locations.</td>
<td>Overhead is high</td>
</tr>
<tr>
<td></td>
<td>Self organizing protocol.</td>
<td></td>
</tr>
<tr>
<td>Joint Mobility and</td>
<td>Well balanced nodes.</td>
<td>Energy sink-hide problems</td>
</tr>
<tr>
<td>Routing</td>
<td></td>
<td>Consumes more power.</td>
</tr>
<tr>
<td>Data MULES</td>
<td>Low cost infrastructures.</td>
<td>Less fault tolerant systems.</td>
</tr>
</tbody>
</table>
MULTIPATH BASED PROTOCOLS
The data packets are forwarded to the sink node directly from the sensor nodes. The sensor nodes consumes more battery, to eliminate this action ‘multipath routing protocol’ is introduced [74] [75]. In any type of architecture, the node with maximum energy is processed and the nodes with minimized energy yield more time which leads to heavy computational costs. The multipath routing protocols produces less scalability and simplicity.

A hierarchical multipath routing protocol possesses many paths to exploit data forwarded to the sink. The candidate parent nodes are created to the sink node based on the layered network [86]. Some multipath routing eliminates the topology exposure protocols. The idea is to hide the topology and does not carry the packets information. The protocol can likewise set up numerous hub disjoint path in a route invention endeavor and reject the slower action before transmitting parcels [87]. In mobile ad hoc networks, the location prediction concurrently reduces the global multipath route and average of hop size from source to destination is reduced.

Based on the node disjoint paths, the collected location and mobility information are utilized to reduce the control message overhead and energy consumption [88]. Frequent connection disappointments are brought in mobile ad hoc systems because of hub's portability and utilization of untrustworthy remote channels for information transmission. In the route discovery system, the path with optimal connection quality and path expiration time is chosen as the essential routes. On the off chance that there is any route failures amid the information transmission through essential way, the following accessible backup path with high connection quality and path expiration time is chosen [89] [93]. To eliminate the traffic of the nodes, multipath routing is used. The load is balanced among the nodes with the use of node disjoint paths. It is a sink initiated routing protocol. The number of hop count and uplink neighbor’s information, multiple paths between the source and destination is discovered [90]. The overhead problem is not recovered in demand routing protocols [91] [92]. The multipath routing extracts the benefits of AODV protocols. It discovers three node disjoint routes from source to destination which enhances the packet delivery ratio and end to end delay. If any route discovery fails, the backup route helps to transfer the packets without incurring extra energy [94].

In heterogeneous WSNs, a distributed fault tolerant topology algorithm is developed. The super nodes are generated according to k vertex disjoint paths. It reduces the power consumption and efficient super nodes creation [95] [100]. In service oriented architecture, the link disjoint based multipath routing is introduced which eliminates the packet forwarding to the unwanted nodes. So, each node should possess the capability of detecting reliable and related nodes [96]. The node coverage issue is solved by deploying the stationary sensors. A weight barrier graph has been introduced to determine the minimum number of mobile sensor nodes that reduces the effect of overhead [97]. There is a chance of gathering unrelated data from the mobile sensor nodes. In [98], the authors introduced a concept of resiliency networks in a distributed manner. A relative indexing is used to reduce the coding coefficients to cope up with node or link failures.

The actor failure in the network may cause partition tasks into disjoint blocks. If any node fails, it will consult with the neighbor to define the role and actions to the network connectivity [99] [101].

<p>| Table 7: Multipath based protocols [101] |</p>
<table>
<thead>
<tr>
<th>Protocols</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disjoint Path</td>
<td>Primary preference is given to the smallest path. If any failure occurs, it remains stagnant.</td>
</tr>
<tr>
<td>Braided Path</td>
<td>Partially disjoint path. Primary computation path is done. It builds in localized manner.</td>
</tr>
</tbody>
</table>

QOS BASED PROTOCOLS

The Quality of Service (QoS) prerequisites varies in the different applications in the view of delivery ratio and packet loss. Based on the QoS requirements, the network protocol should be designed [74] [75].

A cross layer design was introduced [102] to increase the energy consumption and network bandwidth of IEEE 802. 15. 4. In route discovery process, the location of the mobile sensor is embedded. This information is utilized by the network layers in later that simultaneously reduce the transmission range and power consumption. The Cognitive Radio Networks (CRN) in Federal Communication was also suffering from this issue. The cognitive radio methods such as interleave and underlay was designed to meet the Primary User (PU) QoS requirements [103]. The pair-wise nodes are utilized to collect the geographical information of the nodes in data gathering approach that balance the energy consumption and end to end delay [104] [113]. Actually, the presence of energy holes that exhibit on Primary User (PU) movement is insufficient to empower transfer speed in the inter-communication. It additionally needs to consider the Quality of these gaps [105]. To minimize the interruption to the high-transmission capacity streams, the spectrum detecting process needs to recognize stable unmoving channels, i.e., ones that are relied upon to stay unmoving for a broadened span of time. Different levels of QoS requirements needed for smart grid which symbolizes the packet delay in end to end systems [106]. Without picking a specific program, the system should ensure the packet delay quality in a specified range. This study seems to be quite trickier tasks in wireless NAN without compromising the QoS. In demand routing protocols, the QoS in terms of packet delay, packet error probability and node outage issue are solved.

Tuning the data rate of a sensor is also a QoS requirement. The utility concept, Nash Bargaining Solution (NBS) that operates on co-operative game theory that saves the power consumption and network specific parameters [107]. Sensor hubs are defenseless to different wellsprings of failures. These incorporate malware assaults, software corruption and programming defilement which can decrease hubs' role and seriously influence the most WSN operations. These attacks could prompt discriminating disadvantages, for example, incomplete or complete hub failures that causes dangerous impacts on the fundamental observing applications. Hubs encountering such failures or glitch can be named.
contaminated and will typically neglect to perform normal detecting and communication [108]. Consequently, they are not able to watch the auspicious time delivery service which is critical to look after high QoS in WSN. This is handled by fuzzy based approach in twofold. The network with light load possesses high sensitive and high integrity applications. And these can lead the system from congestion problem which enhances the end to end delay. The concept of potential field can reduce the end to end delay [109] [112]. In cluster tree based data collection, a velocity based data collection scheme is introduced [110]. This reduces the issues of coverage distance, mobility, traffic delay and tree density. It works on collecting the information from the CH and transfer to the sink. Coming to the Resource management and Resource allocation in terms QoS requirements, the maximum and minimum number of channels is utilized. To overcome this issue, the integer programming and convex optimization method is utilized [111]. It reduces the effects of imperfect channel sensing by generating co-tier interference between the nodes.

### Table 8: QoS-based protocols [113]

<table>
<thead>
<tr>
<th>Protocols</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core Extraction Distributed Ad hoc routing (CEDAR)</td>
<td>It’s a medium sized routing scheme. Dynamic Networks Routing computing is simple. Less overhead.</td>
</tr>
</tbody>
</table>

### Table 9: Summary of the Basic characteristics of Data Aggregation protocols

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggregation Method</td>
<td>Tree Based Online Driven by the sink</td>
<td>Tree Based Online Driven by the sink</td>
<td>Chain Based Centralized or Distributed</td>
<td>Completely distributed, asynchronous</td>
<td>Cluster Based online distributed</td>
<td>Cluster Based online distributed synchronous</td>
<td>Grid Based</td>
<td>Cluster Based</td>
</tr>
<tr>
<td>Resilience to Link Failure</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
</tr>
<tr>
<td>Overhead to setup/maintain the Aggregation Structure</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td>Scalability</td>
<td>Low</td>
<td>Medium</td>
<td>Very Low</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Resilience in case of node mobility</td>
<td>Low</td>
<td>Medium</td>
<td>Very Low</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
</tr>
<tr>
<td>Energy saving methods</td>
<td>Sleeping Periods</td>
<td>None</td>
<td>Rotation of the Leader</td>
<td>Rotation of the CH, Sleeping Periods</td>
<td>Local Route repairs</td>
<td>None</td>
<td>Placing high Energy nodes as CH</td>
<td>None</td>
</tr>
<tr>
<td>Timing Strategy</td>
<td>Periodic per Hop adjusted</td>
<td>Asynchronous</td>
<td>Periodic per Hop</td>
<td>Asynchronous</td>
<td>Periodic per Hop</td>
<td>Periodic per Hop</td>
<td>Event Driven</td>
<td>Periodic and event Driven</td>
</tr>
</tbody>
</table>

Although these protocols promise the energy efficiency, further research is required to address issues such as quality of service (QoS) in real-time applications. The applications include real time military applications, event triggering in precision agriculture etc. require energy-aware QoS routing. This will guarantee the bandwidth of connection as well as provides the energy efficient path. Further possible research issue for routing protocols is to consider the node mobility. Most of the protocols we discussed above (except cluster-based comb-needle model) assume all the sensor nodes and sink are stationary. Further research is necessary in this direction to design energy efficient protocols specific to a particular application. For example, in battle field it may be necessary for the sink to move for detecting the events such as identification of enemy tank, exhaustion of ammunition at a soldier, etc. The mobility of a node requires more number of updates in terms of its position, and propagation of that information to other nodes may consume more energy. The battery of the nodes may drain out fast. To handle this situation and overhead of the mobility, new routing protocols need to be designed in an energy constrained environment.

### CONCLUSION

Data aggregation in sensor networks has attracted a lot of attention in the recent time. In this paper, we summarized recent research results on data aggregation and routing in WSN. We surveyed routing protocols by taking into account several classification criteria, including location information, network layering and in-network processing, data centricity, Mobility-based, Multipath-based Protocols, network heterogeneity, and QoS requirements. The Table 9 shows the summary of the Basic characteristics of Data Aggregation protocols. If we consider the parameters Aggregation Method, Resilience to Link Failure, Setup Oversead, Scalability, Resilience in case of node mobility, Energy saving method, and Timing strategy, the Cluster based data aggregation protocols perform well compared with other protocols and we can build energy efficient WSN with these protocols.
REFERENCES


Authors detail

Dr. O. B. V. Ramanaiah, Professor, CSE Department, JNTU College of Engineering, Hyderabad, India He received Ph.D. in Computer Science from University of Hyderabad in 2005, M.Tech(CS) from JNTU, India, and B. Tech(CSE) from SVUCE, Tirupathi, India. His Total Teaching Experience is 22 Years with Total Research Experience of 13 Years. He has published research papers in national & International Journals and Conferences. He visited USA for paper presentation in an IEEE Conference, ITCC 04, held during April 5-7, 2004. Currently he is providing Research Guidance to 8 students and organized many Refresher Courses. His research interests include Computer Networks Distributed Systems, Mobile Computing, Ad hoc & Sensor Networks.

Ms. M. Shanmukhi is currently pursuing her Ph.D degree in CSE at JNTU, Hyderabad, India. Received M.Tech (CS), and B.Tech (CSE) from JNTU, Hyderabad, India. She has more than 11 years of teaching experience, currently working as an Associate professor in the department of CSE, Malla Reddy college of Engineering, Affiliated to JNTU, Hyderabad, India. She has published research papers in national & International Conferences. Her main research interests include wireless Adhoc & Sensor networks, Multimedia communications, TCP/IP Networks and Protocols.