TEKNISKA HÖGSKOLAN
HÖGSKOLAN I JÖNKÖPING

Data Aggregation and Gathering Transmission in Wireless Sensor Networks: A Survey

PHANI PRIYA KAKANI

THESIS WORK2011-2013

SUBJECT

Master of Electrical Engineering:
Specialization inEmbedded Systems

Postadress:
Box 1026
551 11 Jönköping

Besöksadress:
Gjuterigatan 5

Telefon:
036-10 10 00 (vx)

TEKNISKA HÖGSKOLAN

HÖGSKOLAN I JÖNKÖPING

This exam work has been carried out at the School of Engineeringin Jönköping in the subject area electronics. The work is a part of the two-year Master of Science programme.

The authors take full responsibility for opinions, conclusions and findings presented.

Examiner:Dr.Youzhixu

Supervisor: Dr.YouzhiXu

Scope: 30 credits (D-level)

Date:

Abstract

Wireless sensor networks have many sensor devices that send their data to the sink or base station for further processing. This is called direct delivery. But this leads to heavy traffic in the network and as the nodes are limited with energy, this decreases the lifetime of the network. So data aggregation technique is introduced to improve the lifetime. This technique aggregates or merges the multiple incoming packets in to single packet and forwards it to sink. There is different data aggregation techniques based on the topology of the network.

This report clearly explains the purpose of data aggregation and gathering in WSN, data aggregation in flat networks and data aggregation in hierarchical networks, different data aggregation techniques in cluster based networks, chain based, tree based and grid based networks.

Data aggregation technique can successfully minimize the data traffic and energy consumption only when it is carried out in a secure manner. Part2 of the survey explains the possible attacks that affect data aggregation in wireless sensor network. The secure data aggregation techniques in wireless sensor networks are also discussed in this report.

Summary

Wireless sensor network includes many sensor devises which work on a common objective. It also includes one or more base stations (sinks) that gather data from all the nodes. There are different techniques to transfer data between the nodes and to the sink. These sensor devices are limited with battery power and limited storage capacity. So, we need to increase the lifetime of the network by implementing techniques like in-network data aggregation techniques.

Data aggregation and gathering technique decreases the data traffic and further saves energy by merging multiple incoming packets into a single packet and then forwarding it to sink.

There are different data aggregations techniques based on the network topology which are discussed in detailed and compared in this report.

It's very essential to provide security to the network while implementing the data aggregation process so that one can receive the original information from data owner within a short span of time. There are many techniques that provide security to the data aggregation process in the network and they are discussed in this report.

The data aggregation algorithms discussed in this report mainly focuses on three concepts which are efficient routing, organization and data aggregation tree construction. This report described the main features, benefits and limitations of different data aggregation algorithms.

Keywords

Data aggregation in hierarchical networks,

Secure data aggregation,

Purpose of data aggregation and gathering in wireless sensor networks,

Impact of timing in data aggregation.

Contents

1. Introduction

Wireless Sensor Network (WSN) is a fast growing and stirring research area which has attracted considerable research attention. These networks are widely used in various applications such as industrial, medical, military and home networks. Simply these networks are a new class of distributed systems which is an essential part of physical space they inhabit [52]. Generally the wireless sensor devices are considered as the battery-operated device and have an ability of sensing physical quantities. Apart from sensing, this network is capable of data storage, wireless communication and partial amount of computation and signal processing. WSN mainly includes a huge number of wireless-capable sensor devices that works collaboratively in order to attain a common objective. This network will either include one or more base stations (or sinks) that gathers information from all sensor devices. Moreover, these sinks are the interface by which WSN interacts with outside world [73]. Within this network there are numerous techniques to transfer the data from one sensor node to another. In-network aggregation is one of the important techniques in WSN, which process the data at intermediate nodes and routing the information through the network. The present study will clearly give an in-depth knowledge on WSN by conducting survey on numerous papers.

1.1 Background of Technology

Wireless sensor networks comprise small computing devices where they have the capability of producing digital representation of real-world phenomena. All these devices will have inadequate energy resources, limited storage capacity and limited network bandwidth. The information that is being produced by nodes in network propagates via network through wireless links. Transmission through wireless is expensive when compare to local processing of information. A research conducted by University of California has estimated that sending a single bit over radio is at least three orders of magnitude more expensive than executing single instruction [47]. Inadequate amount of bandwidth, storage capacity and energy available to sensor nodes calls for particular optimizations of queries injected into network.

At a host node, the query requesting aggregate data is inserted into sensor network. This node is also referred as sink. In the network, host forwards the query to other

nodes. The least optimal query plan would need each node to report its own readings back to host node for processing. Once all the data packets are received by host node from source node, then it would aggregate all of the information into final value and report the value back to user. This process is called as direct delivery; however, this approach has numerous drawbacks. One of the main issue is that large number of packets need to be sent to host node as each of the node send its own information to host and at least there must be a single packet sent to a node. Moreover, as few of the nodes fails to directly communicate with host so these data packets will be forwarded by other nodes until it reaches the host. The other issue is that each of the packet size is relatively small as it only includes readings from one sensor. Due to this reason, the lifetime of a network is decreased.

To preserve both bandwidth and energy, it is useful to move filtering and integration of sensor information into network itself. In-network aggregation is a method for diminishing overall amount of power and bandwidth needs to process the users query by allowing sensor readings to be aggregated by intermediate nodes. In general WSN with 100 nodes, minimum of 100 packets will be forwarded to sink, which conserves energy and decreases the lifetime of the network. So, to reduce the energy consumption, data aggregation or data merging should be implemented [54]. Many of the researchers proposed various algorithms for in-network aggregation to avoid the problems.

1.2 Overview of Wireless Sensor Networks

Now-a-days, WSN is considered as one of the emerging technology since it greatly helps people by offering sensing, computing and communication capabilities and enable humans to have a closely interaction with the environment wherever they go [75]. This network is simply defined as the collection of nodes which are organized into a cooperative network. Here, each of its nodes include processing capability, multiples types of memory, RF transceiver, power source and further it even accommodate a variety of actuators and sensors. The communication among all these nodes will be carried out wirelessly and is often self-organize once it is deployed in an ad-hoc fashion [31]. As per [28], the main concept of WSN is based on simple equation that is

Sensing + CPU + Radio => Thousands of potential

Sensor Network Communication Architecture

According to [49], basically the sensor network comprises sensor field, where the sensor devices or nodes are scattered in this field. Here, each of these nodes will have the capability to gather information and then route information back to sink and end users. With the help of multi-hop infrastructure and less architecture the information is routed back to the final user through sink as shown in figure 1.

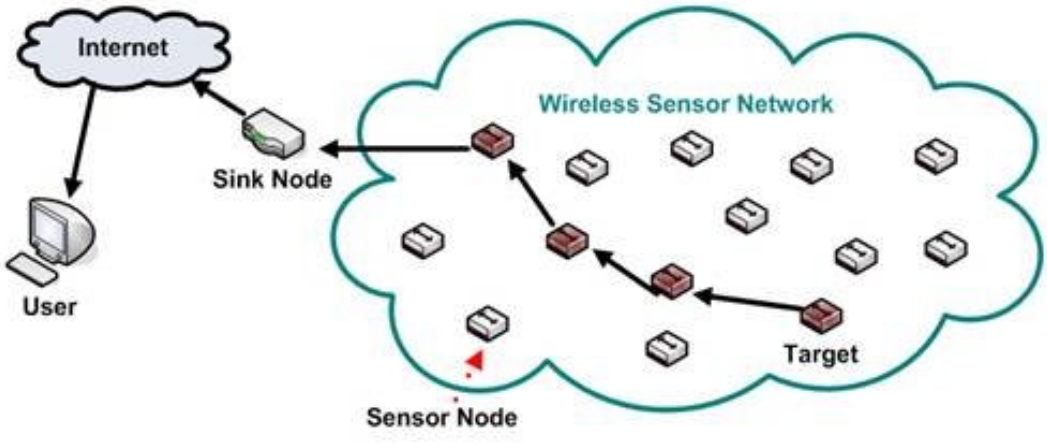

Figure 1: Wireless Sensor Network [48].

Here in this network the sink send commands or queries to other sensor node in sensing region, on other hand sensor node collaborate to achieve the sensing task and send sensed information to sink. In the meantime, sink even serve as gateway to outside networks. Further sink gathers information from sensor nodes, and executes simple processing on collected information and then sends relevant data to end user through internet whoever made request to make use of the information. Each of the sensor nodes makes use of single-hop long-distance transmission to send information to sink [48]. The sink communicates with the user either through internet or satellite. Both sink and nodes make use of protocol stack where it merges power and routing awareness, combines information with networking protocols, communicates power efficiently via wireless medium and promotes cooperative efforts of sensor nodes. The stack of the protocol includes application layer, transport layer, network layer, data link layer, physical layer, power management plane, mobility management plane, and task management plane [1].

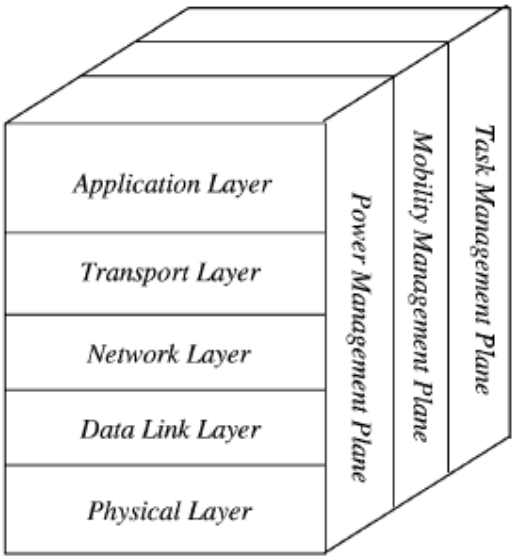

Figure 2: Sensor networks protocol stack [1] p 405.

Yet, this method is costly in terms of energy consumption for long-distance transmission [33]. Hence, from the above context it can be stated that sensor networks consists of large number of small nodes with computation, sensing and wireless communication capabilities. Apart from these the network even produces high-quality data because of its coordination of sensor nodes.

Applications

In the real time environment, WSN is being used in various fields where some of them are discussed below [32]

Industry applications –based on the industry specifications the WSN was separately designed and implemented. With this network, numerous applications can be recognized in this framework. Moreover this network has an ability to monitor the quality of air and even the temperature of building. Apart from all these, it even manages the complex machinery set, produced goods, and conditions of production system of particular factory or group of factories.

Medical applications–wireless sensor networks are used to form BAN (Body Area Network) where it includes numerous sensors which are located closely to human body for measuring signals like breathe rate or heartbeat.

Military applications – WSN was widely used in military applications as it is very hard to install a communication infrastructure in theatre of operation.

1.3 Importance of WSN

In recent years, due to advances in hardware technology the demand for the WSN is increasing rapidly. These networks are changing the way the users acquire the data from physical environment. For instance, wired LAN (Local Area Networks) utilizes network adapters and Ethernet cables for transferring the data from one system to another. However, users are facing problems in transferring the data through wired networks. Some of the major limitations of wired networks include running wires from each room is a difficult task, adding more computers to wired network result to unexpected expense, etc. [3]. Because of these reasons, most of the usersprefer WSN as these wireless networks operate radio waves or microwaves to sustain interaction channels linking computers. One of the main benefits of WSN is that it provides better flexibility to the users and enables them to connect the network easily around the house without any problem.

1.4 Industrial Wireless Sensor Network Standards

1.4.1 Wireless-HART

WirelessHART is introduced in the year 2007 as the first open wireless standard for process control industry. Right from its establishment the demand of this wireless network has been increased rapidly and drawn the attention of industry. "*The WirelessHARTtechnology was designed to enable secure industrial wireless sensor network communications, while ensuring ease-of-use is not compromised*" [21] p.3. At the end of year 2008, the products of WirelessHART have started emerging in the market.

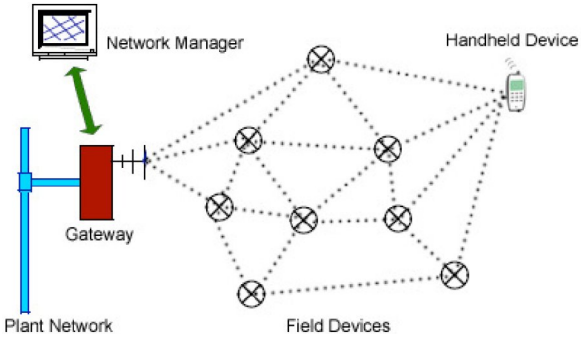

Figure 3: Typical WirelessHART Network [70] p.3.

Network manager acts as the control centre for entire network. Here, manager is responsible for managing all the resources and scheduling communications for all the devices in network. Network manager is a part of database (such as software) but not a physical device. The four types of devices included in WirelessHART are described below [70]

Plant network – this network has a connection with gateway where it indirectly gathers and sends data to sensor nodes.

Gateway –it is similar to access point where it allows communications among host applications in plant network (along with network manager) and field devices. Here the network manager needs to have secure connection with gateway.

Field devices–these devices are sensors fixed all over the plant. These are both consumer and producer of messages and are in charge for gathering the required data and transmit back to plant network. Here each of devices has an ability to route packets to other devices in network.

Handheld device–it is a special device where it is mainly utilized for configuring and monitoring field devices. This device is used to enable access of the plant to the user. Further, it is even used by wirelessHART network to recognize the employees and their position in plant.

Each of the WirelessHART network is recognized with the help of a unique network ID. In case, if the packets have a dissimilar network ID then it cannot be decoded and further it will be discarded.

Wireless HART is the initial wireless sensor network standard to emerge where it is particularly designed to support the particular needs posed by industry. One of the main requirements of automation industry is that WSN should have a very long life time nearly 5 to 10 years without change in battery [46]. Moreover, implementing in-network aggregation of data in wireless HART networks will save energy and thereby prolong the lifetime of network.

WirelessHART does not allow security to be optional, which prevents mistakes that can compromise the system. That is why this standard doesn't allow data aggregation.

1.4.2 ISA-100.11a

ISA100.11a is an industrial standard proposed by International Society of Automation (ISA). It is a mesh networking standard with frequency hopping which takes gain of

IEEE 802.15.4 radio, and offers high security and reliable wireless communication and interfaces with numerous existing industrial standards. It covers different applications such as process applications, asset tracking and identification and so on. ISA 100.11a is superior to Wireless HART since ISA 100.11a was mainly designed from the ground up for control whereas; on the other hand wireless HART was never designed for control and it has been always a configuration protocol [80]. WirelessHART specifies that all communication must be encrypted Whereas, ISA100.11a allows for communications to not be encrypted. Users concerned about security will need to ensure ISA100.11a devices can support secure communications and must be careful to ensure that security is maintained.

Radio Technology	ISA 100.11a	WirelessHART
Operating frequency band	2400-2483.5 MHz	2400-2483.5 MHz
MAC-DLL mechanism	CSMA/CA, Channel hopping, blacklisting, super frame optimization	Channel hopping, blacklisting, TDMA
Range	Indoors: 30m Outdoors: 90m	Indoors: 30m Outdoors: 90m
Channel bandwidth	2 MHz	2 MHz
Supported number of devices	Hundreds	Hundred
Maximum data rate	250kbps	250kbps
modulation	O-QPSK, DSSS	O-QPSK, DSSS
Network architecture	Star, Mesh	Star, Mesh

Table: Shows the performance parameter comparison of ISA 100.11a and Wireless HART [81].

1.4.3 WIA- PA

Wireless Networks for Industrial Automation-Process Automation -

It is a new Chinese industrial wireless communication standard for process automation. It is mainly designed to fulfil the requirements of process automation applications as an IEC 62601 standard with features of HART compatibility, data aggregation and so on. It is a hybrid star-mesh network and allows aggregation.

This protocol consists of 7 logical roles, which are gateway, redundant gateway, network manager (NM), security manager (SM), cluster head, redundant cluster head

and cluster member. In order to guarantee real time and reliable communication, resource reservation, priority-based CSMA/CA and queuing mechanisms are put forward for automatic tasks to access the MAC layer channel and to transmit network layer packets. Congestion control is used to prevent the network from collapse in heavy traffics [82].

1.5 WSN Drawbacks

Apart from its various benefits, even WSN has some limitations. One of the most limiting factorsis the security as these sensor networks are less secure as hacker's PC (Personal Computer) or laptop can act as AP (Access Point). If the user gets connected to hacker laptop then these people can easily read all the users data. The second drawback is that WSN has less speed when compare to wired network. Apart from all these limitations, even these sensor networks affected by surrounding like walls, far distance (attenuation), microwave oven (interference), etc. [36]. However, most of the people are making use of this technology by taking some measurements in securing their networks as WSN offers numerous benefits to the users.

1.6 Thesis Outline

The thesis is organized as follows**:**

Chapter 2: Theoretical Background

This section deals with the review of comprehensive literature and this chapter comprises a review on purpose of data aggregation in WSN, data aggregation in WSN, comparison of WSN with and without data aggregation, data gathering transmission and summary to this section.

Chapter 3: Survey and Discussion

This chapter will make a survey on numerous papers related to research topic of various authors and just contrast their ideas and the approach they followed to reach their objectives.

Chapter 4: Survey Part2

This chapter will make a survey on different papers related to the secure data aggregation and their approach to reach the objective.

Chapter 5: Conclusion

This chapter concludes the study.

2. Theoretical background

2.1. Overview

This section enables the readers to acquire an in-depth theoretical knowledge on the topics that are related to research. Further this chapter is sub classified into four sections where the first section gives a brief idea on purpose of considering data aggregation in WSN.The second section discusses about data aggregation in WSN and further explains how the process of data aggregation carries out in WSN. Third section explains comparison of WSN with and without data aggregation and fourth section clearly illustrates data gathering transmission. Finally it summarizes the chapter.

2.2. Purpose of Data Aggregation in WSN

In current scenariosthe demand for WSN had rapidly increased in various applications like weather monitoring, petroleum and military due to low power, small size, light weight, and wireless sensors. However these inexpensive sensors are equippedwith limited battery power and thus constrained in energy [10]. One of the major issueswith WSN is that one need to increase the lifetime of network. Generally, lifetime of network is defined as the time whenever the first node fails to send its information to base station. This issue can be resolved by implementing data aggregation technique as it decreases data traffic and further saves energy by merging multiple incoming packets into a single packetwhenever the sensed information are highly correlated [66].Numerous researches have been carried out to further extent network lifetime.

2.3. Data Aggregation in Wireless Sensor Networks

It is just a process of aggregating the sensor information through aggregation approaches. This technique was mainly utilized to resolve both overlap and implosion problem in data centric routing [22]. In data aggregation the sensor network is generally supposed as reverse multicast tree. Here, sink request the sensor nodes to report ambient condition of phenomena. In this process, generally the information that

is coming from several sensor nodes are aggregated in such a way that they are about same attribute of phenomenon once it reach the same routing node on way back to sink.

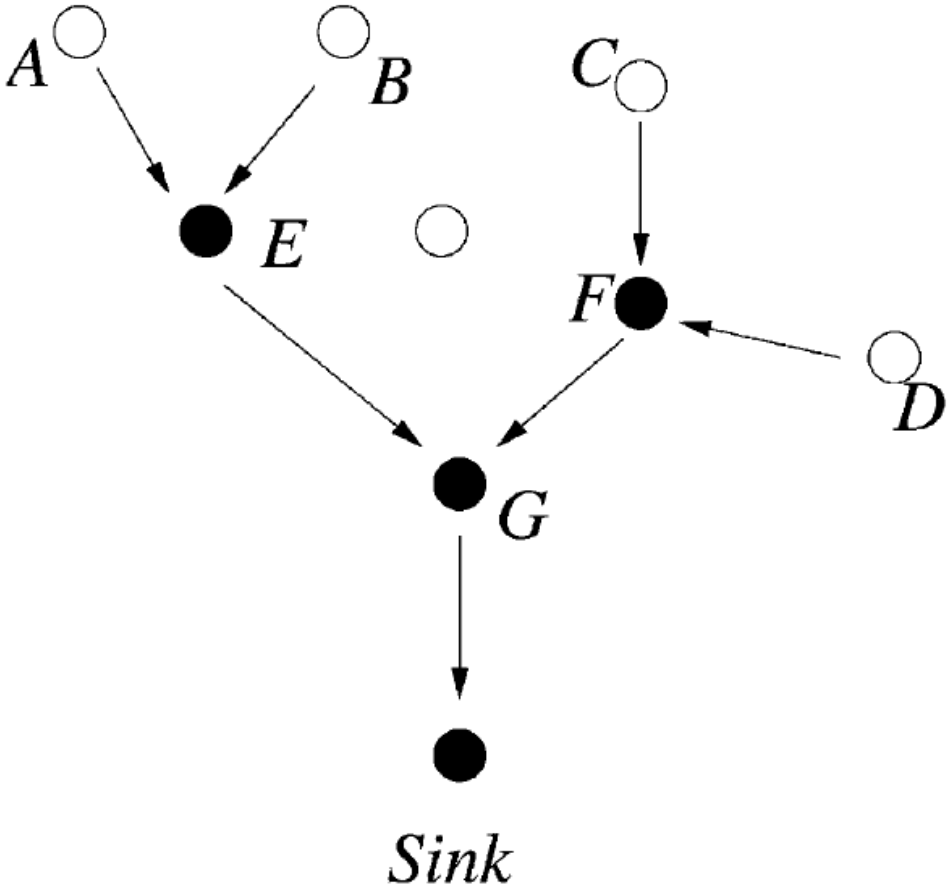

Figure 4: An example of data aggregation [1] p.409.

In order to get a clear idea on data aggregation let us consider seven sensor nodes 'A', 'B', 'C', 'D', 'E', 'F', 'G' and a sink node. Initially, sensor node 'E' aggregates the information from both 'A' and 'B' sensor nodes. In the same way node 'F' aggregates data from 'C' and 'D' sensor nodes. According to [23], data aggregation is professed as the set of automated methods of merging the information which comes from several sensor nodes into set of meaningful data. In other way this process is also called as data fusion. Based on this point, 'G' node gathers information from 'E' and 'F' and finally sends the data to sink node. Here, data aggregation takes place through some aggregation algorithms like TAG (Tiny Aggregation), LEACH (Low Energy Adaptive Clustering Hierarchy), etc.

2.4. Data Aggregation Techniques

Generally the data aggregation can be classified on basis of network topology, quality of services, network basis and many more. For the current research, the techniques are being discussed based on the network topology. In this topology, the data aggregation technique is categorized into two parts which are flat and hierarchical network. Further hierarchical network is sub-classified into four parts which are cluster based, chain based, tree based and grid based.

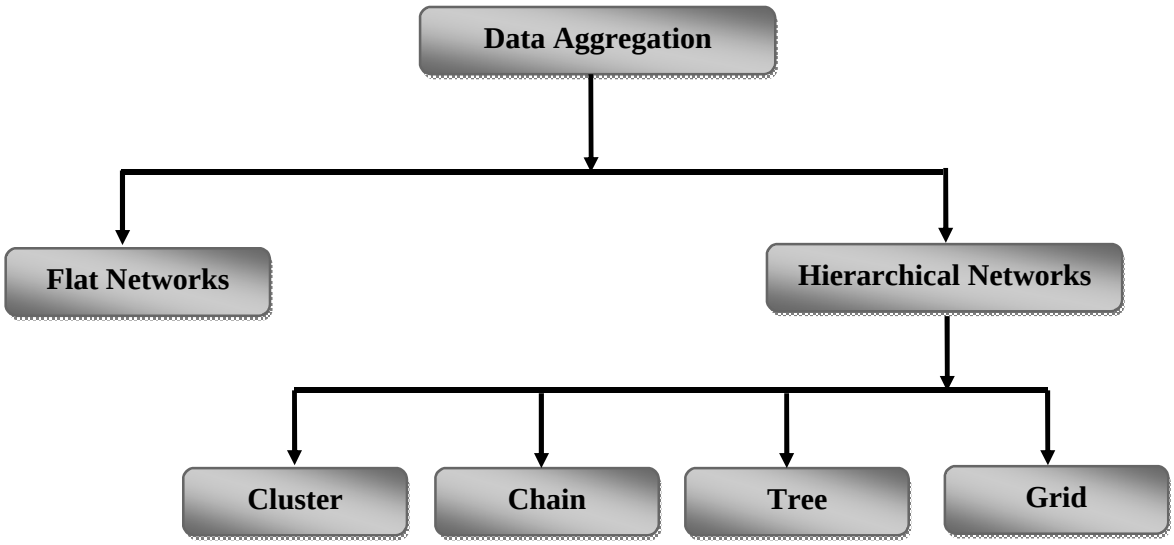

Figure 5: Shows taxonomy of Data Aggregation [64] p.37.

2.4.1. Data Aggregation in Flat Networks

In these networks, all of the sensor nodes play the same role and further these nodes are equipped with same battery power. In this type of network data aggregation is achieved through data centric routing where the sink transmits a query message to the sensors through flooding and in case if the sensors have data matching with the query then that particular sensor send response message back to sink [65]. Due to excessive computation and communications in sink node it results faster depletion of its battery power. Thus the failure of sink node results in breaking the functionality of network. Due to this reason various hierarchical data aggregation techniques have been introduced for scalability and energy efficiency [64].

2.5. Data Aggregation in Hierarchical Networks

In hierarchical networks, data aggregation engages with the data fusion at special nodes where it diminishes the number of messages that need to be transmitted to sink node. Hence, this process improves the energy efficiency of network [65]. Further of this section discusses the different hierarchical data aggregation techniques and protocols.

2.5.1. Cluster based Data Aggregations Technique

In energy constrained sensor networks of large size, it is very difficult for the sensors to transfer data directly to the sink. In these situations the sensors can transmit the information to local aggregator or cluster and then to sink. So in this network, the sensor nodes are organized in the form of clusters. Here sensors transmit the information to cluster head and there by this cluster head aggregates all the data received from sensors and then transmits the concise data to sink. The cluster head can communicate with the sink either through long range transmissions or multi-hoping via other cluster heads [53]. Thus this process results in saving the energy and mainly useful for energy-constrained sensors.

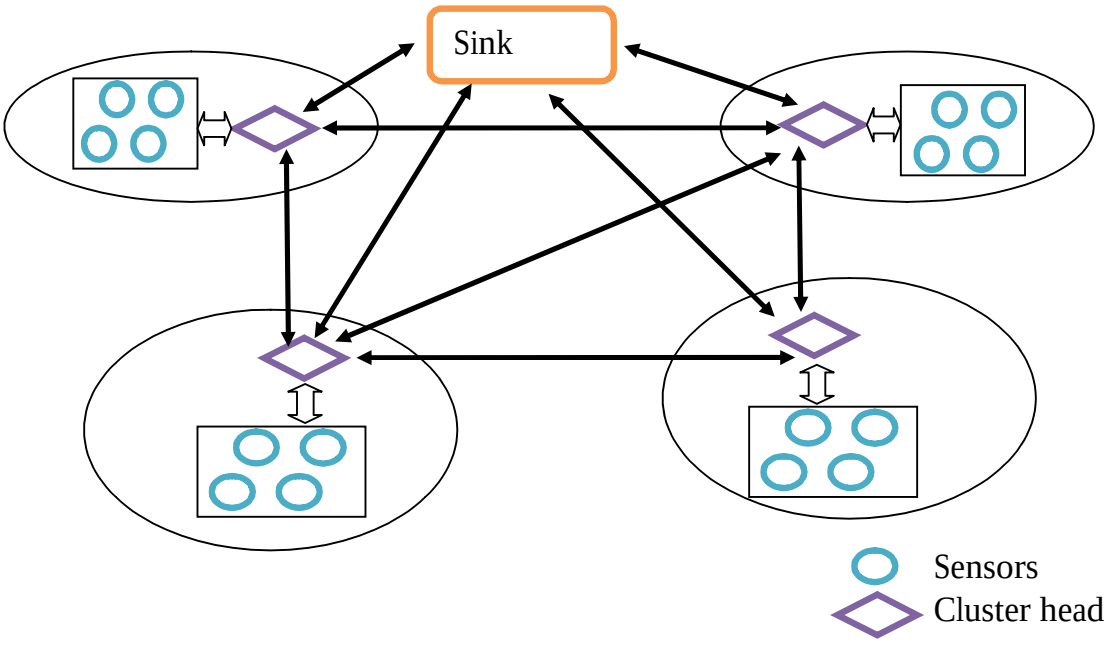

Figure 6: Shows the Cluster based data aggregation [53] p.6.

Various cluster based protocols have been proposed by numerous researchers such as LEACH (Low-Energy Adaptive Clustering Hierarchy), EECDA (Energy Efficient Clustering and Data Aggregation) and CAG (Clustered Aggregation). These protocols are clearly are discussed in section 3.

2.5.2. Chain based Data Aggregations Techniques

In the previous section, it has been discussed that cluster members send the information to cluster head and further it transfer the aggregated information to sink. In case if the distance between cluster head and sink is far then it consumes more energy to communicate the sink whereas in the case of Chain based data aggregation the information is sent only to its closest neighbor.

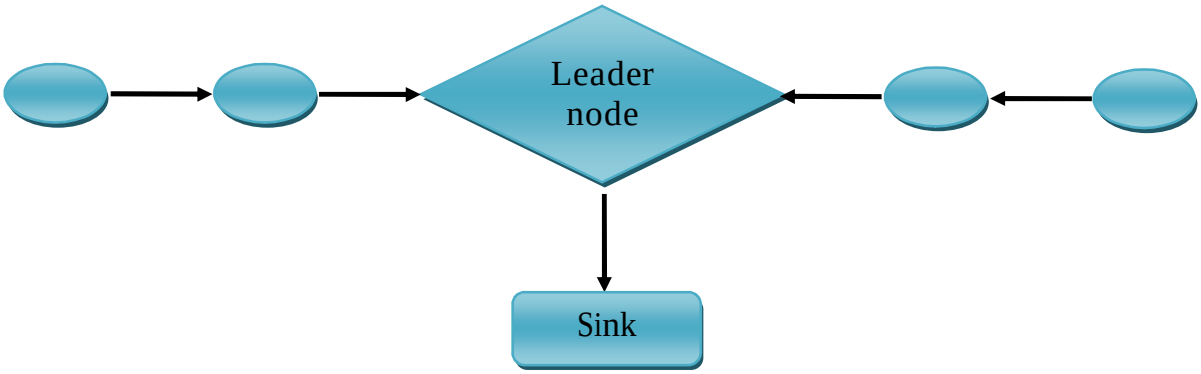

Figure 7: Shows the Chain based organization in a sensor network [53] p.9.

In this process, the nodes are constructed in the form of chains for data transmission to cluster head. Each and every node in the network sends the sensed data only to its neighbor node rather than cluster-head and each of the node aggregates the information to diminish the amount of data transferred [64]. Some of the protocols that are proposed for this network include PEGASIS (Power-Efficient Gathering in Sensor Information System), CHIRON (Chain-Based Hierarchical Routing Protocol), COSEN (Chain Oriented Sensor Network for Efficient Data Collection), and Enhanced PEGASIS which are discussed in section 3.

2.5.3. Tree based Data Aggregations Technique

In tree based network, the nodes are organized in the form of tree topology where sink is considered as a root. Aggregation is carried out by constructing aggregation tree where source nodes are referred as leaves, and rooted at sink. Here data flow starts from leaves nodes and end at sink. So here all the intermediate nodes carry out the aggregation process and finally transfer to root (that is, sink). The main aim of the tree based approach is constructing an energy efficient tree [65].

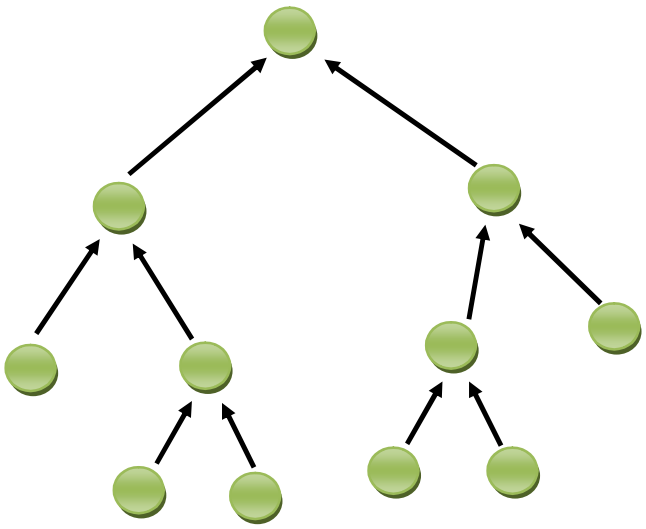

Figure 8: Shows the Tree based data aggregation [64] p.40.

Some of the protocols that are proposed for tree based are TREEPSI (Tree-based Efficient Protocol for Sensor Information), TCDGP (Tree-Clustered Data Gathering Protocol) and PERLA (Power Efficient Routing with Limited Latency) which are discussed in section 3.

2.5.4. Grid based Data Aggregations Technique

In this technique, a fixed data aggregator is placed in each of the grid and further it aggregates the information from all sensors. Thus, within a grid the sensors do not communicate with other sensors. Within a grid any of the sensor nodes can be assumed role of data aggregator in terms of rounds unless and until the last node dies. This technique is mostly useful for weather forecasting and military surveillance [64].

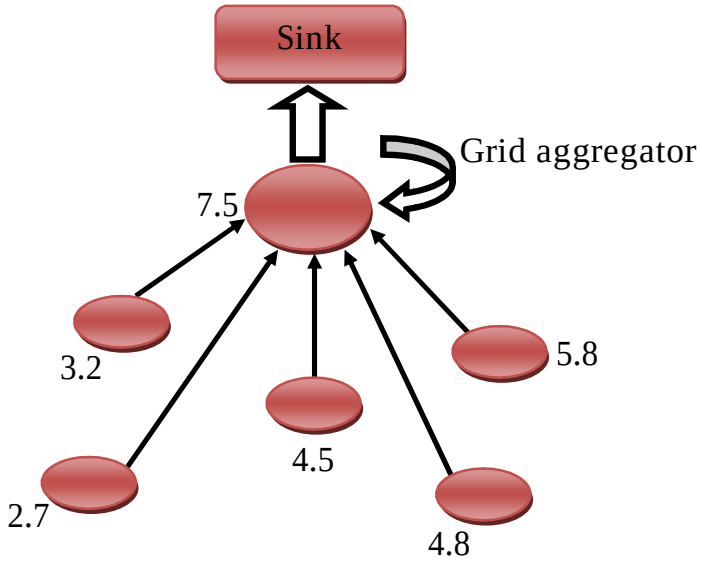

Figure 9: Shows Grid based data aggregation. Here the arrow represents the data transmission from sensors to grid aggregator [53] p.13.

Protocols that are being introduced for grid based include ATCBG (Aggregation Tree Construction Based on Grid) and GROUP which are discussed in section 3.

2.6. Impact of Timing in Data Aggregation

After reviewing many of the secondary resources it has been proved that timing has more effect in data aggregation for sensor networks. In most of the cases, the timing has significant impact on both freshness and accuracy of information delivered by aggregation algorithms. Basically, this model defines only when to clock out the data since it is aggregated by nodes on its way to information sink. In recent energy-accuracy tradeoff study carried out by Boulis, A. et al. [5] had identified the importance of timing models for efficient data aggregation. Further the author introduced a data collection mechanism where nodes has an ability to make a decision on whether to divide up their own readings depending on the estimates that they receive from other nodes. However, this proposal is only helpful for operations such as reporting the minimum or maximum value and further it doesn't suit for more general network monitoring applications. [26] Researcher has investigated in-network aggregation as a power-efficient mechanism for propagating information in wireless sensor networks. In their study, they evaluated the performance of various in-network aggregation algorithms along with cascading timeouts and further characterize the

tradeoffs among energy efficiency, freshness and data accuracy. Thus their result proved that timing (time taken for a node to receive data from its children) before forwarding information onto next hop will play an important role in performance of aggregation algorithms in context of periodic data generation.

2.7. Data Aggregation with Time Synchronization

As per [27], "*Time synchronization is the ability to maintain the same time across servers and the computers that are attached to them*" [27 p 48]. The time synchronization phase is started up by a root by broadcasting time synchronization command (that is SYNC command).

hop	nodeId	sendTimeStamp	delay

Figure 10: Format of SYNC command [12] p.1318.

Here, hop field signify the hop count from the node that sends the command to root. The field nodeID records ID of node that sends the command. The next field is sendTimeStamp where this field records the local time whenever certain byte of command is sent. The final field is delay where it represents the accumulated delay at the time of command transfer from root to current node and it is further initialized to zero by root. Initially, sender sets the sendTimeStamp field at the time when particular byte is actually transmitted at MAC layer. The adjacent node of root records the local time whenever it receives corresponding byte of SYNC command as receiveTimeStamp and further synchronizes itself to root by means of adjusting its local time based on the equation (1) [12].

$$T_{revised} = T_{current} + (SendTimeStamp - receiveTimeStamp) \quad \ldots\ldots\ldots\ldots (1)$$

Along with sensor fusion, even the other coordinated actions have the need for proper time synchronization. Most of the applications in WSN often require higher clock accuracy, so time synchronization in this network has to cope with numerous additional challenges.

Apart from these reasons, in a distributed system there is no common clock. Each node in the network has its own local clock. In order to determine the time of event happened (with respect to time) or the time difference between the events happened in the network, the clocks have to be synchronized. It is difficult to synchronize the clocks because clocks tend to drift over time. So one need to synchronize frequently

to avoid local clock drift, which is a drawback. Also, this is not sufficiently accurate to determine the relative order of events in a distributed system. Hence, because of all these reasons time synchronization is not preferred by many of the researchers.

2.8. Comparison of WSN with and without Data Aggregation

In wireless sensor networks the data communication among the nodes consumes a large portion of total energy. Due to this reason, data aggregation techniques are being introduced, where it greatly helps in reducing the consumption of energy by eliminating the redundant information that is being travelled back to Base Station (BS) [19].

Now let us see the difference between the WSN with and without data aggregation.

No.	**WSN with Data Aggregation**		**WSN without Data Aggregation**	
	Pros	**Cons**	**Pros**	**Cons**
1.	Eliminates redundant data and conserve energy of sensors	Not applicable in all sensing environments. For instance, if the data transmitted by all the nodes are different then data aggregation cannot be carried out.	Can be applied in all the sensing environments.	The number and size of data transmission is more.
2.	Reduces the size and number of data transmissions	Aggregator or cluster head may be attacked by malicious attacker	As there is no in-network aggregation the data routing problem occurred in data transmission can be referred as maximum flow problem and can be easily resolved with integer program with linear constraints.	Sensors consumes more energy while transferring data from one node to another
3.	With this process sensor has an ability to aggregate multiple	In some cases, nodes consume more power whenever the	-	Sensor node may collect same data from different nodes

	incoming packets into single packet	aggregate result is sent to sink through uncompromised nodes.		

Table: Shows the pros and cons of WSN with and without data aggregation [37] and [35].

2.9. Summary

WSN is being widely used by many of the users and the demand of this technology is increasing day by day due to its various functionalities. This network includes numerous sensor nodes where it transmits the information from source to sink by passing through various intermediate sensor nodes. WSN is being widely used in various applications and mainly in industrial applications. Because of this reason the concept of data aggregation is being introduced to transmit the information to its destination without transforming any redundant data.

3. Survey1: In-network Data Aggregation

3.1. Overview

This section gives a clear idea on the survey related to the research by reviewing numerous papers and then compares various thoughts in the area of data aggregation in WSN. This task has been completed by gathering the accurate information for the research. Basically, there are two ways for gathering data to the research which are primary data type and secondary data type. Primary data type enables the researcher to collect the information from various respondents through some methods like questionnaire, survey, interview, etc. On other hand, secondary data type allows the researcher to gather information by reviewing numerous secondary resources like journal, authorized pdf's, online books, etc. For the present research, secondary data type has been selected for acquiring an in-depth knowledge on the research topic. Here, with this method we gathered around 30 papers on research related work and then clearly discuss and contrast their proposed methods.

3.2. Survey and Discussion on WSN

To acquire an in-depth knowledge on WSN in terms of data aggregation, a survey has been conducted on 30 research papers. The aim of all these research papers differs from one paper to another so this survey section has been sub-categorized into two parts, which are listed below:

- Data Aggregation in Hierarchical Networks
- Secure Data aggregation

Part 1: Study on Data Aggregation in Hierarchical Networks

As discussed in section 2.5. There are four different hierarchical networks which are

- Cluster
- Chain
- Tree and
- Grid

Here in each of these data aggregation techniques various scholars introduced different protocols for reaching their objectives.

3.3 Cluster based Data Aggregation Technique

In this data aggregation technique, three protocols namely LEACH (Low-Energy Adaptive Clustering Hierarchy), EECDA (Energy Efficient Clustering and Data Aggregation) and CAG (Clustered Aggregation) had been introduced to provide better features for the users in terms of energy efficiency, better life time, stability, etc.

3.3.1. LEACH

Approach

LEACH is a self-organising and the first dynamic clustering protocol for WSN that makes use of randomized rotation of Cluster Head (CH) to share out the energy load evenly between the sensors in the network. Further it even incorporates the data fusion into routing protocol to diminish the amount of data which should be transmitted to BS. [23] This protocol is mostly suitable for the applications which have periodic data reporting and constant monitoring. It runs in many rounds and each of the rounds includes two phases which are *cluster setup phase* and *steady phase*. In the first phase that is cluster setup, LEACH carry out the cluster organization and further chooses CH.The CH which are being selected will further broadcast the message to all other sensor nodes in the network using CSMA MAC protocol and informs that it is the new CH. Except the CH nodes, the remaining nodes receives the message from CH and decides to which cluster it belongs based on the signal strength of the received message. CH node is very energy intensive than being non CH node. Hence, LEACH protocol includes the randomized rotation of high-energy CH position between the sensors. In the steady phase, cluster creates a Time Division Multiple Access (TDMA) schedule telling each node when it can transmit. The data collection is centralized and is periodically executed by making use of the schedule created by every CH. Now all these nodes transmit the information to CH and CH performs data aggregation and transmits the aggregated data to remote BS. Hence, it can be stated that the proposed protocol improves the performance of system lifetime and networks data accuracy. However the protocol includes few limitations like the chosen CH will be concentrated only on one part of network and after certain number of iterations the clustering process terminates.

Result

[23] Simulation is carried out by MATLAB (MATrixLABoratory). The simulation results had showed that LEACH decreases the communication energy by as much as 8times when compared to direct transmission and minimum transmission energy routing. The other point that had been proven by LEACH is that in the proposed protocol the death of first node occurs over 8times after the first node death in direction transmission, clustering protocol, and minimum transmission energy routing. While in the case of last node death, in LEACH it occurs over three times after the last node death in other protocols. Thus based on MATLAB simulations [23] are more confident that the proposed protocol will outperform the conventional communication protocols in terms of ease of configuration, system lifetime, quality of network and energy dissipation. By making use of this proposed protocol one can attain low-energy and provides the future for micro sensor networks.

3.3.2. CAG

Approach

CAG algorithm forms the clusters of nodes by sensing the same values within the given threshold and these clusters remain unchanged until all the sensor values reach within the threshold over time. With the proposed scheme only one sensor reading per cluster is transmitted on other hand in TAG (TinyAGgregation) the entire nodes in network transmit the readings of sensor. The experimental results of [62] had showed that CAG provides better performance than TAG. CAG calculates and gives accurate answers to the queries by making use of spatial and temporal properties of information. One of the main disparities between CAG and LEACH is that LEACH fails to offer a mechanism to calculate aggregate by making use of CH values. In the proposed scheme, the clusters are formed by sensing the same values. By making use of spatial and temporal correlations the protocol doesn't pay any attention to the redundant data and moreover provides significant energy savings. The proposed protocol can work in two modes, which are illustrated below:

- **Interactive mo*de*** - this mode make use of only spatial correlation of sensed information. The proposed protocol creates a forwarding tree whenever the query is sent out. Further the forwarding path is set along the reverse direction of query propagation. Whenever the users requires the new data from network

this mode need the overhead for broadcasting the query for every time. Frequently rebuilding the tree is waste if the sensed information is nearly same. In case if the information is unchanged then CH nodes and forwarding tree are nearly same. Thus, in interactive mode CAG produces a single set of responses for a query.

- ***Streaming mode*** - this mode takes the benefits from temporal and spatial correlations of information. For query, this mode utilizes the clause epoch duration i for defining the sampling frequency. Now the query is inserted into network only for one time with this clause and in turn for every 'i' seconds it produces a query reply. CAG algorithm generally works in two phases which are query and response. In streaming mode the query phase of CAG algorithm is the same as that of clustering algorithm in interactive mode. One of the major difference between streaming mode and interactive mode response phase algorithm. Thus, in this mode for a query the periodic responses are generated.

Hence, the proposed protocol has an ability to save the significant number of transmissions by avoiding global communications and fine-tuning the clusters by making use of local communications. CAG calculates the results efficiently with high accuracy and assure that the results are within the user-provided thresholds in spite of data distribution.

Result

[62] Computed the results for both interactive and streaming modes. From the results it had been found that the proposed protocol in interactive mode generates results with tiny and frequently bounded errors with dramatically diminished the message overhead than TAG. In streaming mode, the efficiency that is compared to TAG is higher and at the same time it further guarantees that the errors in results are bounded always by user-provided threshold. When the experimental results are computed the average of sensor readings in a network by making use of CAG interactive mode with user-provided error threshold of 20% one can save 68.25% of transmissions when compared to TAG with only 2.46% inaccuracy in result.

3.3.3. EECDA

Approach

EECDA is a protocol for heterogeneous WSNs that merge the data aggregation and energy efficient cluster based routing in order to attain the better performance in terms of stability and lifetime.The main aim of designing the EECDA protocol is to efficiently maintain the sensor nodes energy consumption by relating them in single-hop communication within a cluster. Here, both the data aggregation and data fusion techniques decrease the number of transmitted messages to BS in order to prevent the congestion and save energy. While implementing the proposed protocol, [39] made assumptions which are:

- 'n' number of sensor nodes are distributed within the square field
- after the deployment both BS and all the sensor nodes are immobile
- WSN includes the heterogeneous nodes in terms of node energy.
- CHs executes data aggregation
- BS is not energy limited as in case of other nodes in network.

The aggregation for the proposed protocol is carried out in three phases, which are illustrated below:

- ***Cluster head election phase*** -the proposed protocol makes use of three different types of nodes (that is, normal, advanced and super) that are deployed in a wireless environment where not even a battery can be replaced. The nodes having the higher battery energy are referred as advanced and super and remaining nodes are considered as normal nodes. When compare to normal nodes, in most of the cases super and advanced nodes become CH.
- ***Route selection phase*** - in a particular round once all the cluster heads are chosen then with the help of weighted election probability each of the CH identify its energy residue and further broadcast the data with its CH role to its neighbour nodes.
- ***Data communication*** - in this phase each of non-clusterhead node transmits its information to their CH. Each of CH in the sensor field receives the data from the remaining nodes and further transmits the data to BS.

Hence, EECDA provides better performance in terms of stability, energy efficiency, and network lifetime than LEACH protocols.

Result

[39] Conducted simulations for two sensing fields in order to compare the performance of proposed protocol with EECHA, EDGA and LEACH. In the first

scenario, 100 nodes are deployed in the 100 x 100 square meter area size and in second scenario 200 nodes is deployed over 200 x 200 square met ter. From the simulation results it can be stated that both LEACH and EECHA fails in taking benefit from heterogeneity in both of the scenarios where first and last node dies than EDGA and EECDA. Thus EECDA improves the lifetime of network by 51%, 35% and 10% than LEACH, EEHCA and EDGA. Another benefit with the EECDA is that the residual energy of the proposed protocol is more than other schemes since initially proposed protocol and other schemes has the same initial energy but as the rounds increases EDGA, LEACH and EEHCA the residual energy decreases for both the scenarios. Thus, the more the residual energy the system is more efficient.

3.3.4. Comparison of LEACH, CAG, and EECDA Protocols

No.	Protocol	Proposed by	Description	Pros	Cons
1.	LEACH	Heinzelman, W., Chandrakasan, A. and Balakrishnan, H. [23]	This protocol is introduced to minimize the global energy usage by distributing the load to all the nodes at different points in time.	Randomized cluster head selection	Cost to form cluster is expensive Terminates in a constant number of iterations
2.	CAG	SunHee Yoon and Cyrus Shahabi [62]	Proposed to reduce the number of transmissions and offers approximate results to aggregate queries by	Offers energy efficient aggregation results with small and negligible error	Saves energy only when few nodes change clusters. Resilient to the packet loss

			making use of spatial correlation of sensor data		
3.	EECDA	Kumar, D., Aseri, T. C. and Patel, R. B. [39]	Introduced the protocol to maintain the energy consumption by single hop communication within the cluster	Improves the performance of a network by making use of few heterogeneous nodes in network	Election process of CHs makes network unstable

Table 1: Shows the comparison of Cluster based routing protocols [23], [62] and [39].

Thus, each of the protocol has its own benefits and limitations. Among all the protocols EECDA provides a better lifetime, energy efficiency and stability to network compare to LEACH protocol.

3.4.Chain based Data Aggregation Technique

Some of the protocols that are proposed for this network include PEGASIS (Power-Efficient GAthering in Sensor Information System), COSEN (Chain Oriented Sensor Network for Efficient Data Collection), Enhanced PEGASIS and CHIRON (Chain-Based Hierarchical Routing Protocol) which are discussed in below sub-sections.

3.4.1. PEGASIS

Approach

PEGASIS is a near optimal chain-based protocol and it is designed as an improvement over LEACH. In this scheme each of the nodes will just communicates only with its neighbour node and waits for its turn to transmit the data to Base Station (BS);thus reduces the amount of energy that is spent for one round. Nodes must take turn to become a leader for transmitting data to BS. In a network this scheme evenly

distributes the energy load between the sensor nodes. In a play field, initially [61] located the nodes randomly and thus the i^{th} node is at random location. Here the nodes are organized such that it forms a chain where either it can be accomplished by sensor nodes by with the help of a greedy algorithm by starting from some node. Then again BS calculates this chain and further broadcast to all sensor nodes.

In PEGASIS for building a chain, [61] assumed that in a network all the nodes have a global knowledge about the network and then make use of greedy algorithm. By using greedy approach one can easily construct a chain and it is carried out before first round of its communication. To build the chain, [61] initially start with the node that is located farthest from BS. The chain process is started with this particular node to ensure that the nodes that are farther from BS have close neighbours. In each of the round, for collecting the information each node receives the information from its neighbour node and then fuses the data with its own information. Further, on chain it transmits the fused data to other neighbour. In each round of communication, the leader will be placed at random position as it's essential for nodes to die at random locations. The idea behind the nodes death at random places is just to enable the network robust against the failures. In a round, [61] used a simple token approach which is initiated by leader for beginning the transmission of data from chain end. Since the size of token is small the expenditure on these tokens is also low.

In order to get a clear idea on this scheme, [61] illustrated an example which is shown in figure ().

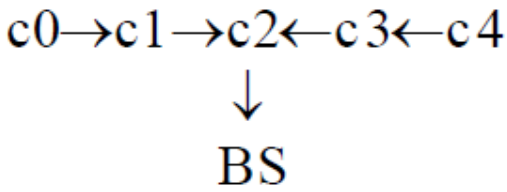

Figure: Shows the token passing approach.

Here, node C_2 is considered as leader. So, C_2 will pass a token to C_0 along the chain and in turn C_0 will pass its information towards C_2 node. Once C_2 node receives the information from C_1 node, it further passes the token to C_4 node. Finally, C_4node passes its gathered data towards C_2 node. In PEGASIS, the data fusion process is carried out at each of the node except the nodes that are present at the end of the chain. Each of the node will fused the data received from its neighbour node along with its own data in order to finally generate a single packet of same length. Hence, in

PEGASIS scheme each of the nodes will receive and transmit one packet in each round and that particular node will become a leader for every 100 rounds. Thus with the implementation of PEGASIS scheme one can easily save energy in various stages like at the time of local gathering, leader receives only two messages rather than receiving 20 as in case of LEACH, etc.

Result

The PEGASIS performance is evaluated through simulation process along with LEACH by making use of 100 nodes in a network. Two sensing fields are being considered which are 50m x50m and 100m x 100m. In 50m x 50m, the BS is located at (25, 150) and in case of 100m x 100m the BS is located at (50, 300). [61] carried out simulation process to identify the number of rounds of communication when 1%, 20%, 50% and 100% of nodes die by making use of direct transmission, LEACH and PEGASIS where each of the node have same initial energy (IE) level. In case if a node dies then it is referred as dead for rest of simulation. [61] Simulation proved that PEGASIS has an ability to attain the following:

- For a 50m x 50m network, PEGASIS achieves nearly 2x the number of rounds than LEACH when 1%, 20%, 50% and 100% of nodes die.
- For 100m x 100m network, PEGASIS attain 3x the number of rounds when compare to LEACH when 1%, 20%, 50% and 100% of nodes die.
- In order to have a full use of entire sensor network, balanced energy dissipation is maintained between the sensor nodes.
- performance is optimal

In a sensor field generally nodes die at a uniform rate after 20% die since the distance among the nodes become greater and frequently enable them to become leaders thus results in causing the energy to drain quickly. As expected, the number of rounds becomes twice the energy/node also becomes twice for the given network size.

PEGASIS provides nearly twice better results performance when compared to LEACH in all of the cases for 50m x 50m and thrice better performance is offered by PEGASIS than LEACH in 100m x 100m.

3.4.2. COSEN

Approach

The operation of COSEN includes two phases which are illustrated below:

- Chain Formation Phase
- Data Collection and Transmission Phase

[45] Considered that all the sensor nodes have a capability of adjusting the dynamic power. So the nodes can fine-tune the amplifier electronics to accommodate/adjust for any required distance.

Chain Formation Phase

In a target field, the sensor nodes are randomly deployed. Now COSEN protocol enables all the nodes to form several lower level chains. The length of the chain formed with this scheme is fixed. By make use of greedy algorithm, the formation of chain begins from the node that is located furthest from BS. Now this node chooses another live node that is nearer to it and verifies whether this live node is connected with any other chain or not. If it is not part of any other chain then the initial node adds the live node to its chain. This chain process continues until its length exceeds the fixed length. If it exceeds the length then the formation of new chain begins. In this way the chain formation goes on until all the nodes in the field are grouped into chains. By making use of triangulation, one can identify the nodes positions. After chain formation now it's time to recognize the leader in a chain. COSEN choose the leaders for each of the chain depending on its remaining energy that is stored in each sensor of chain whereas in case of PEGASIS leader node is selected in each of the round. Additionally, COSEN will not alter its leader for every round instead it just changes after 'n' number of nodes. After the leader selection again a higher level leader is chosen among the leader nodes by making use of greedy algorithm. Higher level leader is only one node which transmits information to BS.

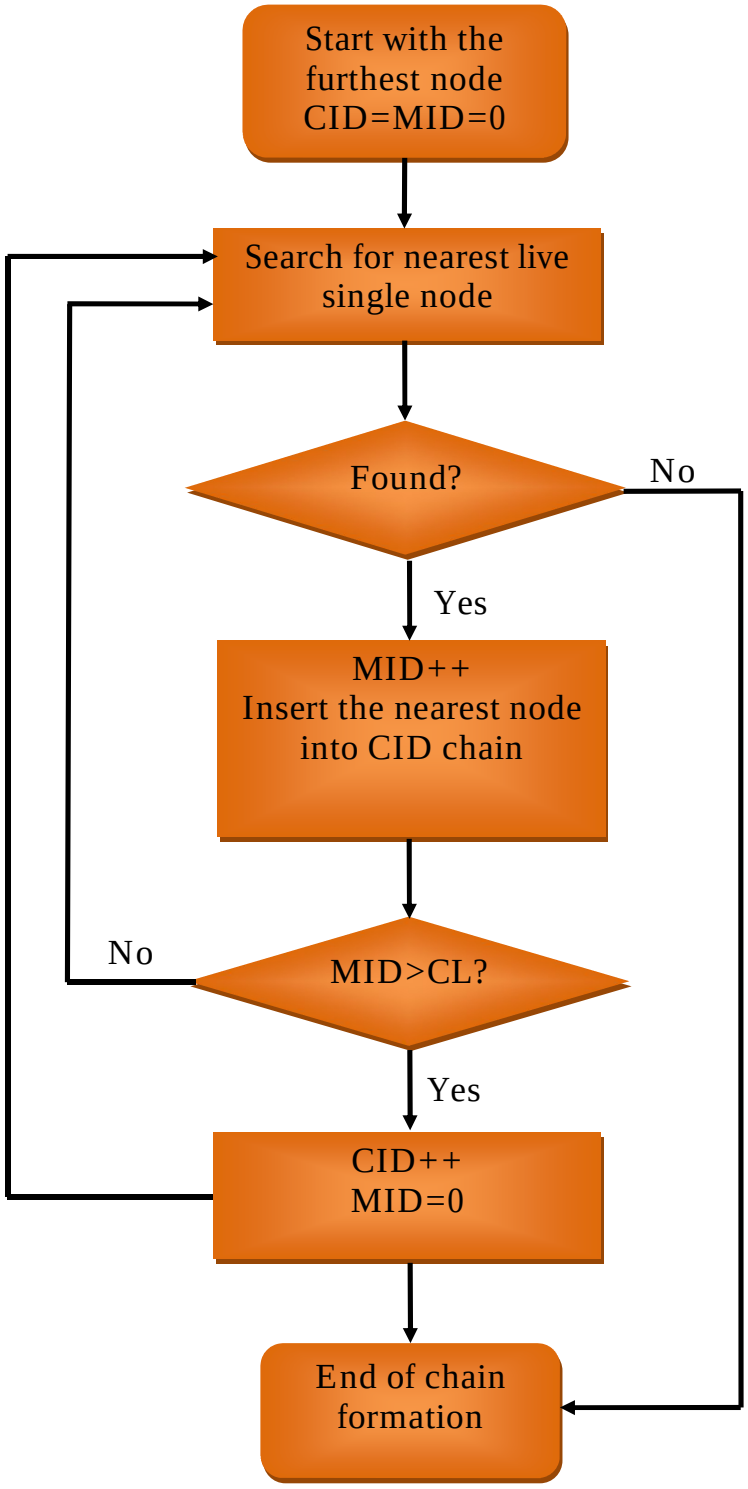

Figure: Shows chain formation algorithm [45, p.4].

Data Collection and Transmission Phase

Once the formation of chain and selection of leaders are completed, the sensor nodes begin the data collection process. The main thing that needs to be noted is that the chain formation phase shouldn't be head of data collection phase always. Chain formation phase will lead only when it is essential to reconstruct the new chains. [45] Assumed that all the sensor nodes have information to send to BSs so the information is aggregated before transmission at each of the node. The token mechanism followed by COSEN is same as PEGASIS.

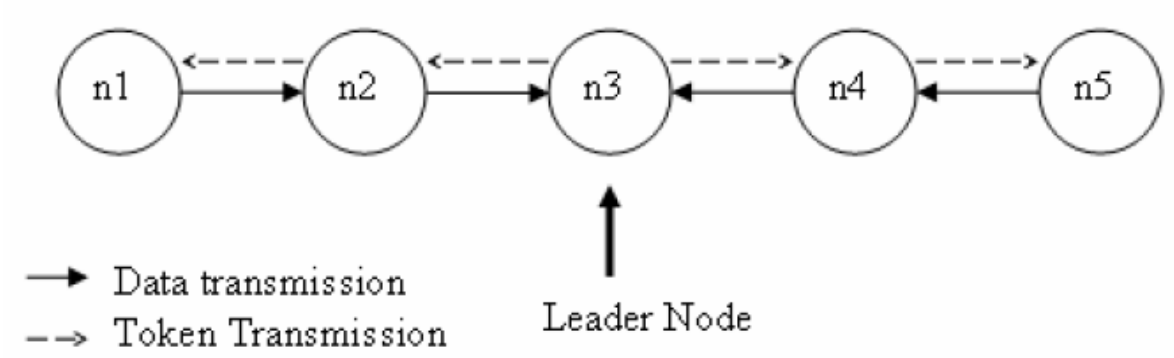

Figure (a): Shows the token passing approach [45, p.4].

From figure (a), n3 node is chosen as leader node and it transmit the token to end of the chain. Now each end node starts transmitting the data to its next node. In this node it receives the data and further fuses with its own data and then sends the information to next node. In this way the data is propagated from last node to chain leader. Now all the leaders transmits the data to its higher level chain by using the same process until the higher level leader receives all the data. Finally, this higher level leader transmits the data to BS after data fusion.

Result

In simulation process [45] considered 100 nodes and is placed randomly in a field of 50m x 50m. [45] Make use of Cartesian coordinates to identify the sensor locations. BS is placed at (25, 150). In simulation, [45] compared the proposed protocol with PEGASIS. Initially the simulation is done on energy consumption.

After several hundreds of rounds the energy consumed is same for both. However, COSEN spend the total energy in distributed way such that the network can work more number of rounds before the first sensor node dies. On other hand, in PEGASIS the first node dies at 350 rounds and in COSEN the first node dies at nearly 450

rounds. The next simulation is carried out on lifetime pattern and time required for completing one and multiple rounds. It can be noticed that COSEN outperforms LEACH by evade from the overload that is caused due to dynamic cluster setup and reducing the number of long distance transmissions. Even it causes much less delay to deliver the data to base station from distant nodes than PEGASIS. Hence, it can be stated that there is an ultimate improvement of COSEN from PEGASIS and moreover the delay is very low in COSEN.

3.4.3. Enhanced PEGASIS

Approach

The proposed scheme enhanced PEGASIS makes use of multiple-chaining and concentric-clustering scheme. This process mainly includes four phases which are defined below [34]:

- ***Level assignment*** – in WSN each of the sensor nodes allocates its own level from external BS. Here the level is represented in the form of concentric circle by making use of signal strength of BS. In the sensor field, base station assigns level to each of the sensor node after computing the BS based on number of levels.
- ***Chain construction in levels*** – the chain construction of the proposed protocol is similar to that of PEGASIS protocol. In each level, the chain is constructed by making use of greedy algorithm. Later the external BS broadcast the information of chain to sensor nodes in each of its level after completing the chain construction process.
- ***Selection of head node in chain*** – On the chain only one sensor node is chosen as head node which has highest level. In each of the level a head node is responsible for gathering the data from other sensor nodes which are located in same level and head node in adjacent level. Finally, in the first level the head node sends the information to external BS once it aggregating its own data and gathered data.
- ***Data Transmission*** – after constructing the chain the head node is selected in each level. Each of the sensor nodes delivers its own information and the data acquired from its neighbour nodes along with the chain.

Thus, in this way the data is transmitted from sensor nodes to BS through chain.

Result

To prove the efficiency of the EPEGASIS the total residual energy is computed for all the sensor nodes. [34] drawn a graph on residual energy for all the nodes from 1st round to 1500 round in a field size of 100m x 100m and in case of 200m x 200m it is from 1st round to 3000 round. The results acquired are compared between PEGASIS and Enhanced PEGASIS.

From the simulation result it can be found that the Enhanced PEGASIS provides better performance compared to PEGASIS as the network size becomes larger. Thus, by making use of this proposed protocol one can save energy for large network size in WSN.

3.4.4. CHIRON

Approach

The working of CHIRON protocol mainly includes four phases, which are illustrated below [40]:

- Group Construction Phase
- Chain Formation Phase
- Leader Node Election Phase
- Data Collection and Transmission Phase

Group Construction Phase

This phase classify the sensing field into numerous small areas with the intention that the proposed protocol can build multiple shorter chain to decrease the data propagation delay and redundant transmission path in remaining phases. CHIRON makes use of Beam Star technique to organize its group rather than using concentric cluster as in case of EPEGASIS. In a field once the nodes are scattered then the base station removes gradually the whole sensing area by sequentially changing different directions of antenna and levels of transmission power in order to send the control data to the entire nodes present n field [40]. Once the entire node receives the control packets then each of the nodes can easily recognize which group they belong to. For instance, consider a group where R (the transmission range of BS) =1..3 and θ(the beam width of directional antenna) = 1..2.

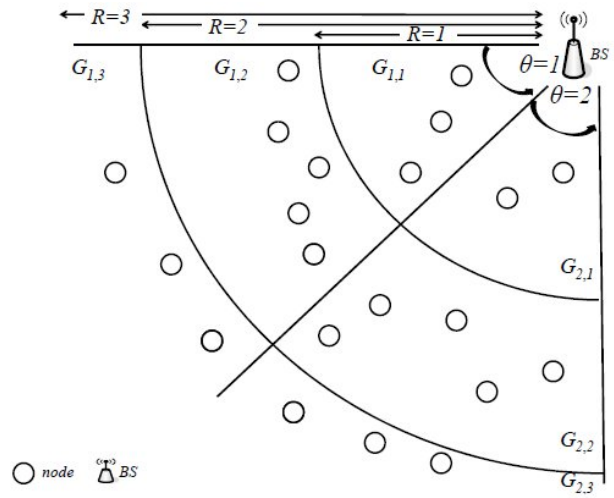

Figure a: Shows the grouping example with R=1..3 and θ = 1..2 [40].

Chain Formation Phase

This is the second phase where the all the nodes link together to form a chain. This process of forming chain is as same as in case of PEGASIS. Greedy algorithm is used in this phase to identify the nearest node to link that node and then form as a newly initiate node to link with other node. This process goes on until all the nodes connected and at last forms a group chain.

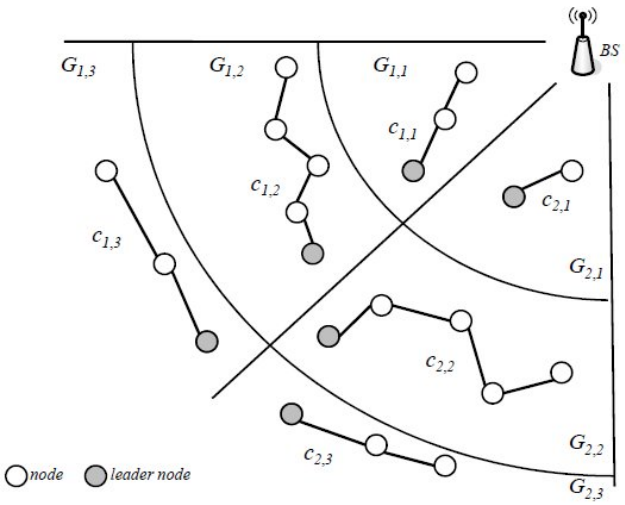

Figure: Shows the group chains that are built from figure a [40, p.3].

Leader Node Election Phase

In the third phase it's essential to choose a leader node in every group chain for gathering and forwarding the aggregated information to BS in data transmission process. In PEGASIS and EPEGASIS the leader for each of the chain is chosen in

round-robin whereas in CHIRON it selects the chain leader depending on the group nodes maximum value. In each group primarily the node that is present at farthest from BS is chosen as the group chain leader [40]. Later, for each data transmission round the leader node is selected based on maximum residual energy. In order to recognize which node to be the leader for next transmission round the residual power data of each of the node is piggybacked with fused data to chain leader.

Data Collection and Transmission Phase

This phase starts the data collection and transmission phases. In CHIRON, the transmission process is same as PEGASIS. Nodes in each of the group transmit the gathered information from their nearest nodes and transmit to chain leader [40]. From farthest groups, the leaders of chain collaboratively relay their aggregated sensing data to BS in multi-hop.

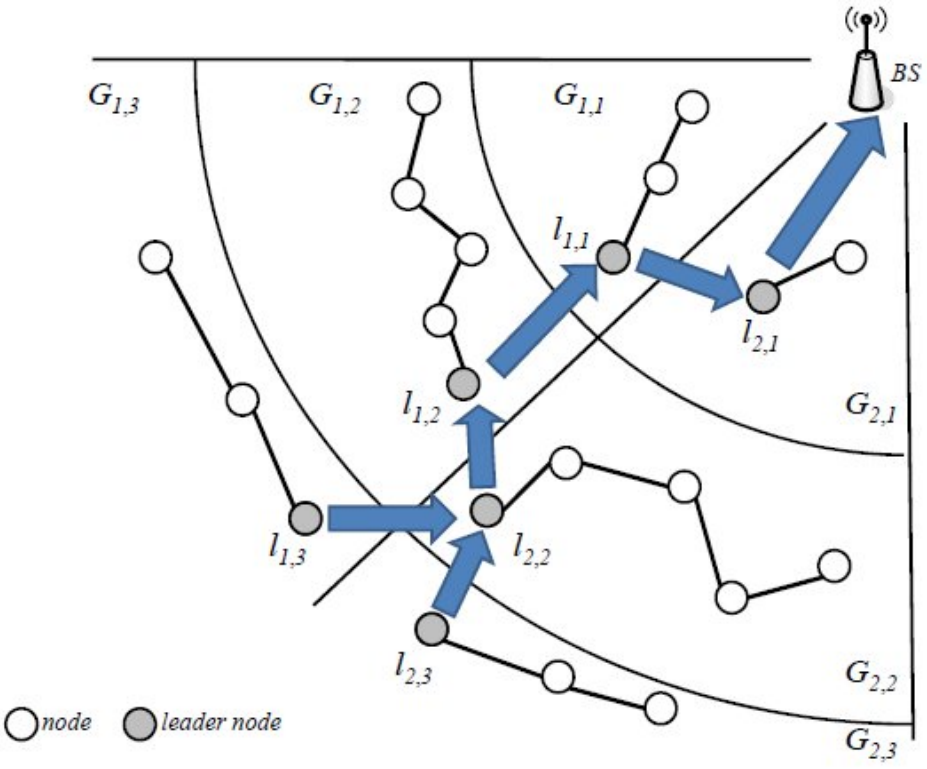

Figure: Shows the data transmission flows [40, p.3].

Thus, in this way the data is fused from all nodes and send to BS through various chain leaders.

Result

To evaluate the performance of CHIRON, MATLAB simulation tool is chosen and compared its results with the two schemes which are PEGASIS and EPEGASIS in terms of redundant transmission hops, network lifetime and average transmission delay. Here, three different sizes of sensing area are chosen which include 100m x 100m, 200m x 200m and 300m x 300m. In each of the area 100 sensor nodes are randomly deployed and BS is placed on corner of sensing area [40]. Initially, the average propagation delay and redundant transmission path is simulated for three schemes which are CHIRON, EGEGASIS and PEGASIS.

The performance of CHIRON is better when comparing to EPEGASIS and PEGASIS schemes. The average delay is improved by 15% and 1.68 times. In case of redundant path it improved by 30% and 65%. Even in network lifetime, CHIRON provides better performance compare to other schemes. Apart from extending the lifetime of first node death even CHIRON improve the lifetime of network. For large simulation areas the extension is nearly 50% ~ 23% than PEGASIS and EPEGASIS. In case of small simulation area the improvement is nearly 14% ~ 7% than PEGASIS and EPEGASIS. This is mainly because of the reductions in length of the chain and redundant transmission path in CHIRON protocol.

3.4.5. Comparison of PEGASIS, COSEN, Enhanced PEGASIS and CHIRON Protocols

S.No.	Protocol	Proposed by	Description	Pros	Cons
1.	PEGASIS	Stephanie Lindsey and Cauligi S. Raghavendra[61]	Each node communicates only with neighbor node, take turns transmitting to BS hence decreases energy consumption per round.	network energy dissipation is balanced	1. Chain leader is chosen by taking turns 2. improper data delay time
2.	COSEN	NahdiaTabassum	Introduced	improve the	Lot of

		, QuziEhsanulkabirMamun and YoshiyoriUrano [45]	protocol for gathering data in a time and energy constraint sensor network.	consumption of energy and transmission delay	unneeded transmission paths
3.	Enhanced PEGASIS	Jung, S. M., Han, Y. J. and Chung, T. M. [34]	Proposed to resolve the issues faced with the PEGASIS based on concentric clustering scheme	Alleviate network lifetime and redundant transmission path	1. Consume more energy 2.Longer transmission delay
4.	CHIRON	Kuong-Ho Chen, Jyh-Ming Huang and Chieh-Chuan Hsiao [40]	Proposed based on beamstar concept and split the sensing field into number of smaller areas to create multiple shorter chains.	Diminish unnecessary transmission path for saving energy and data propagation delay.	numerous short chains

Table 2: Shows the comparison of Chain based routing protocols [61], [45], [34] and [40].

3.5. TREE based Data Aggregation Technique

Some of the protocols that are proposed for tree based are TREEPSI (TRee-based Energy Efficient Protocol for Sensor Information), TCDGP (Tree-Clustered Data Gathering Protocol) and PERLA (Power Efficient Routing with Limited Latency) which are clearly discussed in below sub-sections.

3.5.1. TREEPSI

Approach

Multi-hop routing is more preferable for densely deployed networks than making use of single hop communication since multi-hop routing consumes less energy. Some of the assumptions that are made for proposing TREEPSI are illustrated below [57]:

- nodes are stationary
- each of its node has the information related to its location
- Sensor network includes energy constrained and homogeneous sensor nodes with uniform initial energy.
- All sensor nodes have power control and have an ability to transmit the information directly to sink or other sensor node.

In TREEPSI, researchers make use of different approach for collaboration among the nodes. In the proposed method, initially the sensor nodes are randomly deployed in playing field. Thus $node_i$ (i^{th} node) is placed at random location. Let us consider that $node_i$ as the root and now a tree-like hierarchical path of nodes is constructed. Basically there are two ways for path computation. The first way is to calculate the path centrally through base station and broadcast the information of the path to the network. The second way is that all of the nodes build the same tree structure locally by making use of common algorithm in every node. Once the tree construction is completed, one can make use of any one of the two different ways for gathering data from field. The two ways are [57]:

- In first method, initially root started data gathering process by sending small control packet to children nodes with the help of standard tree traversal algorithm. Here as the control packet size is minute it consumes very less energy and with little more delay.
- In second method, firstly leaf nodes sense the information and then it just forward this data towards their parent node. Once the parent node receives the data then this node fuse the received information with its own data. The result of this information is forwarded to its parent node. This process continues until the root node receives data.

Among these two approaches, the second method results in less delay so it's essential to make use of some multiplexing scheme such as CDMA (Code Division Multiple Access) or TDMA (Time Division Multiple Access) in order to avoid collision at the

time of simultaneous data transmission by one or more than two nodes to a common parent node. Although the approaches are different, in both the methods the data are collected at root and then this root further take responsibility to transmit the information to BS. This particular path is used for communication until its $node_i$ dies due to low battery power. Once the $node_i$ is dead then a new tree-like path is built by considering $node_i + 1$ as root and further used for several rounds of communication. Here in this process, the path computation is done only once and then it is followed by several rounds of data communication. Due to this reason the overhead per communication round is very less than the energy spent in data collection phase. While designing this protocol, an assumption is made that there is no node mobility thus there will be no change in path until and unless the root is dead.

Routing Algorithm

In sensor field for each $node_i$ do If (*the node is having energy more than what is necessary to transmit a packet to the BS*) then 1) Build a tree structure consisting of all the live nodes with $node_i$ as root node. 2) While ($node_i$ is having more than minimum necessary energy for transmission to base station) do i) Gather the information in tree structure right from its leaf node to root node as explained in step-1. ii) Transmit the gathered information to base station by $node_i$.

Table: Shows the routing algorithm [57, p.3]

Result

To prove the effectiveness of TREEPSI, this protocol was simulated by randomly distributing 100 nodes within the area 50m x 50m and 100m x 100m.

In these sensor fields the sinks are located at (25, 150) and (50, 300). The other aspect that is being considered is different Initial Energy (IE) levels such as 0.25J, 0.5J and 1.0 for each node. The length of the data packet is imagined as 2000 bytes. To build the tree like path, [57] considers a threshold distance for choosing a node as child of another node. If the distance among $node_i$ and $node_j$ is greater than its threshold

value then $node_j$ is not considered as $node_i$ child since there is may be a chance that a $node_k$ may be present nearer to $node_j$ when compared to$node_i$. This assumption is truly valid due to its dense deployment of nodes in sensor field. From simulation study, [57] had been observed that roughly threshold equals to 9 meter works excellent to comprise the entire nodes in tree like path of 50m x 50m sensor field. In the field the density decreases as nodes die. So it's necessary to raise threshold linearly in its path construction.

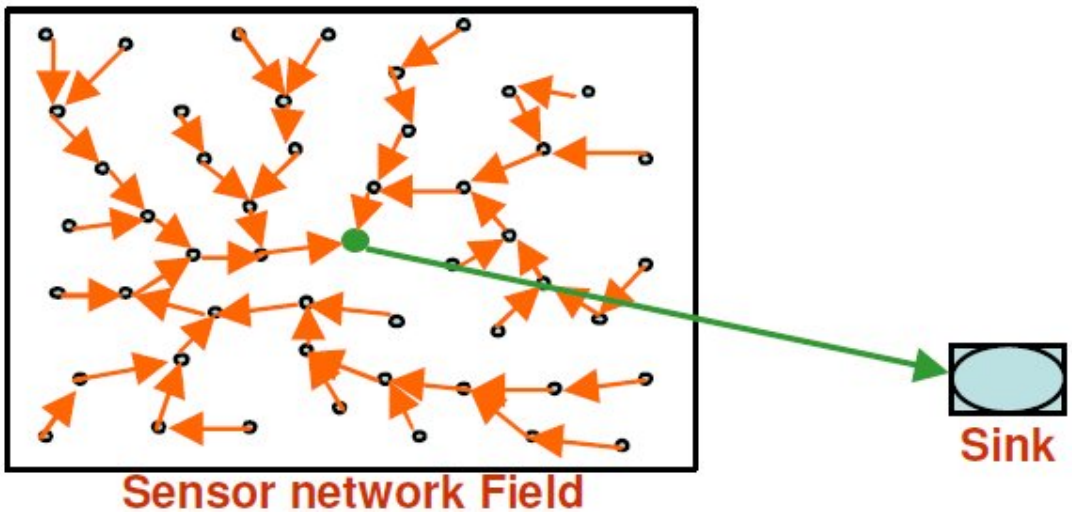

Figure: Shows the path tree constructed for data gathering in TREEPSI [57, p.3].

In order to calculate the maximum possible rounds of communications for different IE levels are considered.

100 nodes are randomly distributed in 50m x 50m sensor field. From initial distribution, after 1000 rounds of communications the remaining live nodes are having 0.25J IE. Thus, it had been proved that the nodes in sensor fields randomly died and TREEPSI protocol helps to gather the information from whole area of field throughout the whole network lifetime.

In sensor networks, after observing the death rate of nodes from simulations of TREEPSI and other existing schemes, [57] gathered the data about the node death rates and plotted a table against communication rounds and maximum possible communication rounds for various energy levels.

Thus, from the simulation result it can be seen that in all energy levels the performance result of TREEPSI is much more better when compare to other existing protocols like PEGASIS.

3.5.2. PERLA

Approach

[43] Introduces PERLA, a network layer protocol which is mainly designed to work on top of IEEE 802.15.4 MAC (Medium Access Control)/PHY (Physical) stack. In [43] research, [43] implements PERLA on top of IEEE 802.15.4 which considers the power management to its account and further address few issues which are related to adoption of it standard like synchronization between the nodes. [43] Make use of an algorithm that depends on spanning tree for ordinary routing operations and also for resorts to develop an alternative path whenever a malfunctioning is identified.

In WSN, [43] mainly considered two causes of communication unreliability that is link failure and node failure. Whenever a node receives a corrupted packet because of channel errors, wrong synchronization or collisions then it results to link failure. Node failure is just a malfunctioning of sensor node either due to battery depletion or any other accidental causes. PERLA follows a particular procedure to identify the link failures and to recover it. So there is no chance of any unnecessary route changes and to decrease overhead traffic. PERLA recognizes the link failure with the implicit acknowledgements whereas the recovery is carried out by retransmission and caching strategy in its neighbour nodes. At initializing phase, each and every node forwards the packets only after choosing a primary node and set of upper-layer neighbour nodes as backup parents. Regarding to multi-path approaches, in a network PERLA decreases the overall traffic by triggering retransmission only if it is essential. One more thing that needs to be mentioned about PERLA is that it doesn't depend on link layer acknowledgements and retransmissions.

To get an in-depth knowledge on the proposed protocol, [43] considered a network with a spanning tree in which all the nodes are allocated a level that represent the hop distance sink node and among themselves. The other assumption that is made is the data gathering from all nodes to sink should be within a time known as epoch which is represented as e. If it is given that the depth of tree is n then the time for each level allocated is e/n. The timing for data forwarding process among the adjacent levels is illustrated in figure (A). Here, sub-epoch are further classified into eight phases with sub-epochs related to adjacent levels are shifted in order to correct the phase coupling.

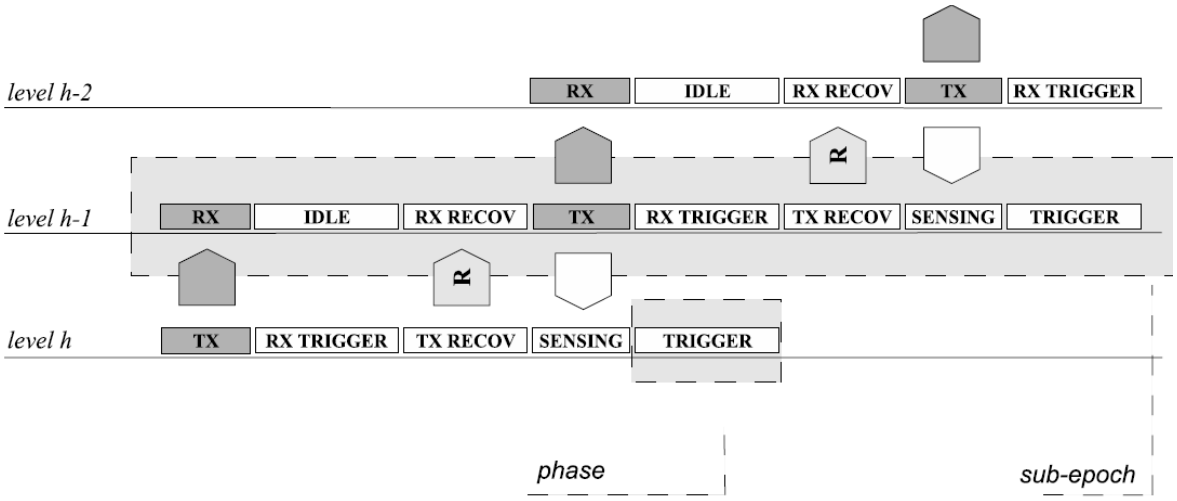

Figure (A): Shows the organization of the phases of PERLA [43, p.2].

If the network is in error free situation then only main transmission (TX) and reception (RX) phases are utilized since the remaining phases are only used to identify the link failures and recovery procedure. In case if the network is affected with link failure then the detection process is carried out at the time of sensing phase. Parent nodes forward the data that is being received from its child nodes. This process overhears by originating nodes related to the transmitted packets and then executes an implicit acknowledgment. Once the error is recognized, a recovery procedure is started and enables the originating nodes to enter into trigger phase and forces to retransmit their previously cached data. To clearly understand this process let us consider a simple network with 14 nodes is shown in figure (B).

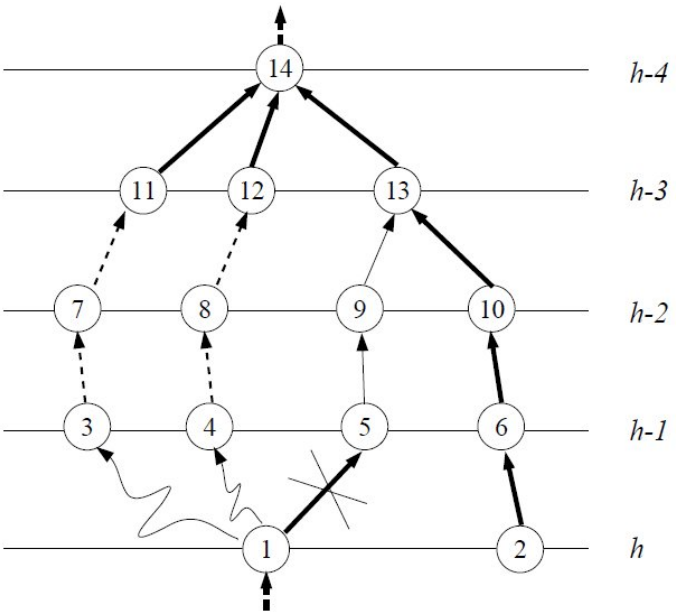

Figure B: Shows the retransmission scenario [43, p.3].

Here, at node 1 data is originated and now this node sends the information through its primary parent node (that is node 5). Node 3 and node 4 are within the hearing range and even these nodes are chosen as backup parents by node 1. Thus there is a chance to overhear and store the data that is being transmitted. In this case node 5 receives a corrupted data packet from node 1. Generally, node 5 must forward the information that is either a sensed data or received data from its TX phase. Due to link failure, this transmission process cannot be carried out. Now, node 1 will understand that the packet was not received correctly and further broadcast a retransmission request at the time of trigger phase. Both node 3 and node 4 react to the broadcast message and retransmit the previous stored information in aid of node 5 misbehaviour. Thus in this way link failure is detected and completed its recovery procedure. As shown in figure (B) the data reached its destination eventually by making use of only limited amount of multiple paths and the data aggregated at node 14.

Result

PERLA was mainly developed for NS-2 simulator by making use of IEEE 802.15.4 implementation. Here, in this process the energy that is needed for PERLA can be easily controlled acting during TX recovery and TX trigger phases. As the duration of the phase becomes shorter then probability will be low and thus all nodes can complete their transmissions within their time. The performance of the proposed approach is assessed based on two alternatives, which are:

- basic protocol which uses unacknowledged transmissions
- protocol which depends on link layer acknowledgments and retransmissions

The data gathering model is adopted as described in the approach section for all the cases. The structure for the two proposed alternatives is displayed as shown in figure (A) and further it is simplified by considering only TX and RX phases and all the nodes are set in sleep state in other phases. In a network, the nodes are placed randomly based on the uniform distribution and all are assumed that it generates information. Here the failures of nodes are not considered and moreover assumed an ideal channel with the intention that the link errors will cause only because of collisions.

For evaluating the performance of PERLA [43] considered two things which are energy efficiency and reliability. The energy efficiency is defined as ratio computed between the energy spent on each of epoch and its reliability. Reliability defined as the ratio among the originators which receives aggregate by BS and total number of nodes. In various scenarios, [43] tested the protocols by considering various network densities and sizes. However the introductions of ACKs and retransmission have effect on its performance. On other hand, PERLA has an ability to do better than the alternative protocols. It can be justified by considering two factors. The first factor is that although the retransmission and links layer ACKs increases the collisions, PERLA will make a separate schedule for retransmission and it will not negatively effect on the first transmission. Apart from it, the retransmission of the proposed protocol will not take place at tree depth where the link failure exist instead the packet may be stored in the nodes that are closer to BS so it start from that particular node. Thus the performance of PERLA shows better scalability when compared to other alternatives.

3.5.3. TCDGP

Approach

Numerous protocols had been introduced for better performance like LEACH (for cluster-based), PEGASIS (for chain-based) and TREEPSI (for tree-based). However, all these protocols have some limitations.So, an improved version is proposed by merging both the cluster and tree based protocols known as TCDGP. From [17] related work, it had been identified that the tree-based protocols (such as TREEPSI)

provides better energy efficiency when compare to chain-based and cluster-based protocol. If few sensor nodes send information to sink then this data will make diversions for nodes. However, this process will result to more power dissipation while gathering the data. This type of situation occurs while constructing the binary tree paths particularly in case of large sensor field. So to improve the reduction of power dissipation [17] introduced a novel protocol which is formed by combining both the tree-based and cluster-based protocol. Before going to discuss about the TCDGP protocol now let us see the flowchart of proposed protocol which was represented in figure (A).

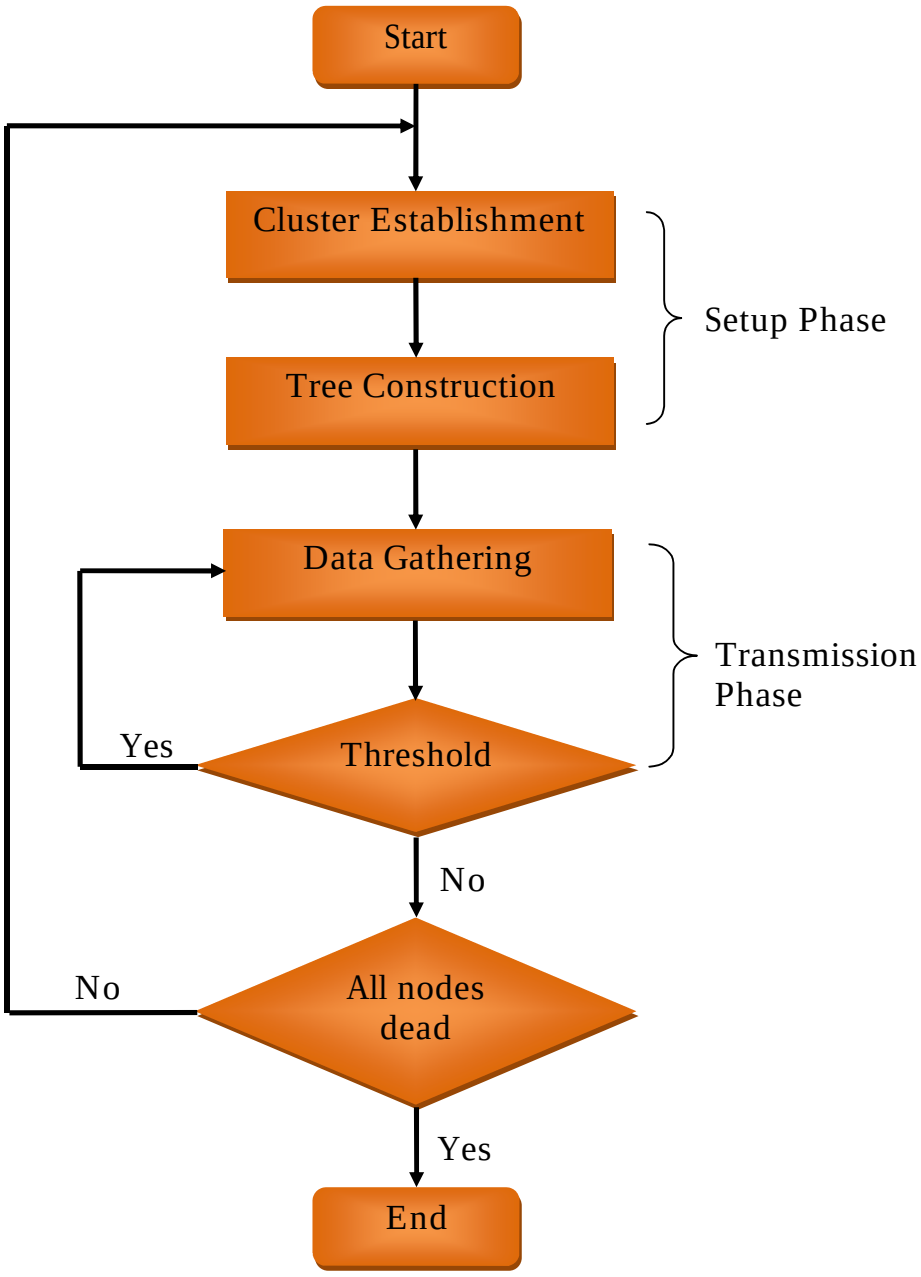

Figure A: Shows the flow chart of proposed protocol [17, p.30].

Some of the assumptions that are made while proposing the protocols are illustrated below:

- Each and every node in a network has an ability to directly transmit the message to any sink or other nodes.
- Based on the transmission distance, each of the sensor nodes can tune the magnitude with the help of radio power control node.
- In wireless sensor networks all the sensor nodes have equal initial power.
- Every sensor node has the data related to their location.
- Once all the sensor nodes are deployed, every sensor node is fixed.
- Humans cannot maintain wireless sensor networks.
- In WSN, all the sensor nodes follow the same process and also they will have same communication ability.
- All sensor nodes are randomly deployed in sensor field.

In a network, the sink has an ability to acquire the information related to sensor nodes location and energy at any manner. In sink, one is recorded at its initial state when the nodes are deployed and other when the sink broadcast the entire network; finally it receives the back message from sensor nodes.

A. Cluster Establishment

Setup Phase:

Generally this phase includes two steps which are:

- Cluster formation
- Cluster head selection

In a network, if BS forms primal cluster then there is no chance to have a much change since all the nodes are stationary. However, in each round the chosen CH may be different in same cluster. In a first round, initially BS classifies the network into two sub-clusters. Then it further continues the process by splitting the sub-cluster again into minute clusters.

Neighbor id	Residual Energy	Distance	Distance to BS	State	Weight

Table: Shows neighbor table information [17, p.30].

If a node wants to be CH then that particular node must be located at center of cluster. Once the sensor node is chosen as the CH then it broadcast the message in a network and further invites other nodes to link its cluster. Except CH, all other nodes in the

network select their own CH's and then send join message to its CH based on the power of received broadcast message. Once CH receives join message from its neighbor node then CH will allot a time slot to transmit information. When the first round is completed and primal cluster topology is formed then BS doesn't have any responsibility to choose CH. Now the mission of forming cluster is shifted from BS to its sensor nodes. So based on the node's weight value a new CH can be chosen within the cluster based. The pseudo code for the operation of all this process is illustrated below:

Initialize
{
1. Base Station: obtains N number of clusters;
2. The network is divided into N clusters;
3. Select CH from every cluster;
4. Notify the node to be CH
}
Repeat:
{
1. Node i: if (node receives the notify message from BS)
2. works in CH mode;
3. If (Receives the broadcast message from CH node)
4. Work in sensing node
}
For CH i:
{
1. Receive information born cluster member j;
2. Calculate the weight value $W_i$and W_j;
3. If (W_i> W_j),W_i work in Ch;
4. Else I work in sensing mode;
5. Notify 'j' to be CH;
}

B. Constructing Cluster Based Tree

Sink will gather the data from each of the cluster (which are labelled by CH) and then construct a path in Minimum Spanning Tree (MST) to calculate the tree path. In

Greedy algorithms, MST is used to work out the undirected weight graph problem. Even though, if some connection links are eliminated still the sub-graph will have the ability of connection. Due to this reason, sub-graph has an ability to decrease the sum of the weights. If a sub-graph has a least sum of weights then it should be tree like framework. Spanning tree enables all the nodes to conform to tree definition that is connected in graph. If a connected sub-graph has minimum sum of weights then it should be a spanning tree. But in converse, this statement is not absolutely correct. There may be chance that there exist more than one minimum spanning tree in a graph and however the sum of weight must be same. In case, if [17] make use of Brute force to identify the spanning tree then it generates enormous computation time. So to overcome from this issue, [17] used Prim algorithm to find MST.

C. Data Aggregation

After the establishment of routing mechanism, each and every node initially gathers the data from its neighbour nodes and then transmits the information to its upper level nodes. Now it's chance for upper level node to fuse both the sensed data and received data. Once the information is fused by upper level node then this particular node send the fused data to its upper level node. This process continues until root node, CH aggregates the information in cluster. The completion of this whole process is considered as round since all the root nodes finishes transmitting data.

Result

From the simulation study, [17] had been demonstrated that the new data gathering protocol for WSN has an ability to improve the evenness of dissipated network energy, reduce the consumption of energy, further extends the life span of network. [17] Proposed method has numerous advantages for data gathering in wireless sensor networks. This protocol reduces consumption of energy by restricting the direct communication among the sensor nodes and sink. By making use of threshold mechanism in this protocol, the number of nodes alive increases and further improves its lifetime when compare to other protocols. Here, each and every node has a chance to be parent so this protocol protected the death of parent node slowly. The proposed cluster-tree based protocol operates in two phases where first phase enables the users to improve the lifetime of a network by simultaneously balancing the nodes energy consumption. Second phase decreases the possibility of forming tree structure overhead.

3.5.4. Comparison of TREEPSI, PERLA and TCDGP Protocols

No.	Protocol	Proposed by	Description	Pros	Cons
1.	TREEPSI	Satapathy S. S. and Sarma, N. [57]	Proposed to give 30% better performance when compare to PEGASIS in terms of energy efficiency	Consumes less power during data transmission	Path has made an alternative route in the topology
2.	PERLA	Messina, D., Ortolani, M. and Lo Re, G. [43]	Introduced to identify and recover from the link failures.	keep away from unnecessary route changes	Requires more energy for identifying error and recovery procedures
3.	TCDGP	Gurpreet Singh Chhabra and Dipesh Sharma [17]	Proposed protocol by considering both cluster and tree based protocols where it improves the consumption of power by improving First Node Death (FND)	Reduces consumption of energy	Require methods for recovery procedures

Table 3: Shows the comparison of Tree based routing protocols [57], [43] and [17].

3.6. GRID based Data Aggregation Technique

Protocols that are being introduced for grid based is GROUP (Grid-clustering Routing Protocol) and ATCBG (Aggregation Tree Construction Based on Grid) which are clearly illustrated in below sub-sections.

3.6.1. GROUP

Approach

GROUP have been introduced in Liyang Yu, Neng Wang, Wei Zhang and ChunleiZheng (2006) research paper which is an energy-efficient and cluster-based routing protocol for WSN. In this designed protocol, one of the sinks dynamically, proactively and randomly builds cluster grid in order to forward query message and data packets [42].

Working

In GROUP, all the sensor nodes are classified into several clusters dynamically. Within the cluster, only one node is considered as cluster head (CH) which is selected dynamically. All CHs form a virtual cluster grid and further the sinks transmit the data queries to all the nodes through cluster heads. Once the sensor nodes receive the query from CH then it will check collected data with the query. If the data matches with the query then it sends out data to its CH via short-range radio. Further the data packet is recursively forwarded by CH to its upstream CH until it reaches its sinks that generated query. Here, in this process CH can make use of data aggregation expediently to save energy and to reduce the number of data packets. Now let us see how the data is forwarded during cluster grid construction.

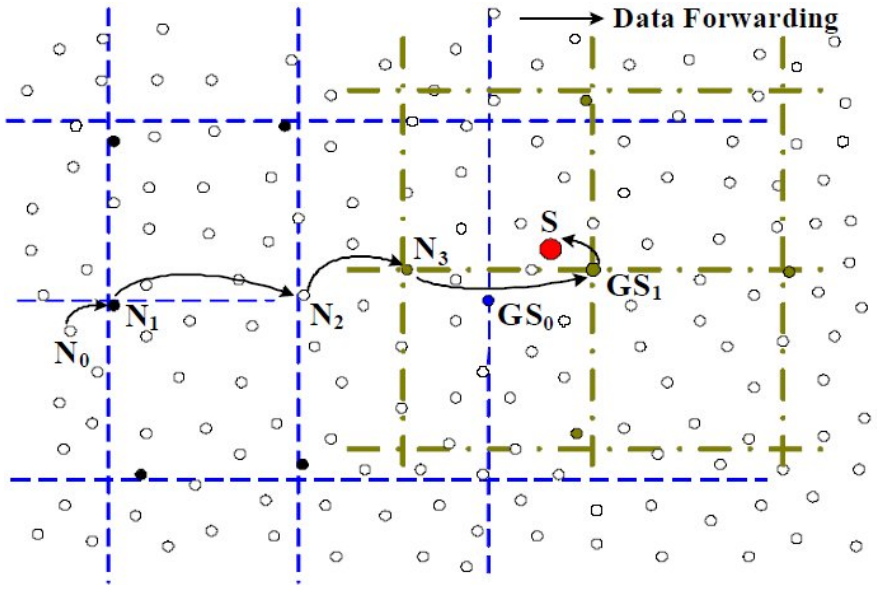

Figure: Shows the data forwarding during cluster grid construction [42, p.3].

To get a clear idea on this concept, let us consider that the network is transferring to new cluster grid whose grid seed isGS_1. Whenever the node N_0 sends data packet to N_2through N_1 then based on previous cluster grid N_2 receives and further transmits to node N_3rather than transmitting to GS_0. Since node N_3 informed its neighbour node that it became a CH of new cluster grid during cluster grid construction. Thus the data packet is transmitted to sink S through GS_1 and further construction of cluster grid will not have any impact on data forwarding.

Generally in WSN, the sensor nodes are prone to failure. In GROUP, only the CH failure will impact on query forwarding and data forwarding. Whenever the CH fails, it broadcasts a CH search packet in order to elect a new CH and further replaced with the old CH. Once new CH is chosen, it identifies its upstream CH and downstream CH by broadcasting CH-recovery packet. Hence, in this way CH forwards the data packets to its sink which are selected dynamically.

Application: GROUP can be used in real-time applications such as forest fire detection. To identify the forest fire, numerous sensor nodes are deployed in forest where these nodes have ability to measure environment temperature, smoke and relative humidity. With the help of this protocol, these entire nodes are organized in the form of cluster in order that each and every node has corresponding cluster head at certain time. Thus sensor nodes receive the query message from sinks and further forward data packets through GROUP to sinks [42].

Result

GROUP is implemented in NS-2.27 and underlying MAC is 802.11 DCF. The performance of the GROUP is evaluated and analysed with three metrics, which are

Energy Consumption

From the simulation result [42], it had been proven that the energy consumption of GROUP is lower than LEACH in the case of 75 sensor nodes. For more sensor nodes, LEACH consumes maximum energy when compare to GROUP.

Thus in all cases, the gap between maximum and average energy consumption is small for GROUP than LEACH. It means that GROUP has better distribution of energy consumption than LEACH since in LEACH as the sensor nodes are increased it has worse energy consumption distribution.

Packet delivery fraction

From its simulation result [42], one can notice that GROUP has better packet delivery fraction when compared with LEACH particularly in the case of more sensor nodes. Since in case of LEACH more collisions are reported in bigger scenario field as all cluster head forward the data packet from cluster members to sink.

Whereas in case of GROUP, only the cluster head forward data packet to its upstream that is closer to sink. Here only radio range varies with its cell size.

Average end-to-end delay of data packet

LEACH has lower average delay when compare to GROUP since it is multi-hop clustering protocol whereas LEACH is single-hop clustering routing protocol. However, GROUP has lower average delay with less sensor nodes.

Hence, it can be concluded that GROUP is more scalable clustering routing protocol and provides better performance with more sensor nodes when compared to LEACH.

3.6.2. ATCBG

Approach

ATCBG is proposed by making some improvements over GROUP. This protocol is aimed at fusing the periodical data gathered by whole network nodes in event-driven WSN. ATCBG is designed by considering some assumptions which are defined below [30]:

- In a network, there exist only one sink which is static.

- All the nodes in the network are fixed and each of the nodes knows its own locations which can be executed by making used of positioning techniques.
- Based on its actual distance, nodes have an ability to adjust its transceiver power.

The main aim of ATCBG is that the aggregation tree is constructed by taking sink as center of grid and cell size is R. The complete network is classified into grids and each of the grid forms cluster. Here CH is chosen by considering distance to grid center, residual energy, etc. and it takes responsibility for data fusion. All CH forms a tree structure and JianShu et al. research the aggregation tree constructed is cluster-tree structure [30].

1. Aggregation Tree Construction

The construction of an aggregation tree is started by sink and it initially broadcasts tree construction message. The tree construction message will process according to its grid coordinates. For instance, consider node A and node B in WSN. Here two cases can be considered, the first case is that two nodes come from same grid and in second case both node A and node B belongs to neighbor grids.

Case 1: If node A and B belongs to same grid

Initially node A join aggregation tree and considers node B as CH. Now it records the information of CH.

Case 2: If node A and B doesn't belongs to same grid

Firstly node A verifies whether it joined the tree aggregation or not. In case, if it is joined then node A will again verify whether node B consider node A as its parent node or not. If node B accepts A as parent node then A will consider node B as its child node. On other hand if node A has not joined the tree then this particular node compete for CH.

Algorithm: ProcessTreeConMsg(msg)
if(A.grid=B.grid) //A and B in the same grid
{
A.clusterHead=B; //A take B as cluster head
A.isJoinTree=1; //mark itself has join the tree
A.recordCHInfo(B); //record information of CH
}
if (abs(A.grid-B.grid)==1) //A and B from neighbor grids

```
{
   if(!A.isJoinTree)                    //A has not join Tree
A.competeForCH();
else
{
  if (B.father=A)    //B take A as father
{
A.child=B;
A.recordChildInfo(B);       //record information of child
}
}
}
```

Table: Shows the algorithm that node A receives tree construction message from node B [30, p.856].

2. Cluster Head Replacing Scheme

In a network, CH consumes more energy since these heads receives and fuses all the information from its member and child nodes. After particular time, CH should be replaced in order to keep away from CH premature death. This replacement is carried out through by reconstructing that is adopted by GROUP and SCT. But, in this case there is a problem in replacing CH as it consumes more energy. However, ATCBG will not follow this way. ATCBG will just replace CH whenever the elected CH residual energy is above half of the CH energy that needs to be replaced. Whenever the residual energy of a particular CH is below its threshold value then that CH sends a replacing CH message request.

3. Data Transmission

In this process, initially the cluster member nodes send the gathered information to its corresponding CH. Now this CH fuses the information after receiving all the data from member nodes and its child nodes. Finally CH sends the fused data to its parent and this process carry on until the information reaches the sink. In transmission process, based on the actual distance all the nodes adjust its own transmission power [30].

4. Failure Recovery

In case, for certain period if the node doesn't receive any information from its child node then that particular node believes that the child node is failed. Now, this node transmits the information to subsequent grid for competing with the new cluster head.

Result

In this research paper, simulation is carried out through OMNet++. JianShu et al. proved the effectiveness of the ATCBG by comparing with the GROUP. Two metrics are considered to assess the ATCBG performance when compared with the GROUP, which are illustrated below:

- ***Average Energy Consumption*** – JianShu et al. experimented by sending 40 packets in various node densities when the cell size is set to 180 and 200 meters to compare the average energy consumption of GROUP and ATCBG.

Here, in GROUP the tree construction interval is set to 10 packets, 20 packets, 30 packets and 40 packets. In figure (), RI-X means 'X' packets is the reconstruction interval and it corresponds to 10, 20, 30 or 40 packets.

It can be noticed that in ATCBG the average energy consumption of sending the 40 packets is decreased slowly with the increment of node density since the aggregation node is closer to centre of grid with higher node density and no matter the cell size is 180 meters or 200 meters.

- ***Packets sent before emergence of first node death*** – The death of first node relates with the nodes that are distributed in the cluster and its algorithm. For instance, consider a network where it has two clusters 'A' and 'B' as shown.

In this case, the nodes present in cluster 'A' will die earlier when compare with the CH in cluster 'B'. It is due to presence of more number of candidate nodes and can be chosen as CH by replacing CH in cluster 'B'. In particular network area, all the nodes are randomly distributed and the time of first node died will vary with various node densities that will cause different node distribution. In [30] experiment the reconstruction interval in GROUP is set to 10 packets, 20 packets, 30 packets and 40 packets. It can be seen that in ATCBG the minimum number of packets sent before first node death is larger when compared with maximum of GROUP.

3.6.3. Comparison of GROUP and ATCBG Protocols

No.	Protocol	Proposed by	Description	Pros	Cons
1.	GROUP	Liyang Yu, Neng Wang, Wei Zhang and Chunlei Zheng [42]	To offer scalable and efficient packet routing for large-scale WSN	Share out the energy load among sensor in the network and offer in-network processing	Periodically aggregation tree is reconstructed Cluster head selection is carried out based on the distance of grid
2.	ATCBG	JianShu et al. [30]	Proposed by considering the sink as the center of grid.	1. Cluster head selection considers energy and distance. 2. Cluster head with energy below half of the energy required will be replaced.	Construction of tree is done only based on energy

Table 4: Shows the comparison of Grid based routing protocols [42] and [30].

4. Survey 2: Secure Data Aggregation

Type of Attacks and need for Study on Secure Data Aggregation

Data aggregation technique can successfully enables the network in minimizing the data transmission and energy consumption only when the data aggregation process carried out in a secure manner. So, it's very essential to provide security to the network while implementing the data aggregation process so that one can receive the original information from data owner within a short span of time. Based on this aspect, numerous researchers examined and proposed various new techniques to provide security to data aggregation process in their own way. Now, let us discuss some of the researcher papers which provide better security to data aggregation process. Before discussing the papers, now let us have a glance on the attacks that mostly occurred in WSN aggregation due to lack of security.

Effect on Network due to lack of security in WSN Aggregation

According to Roosta et al. (2006), Wireless Sensor Networks are easily vulnerable to various types of attacks because of its nature of transmission medium, hostile and remote deployment location and lack of physical security in every node [55]. Some of the common attacks which frequently affect the data aggregation in WSN are illustrated below:

- Denial of Service (DoS) Attack
- Sybil Attack
- Selective Forwarding Attack
- Node Compromise
- Stealthy Attack
- Replay Attack

Denial of Service (DoS) Attack

DoS definition mainly includes three components which are authorized users, shared service and maximum waiting time [2]. On WSN, this attack is referred as standard by transmitting the radio signals which interfere with the radio frequencies that are used by WSN. Sometimes this is referred as jamming. This attack affects the larger

portions of network as the adversary capability increases. Particularly, in case of data aggregation this attack acts as an aggregator and further refuses to aggregate and prevents the information from transferring the data to higher levels.

Sybil Attack

According to [58], when a node illegitimately asserts multiple identities fake IDs, then the network suffers from an attack known as Sybil attack. The node photocopies itself to make numerous copies in order to collapse and confuse the network. This attack goes against the one-to-one mapping and creates a multiple identities.

Selective Forwarding Attack

In this attack, the malicious nodes refuse to forward some messages and just simply drop them and further it make sure that it doesn't propagate further. This attack is referred as the most effective whenever the attacker is explicitly included on path of data flow [4].

Node Compromise

In data aggregation, once the node has been taken over all the secret data stored on it can be easily extracted.

Stealthy Attack

The aim of this attack is to inject the false information into the network without revealing its existence. This leads to false aggregation result.

Replay Attack

This is one of the most common attacks where the malicious node resends the previous sent packets [60]. It means that the attacker records some traffic from network and just replays them in order to mislead the aggregator.

Thus due to all these attacks it’s very essential to provide security for data aggregation process in wireless sensor networks.

The possible types of attacks in the wireless sensor network is mentioned above and in what way a particular attack affects the security of a wireless sensor network is discussed.Since our second part of the survey is about the secure data aggregation protocols in wireless sensor networks, I mentioned the possible attacks and the need to provide better security for data aggregation in wireless sensor network. So, we are just mentioning the solutions to the attacks and discussed a few protocols in the following report.

Type of Attack	Solution	Reference
Denial of Service Attack	Message Observation Mechanism (MOM)	Reference is [77]
Sybil Attack	False Data Detection Algorithm, WDA witness-based data aggregation.	Described in 4.2 of this report and reference [50] Described in 4.5 of this report and reference is [14]
Selective Forwarding Attack	Support Vector Machines (SVMs)	Reference is [78]
Node Compromise	SDAP, A novel security strategy with assistant cluster heads, SSACH	Described in 4.3 of this report and reference is [71] Reference is [79]
Stealthy Attack	Secure Information Aggregation (SIA)	Described in 4.4 of this report and reference is [18]
Replay Attack	ESPDA (Energy efficient Secure Pattern based Data Aggregation) protocol	Described in 4.7 of this report and reference is [44]

Following are proposed secure data aggregation protocols in wireless sensor network by different researchers.

4.1 Fuzzy Based Secure Data Aggregation Technique in Wireless Sensor Networks

Approach

Fuzzy based secure data aggregation technique was proposed in this research paper where it has three phases, which are illustrated below:

Phase 1: Clustering

In the first phase, all the sensor nodes are grouped into different clusters and each of the clusters has one elected Ch. Further this particular CH will identify the distance among the each member and itself by swapping the topology discovery packets. Thus, in this phase it executes clustering and cluster head election process.

Phase 2:

Here, the CHs gather the information from its members. Along with the information each of the members attaches its current power level. Now, CH identifies the trust level of each node by identifying the correctness of information. This was estimated through spatial and temporal changes that are the difference in two consecutive values

and difference in reading the neighbor sensors. In this phase, within each cluster the power consumed, distance and trust values are computed for each member which is represented below [24]:

1. Distance Estimation

For transferring k-bit message over a distance, d the energy consumed is computed by the radio model which is shown in equation (1)

$$E_T = E_{TX}(k,d) = E_{TX-elec}(k) + E_{TX-amp}(k,d)$$
$$= \{kE_{elec} + k\varepsilon_{fs}d^2, (d < d_0)\}$$
$$= \{kE_{elec} + k\varepsilon_{amp}d^4, (d \geq d_0)\} \quad \text{.....(1)}$$

where, E_{elec} = transmitter circuitry dissipation per bit

The receiving cost is calculated by making use of below equation:

$E_R = E_{RX}(k) = E_{RX-elec}(k) = kE_{elec}$ .....(2)

The mathematical model to increase the lifetime of network and reduce the network energy cost in WSN is defined in equation (3).

Min(E_{Total}) .....(3)

$E_{Total} = E_T + E_R + E_I + E_S$ .....(4)

Where,

E_{Total} = total energy cost in the network

E_T = the transmission cost

E_R = the receiving cost

E_I= the energy cost while being in idle state

E_S = the energy cost while sensing

Total Energy Cost

Cost is computed based on transmission cost, so the new equation for computing total energy cost is given in equation (5):

Min(E_T) .....(5)

In wireless model, depending on the cost of transmission equation (5) can be represented as in the form of equation (6)

$$Min(E_{TX}(k,d)) = E_{TX-elec}(k) + E_{TX-amp}(k,d)$$
$$= \{kE_{elec} + k\varepsilon_{fs}d^2, (d < d_0)\}$$
$$= \{kE_{elec} + k\varepsilon_{amp}d^4, (d \geq d_0)\} \quad \text{.....(6)}$$

Where,

k = the number of bits that are forwarding on the distance d

E_{elec} = transmitter circuitry dissipation per bit

ε= transmit amplifier dissipation peer bit

The critical effect of d on the energy cost of network is shown in the equation (6) so the system model can be represented as

$Min\ (d^n)$.....(7)

Where, n = set to 2 or 4.

When n is set to 2 then equation (7) can be given as $Min\ (d^2)$. d_{NtoCH}represents the distance among the node and CH. Thus it is further reduced to

$Min(d_{NtoCH}^2)$.....(8)

Trust Evaluation

Values of trust are computed in order to test the consistency of sensor nodes.

Spatial Values

It is the average of difference among the distances of the nodes in cluster and the obtained value is compared with the threshold value.

Temporal values

Each of the sensor nodes will compare its present reading with the previous reading and an average of the readings of all sensors are computed and then compared with the threshold value.

If both the spatial and temporal values are greater than its own threshold value then that particular nodes are referred as inconsistent and in case if both the values are than its threshold value than those nodes are referred as consistent.

Consistency Factor

This factor represents the reliability of sensor node, and is computed as shown in below equation

$CV_i = \frac{CC_{Si}-IC_{Si}}{CC_{Si}+IC_{Si}}$Where, -1$\leq C_i \leq 1$

Where,

CV_i = consistency value of node $i(1 \leq i \leq k)$

CC_{Si} = consistent sensing counts of node i

IC_{Si} = inconsistent sensing counts of node i

Sensing Communication Factor

This factor maintains the data related to communication ratio which is calculated with the below equation

$$SR_i = \frac{SS_i - SF_i}{SS_i + SF_i}$$

Where,

SR_i = the sensing communication value of node $iwhere\ 1 \leq i \leq k$

SS_i = sensing success count of node i

SF_i = sensing failure count of node i

Battery Factor

This factor specifies the remaining lifetime of sensor node in network. The Combined Trust Value (CTV) of the node i is computed as follows:

$$CTV_i = \frac{W_1 B_i + W_2 SR_i + W_3 CV_i}{\sum_{i=1}^{3} W_i}$$

Where,

B_i = battery value of node i where $1 \leq i \leq k$

W_i = weight that represents the importance of a particular factor from 0 to +1

Power Estimation

Battery value represents the power in the nodes, each of the sensor node broadcasts quantification value of its own B_i

$$B_i : -1 \leq B_i \leq 1$$

Phase 3:

In this phase, [24] make use of fuzzy logic to choose the best nodes for aggregation. The parameters power level, trust level, and distance to the CH of each of the node are considered as input and fuzzy rules are formed. After the rules are applied, the output will be treated in three different forms that is best node or normal node or worst node. Now the CHs try to aggregate only the packets which are best and normal node by rejecting the worst node. Finally the aggregated data from CHs was transmitted to sink. The three parameters which are represented in fuzzy set are represented below:

Parameter 1: Distance, D = Fuzzy Set [{BN, a}, {NN, b}, {WN, c}]

Where,

a = the membership grade for Best node in Distance calculation

b = membership grade for normal node in Distance calculation

c = membership grade for Worst node in Distance calculation

Parameter 2: Power consumed, P = FuzzySet[{BN, e}, {NN, f}, {WN, g}]

Where,

e = membership grade for Best Node in the calculation of power consumption

f = membership grade for Normal node in the calculation of power consumption

g = membership grade for Worst node in the calculation of power consumption

Parameter 3: Trust, T = FuzzySet[{BN, u}, {NN, v}, {WN,w}]

Where,

u = membership grade for Best Node in trust calculation

v = membership grade for Normal node in trust calculation

w = membership grade for Normal node in trust calculation

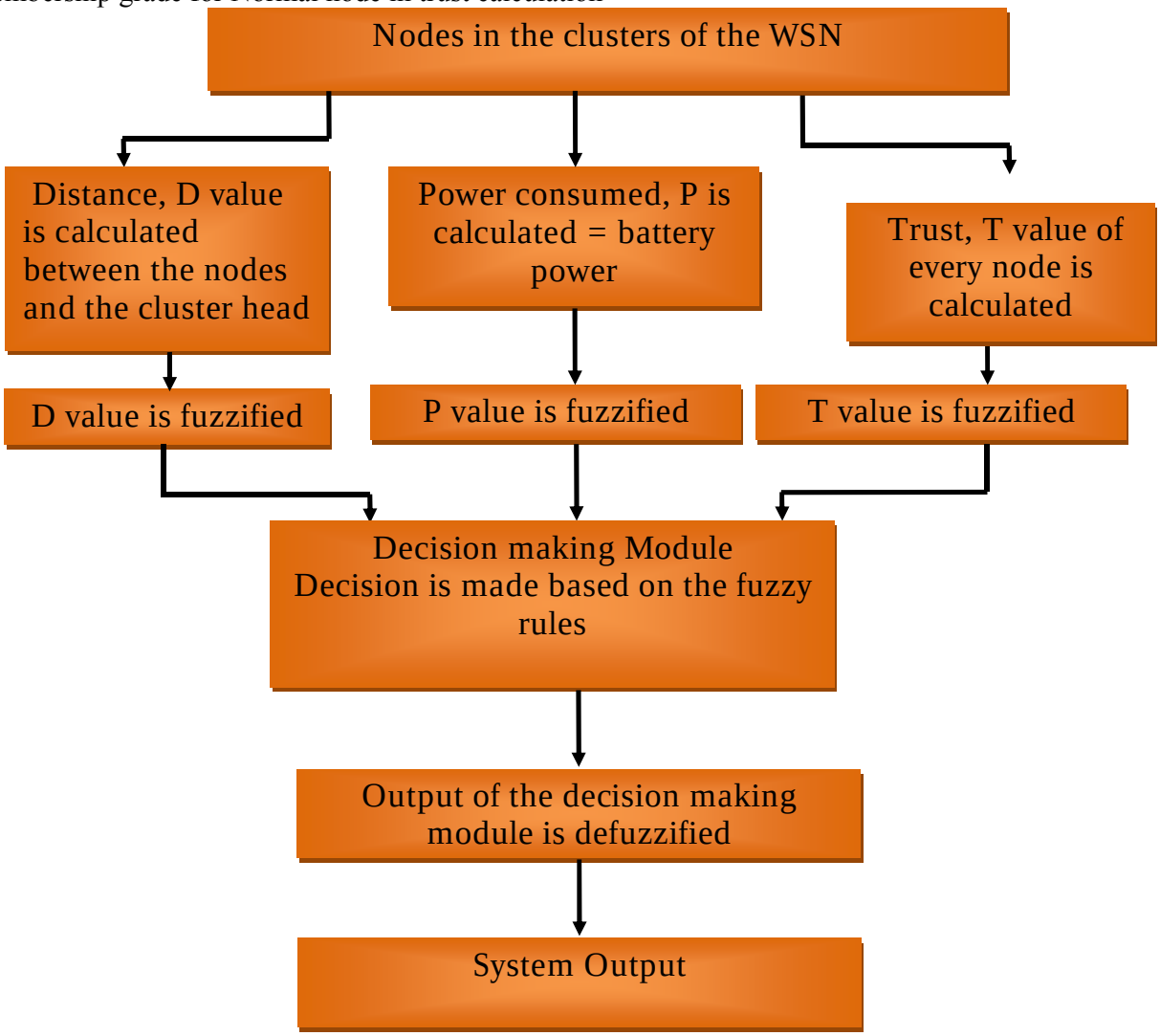

Figure: Shows the Decision making using fuzzy logic [24, p.904].

Distance, D	Power consumed, P	Trust, T	Result
Less	Less	High	Best
Less	High	High	Normal
High	Less	Less	Normal
Less	Less	Less	Normal
Less	High	Less	Worst
High	Less	Less	Worst

High	High	High	Worst
High	High	Low	Worst

Table 1: Shows the Fuzzy rules [24, p.904].

Thus, depending on these parameters the proposed technique chooses the secure and non-faulty node members for data aggregation. As the fuzzy decision rule is made based on the power level of node and trust, [24] approach is secured and power efficient. Apart from it, this approach doesn't involve any cryptographic operations and results in less overhead.

Result

In order to show the effectiveness of fuzzy based secure data aggregation, this technique is compared with the performance of Power-Efficient Secure Routing Protocol (PESRP) in terms of four factors, which are [24]:

- Average Packet Delivery Ratio
- Throughput
- Drop
- Energy

The performance of both these techniques is evaluated through NS2 simulator. A random network is organized in an area of 500x500 m and 30 sensor nodes are placed in square grid area by inserting each sensor in a 50x50 grid cell. The results of the proposed technique are categorized into two parts, which are:

Based on Rate

In the first experiment, the rate was varied from 50, 100, 150, 200 and 250 Kb.

From the above two figures whenever the rate is increased it can be seen that the proposed FBSDA protocol attains good delivery ratio and lower packet drop when compare to PESRP.

When the rate is increased the proposed technique FBSDA protocol has received more number of packets and less energy consumption than PESRP.

Based on Sources

In the second experiment, the traffic flows as 1, 2, 3, and 4.

When number of sources is increased FBSDA achieves a good delivery ratio and lower packet drop when compare to PESRP.

When number of sources is increased FBSDA protocol has received more number of packets and less energy consumption than PESRP.

4.2 False Data Detection in Wireless Sensor Network with Secure Communication

Approach

[50] Introduces a Data Aggregation and Authentication (DAA) protocol to offer false data detection and secure data aggregation against up to T compromised sensor nodes for T≥1. This introduced protocol is classified into three groups which are illustrated below [50]:

- Monitor Node Selection (MNS)
- Forming pairs of sensor node
- Secure data arrogation and false detection

In DAA, one needs to choose the current data aggregator, forward data aggregator and backward data aggregator.

Monitor Node Selection (MNS)

Here, each of the data aggregator is monitored by only T neighboring nodes in total of 'n' neighboring nodes where n≥T. Thus in the initial phase of DAA, T neighbours of data aggregator are chosen as monitoring nodes to execute the data aggregation and to calculate subMACs (Message Authentication Codes) of aggregated data. The monitoring nodes are chosen by the MNS algorithm. Here both the data aggregator and all neighboring nodes are involved with in choosing the monitoring nodes in order to reduce the adverse impact of compromised node.

Forming pairs of sensor nodes

Here, the proposed protocol assumes that there exists a path among the two consecutive data aggregators through forwarding nodes and each of the data aggregator make use of only one outgoing path towards base station with the given time. To set up a pair among the monitoring nodes and forwarding nodes, Af (forwarding node) sends a pairmate discovery message to Au (neighbouring node) along with its neighbouring node list.

Secure data arrogation and false detection

This phase make use of SDFC algorithm to provide false data detection, data confidentiality and secure data aggregation. In order to offer data confidentiality the transmitted information are always encrypted and further the forwarding nodes executes the data verification over encrypted data. To detect the data integrity and

false data injection one pair mate calculates subMAC and other pair mate checks the subMAC. SubMACs are calculated for both plain and encrypted data. Here, subMACs of plain information are utilized to identify the false data injections at the time of data forwarding and subMACs of encrypted data is used to identify the false data injection during data forwarding [50].

Thus, in this way the proposed protocol identifies the false data injection at the time of data forwarding and data aggregation and further secures the data transferred between sink and its destination.

Result

The proposed technique shows better performances when compare to the existing techniques. Since the known techniques do not support data confidentiality and data aggregation although these are essential for WSN whereas in case of proposed DAA protocol which integrates data aggregation, false data detection and data confidentiality.

4.3Secure hop-by-hop Data Aggregation Approach Protocol (SDAP)

Yi Yang et al. [71] proposed SDAP for answering a fundamental challenge that is how the base station acquires a good approximation of fusion result when a fraction of sensor nodes are compromised. The SDAP was designed based on the principle of commit-and-attest and divide-and-conquer. By making use of divide-and-conquer, [71] partitioned the aggregation tree into groups to decrease the importance of high-level nodes in aggregation tree. [71] Uses commit-and-attest with an intention that BS has a way to recognize the aggregates.

Initially this protocol makes use of novel probabilistic grouping technique for partitioning the nodes in a tree topology into many sub-trees of same size. A commitment based hop-by-hop aggregation is executed in every group to produce a group aggregated result. Then the BS recognizes the suspicious group based on the set of group aggregated results. Lastly, every group under suspect participates in an attestation process to prove correctness of its group aggregate. This protocol consists of three phases which are query dissemination, data aggregation and attestation.

For further enriching the protocol, [71] considered the content-based attestation and breadth-based attestation can be included in protocol.

4.4Secure Information Aggregation (SIA)

SIA is a protocol for sensor networks where a part of sensor nodes may be compromised [18]. SIA provides an efficient sub-protocol for securely calculating the median and average of measurements, assessing the network size and identifying the maximum and minimum sensor readings. It includes mainly three types of network components, which are off-site home server (or user), BS (or aggregator) and numerous sensors. [18] Stated that this protocol has a resistance against the stealthy attacks where in this attack the goal of the attacker is to make the user to accept the false aggregation results without revealing its presence [18]. [59] Argues that although this scheme provides secure data authentication but still it delivers the data in plaintext format where it provides no privacy during its transmission from one system to another. [20] Explains that this scheme was proposed under the assumption of a single-aggregator model since in this mode the sensor nodes send their information only to a single aggregator node where it calculates the aggregate and further sends to BS. This process will just decrease the communication link between the aggregator and BS. Further [20] states that this scheme is not scalable for multi-hop sensor deployments.

4.5 Witness-based Data Aggregation (WDA)

Du, W. et al. [14] introduced WDA scheme for wireless sensor networks to guarantee the validation of the information sent from data fusion nodes to BS. In order to prove the fusion results validity, the fusion node needs to give proofs from numerous witnesses. Here the witness refers to a node which also conducts data fusion as a data fusion node however it doesn't forward its results to BS. Now each of the witness calculates the MAC (Message Authentication Code) of result and further sends it to the aggregator node where it should forward the proofs to BS. This protocol provides only integrity property to secure the data aggregation and it is needed to send the numerous copies which are similar to its original aggregated result, to aggregator the point. Finally the aggregator point should forward these collected reports to BS along with the aggregated result. As the point of aggregator is fixed and is responsible to handle a lot of traffic [19].

4.6 Secure Aggregation Tree (SAT)

Traditional solutions it incurs heavy computational overheads although in the network it doesn't include any misbehaving nodes or attackers. So, the computational overhead waste energy and further reduces the available CPU resources for data processing. To overcome from this issue [38] proposed SAT by considering some topological constraints to facilitate the monitoring of the behaviour of each aggregation sensor node. To build the SAT [38] used distributed algorithm and the tree is started from sink node. After building the SAT, each of the nodes must record some of the information which is defined below:

- ID's of both one-hop and two-hop neighbouring nodes
- ID's of its father node in the aggregation tree
- ID's of its sibling nodes in aggregation tree

SAT is very easy method to observe the behaviour of aggregation nodes without resorting to persistent data authentication when compared to traditional solutions. Apart from it, even SAT is essentially different from other existing solutions as it doesn't require any cryptographic operations whenever all the sensor nodes work honestly and in case if in a network cheating is carried out then it is detected based on the topological constraints in aggregation tree.

4.7 Energy-efficient Secure Pattern-based Data Aggregation (ESPDA)

Hasan Cam [44] introduced ESPDA for performing both data aggregation and security for the cluster based WSN. Here, the pattern code is utilized to execute the data aggregation. These pattern codes are considered as specimen data items which are extracted from actual information and these codes are chosen such that these include certain characteristics of corresponding actual information. Based on the type of actual information the extraction process may differ. The pattern codes of sensor comprising multiple sensing units are acquired by combining the individual units. In this protocol, the sensor nodes will not completely transmit the complete sensed information. Initially sensor nodes produce the pattern codes and then send the pattern codes to CH. The distinct pattern codes are identified by CH. Further, CH request only one sensor node to transmit the actual information for each distinct pattern. [29]

Shortly explains that this protocol is applicable for all hierarchical WSNs. Here, in this scheme a designated CH sends requests to sensor nodes in order to send the pattern codes of the information that is being sensed. In case if numerous sensor nodes send the same pattern code to CH then among all only one of the sensor node is permitted to send its information. This protocol is considered as secured as there is no need of any aggregator node to decrypt the information for aggregation. Thus, it has a high computational overhead and achieves both energy and bandwidth efficiency.

4.8 Secure Reference-based Data Aggregation (SRDA)

In the year 2004, [56] developed a new data aggregation technique known as SRDA where it sends only the difference occurred among the sensed data and the reference value (known as differential value) instead of sending the raw information. Here, the difference value is considered as the average value of previous sensor readings. In this proposed scheme, in a sensor field each of the sensor node calculates the differential data then it encrypts the data and further sends the encrypted information to Ch. [56] state that the security level of network must be gradually raised as the information passes to higher level CH. In order to acquire the different levels of security in wireless sensor networks, [56] suggest making use of cryptographic algorithm (RC6) with the adjustable parameters like number of rounds. Here, by increasing or decreasing the number of rounds the security strength of RC6 changes and even it can be measured with the help of security margin. Security margin is nothing but a deviation of the actual number of rounds from minimum number of rounds by which the algorithm is referred to be secured. Hence, this proposed scheme SRDA make use of higher security margin at higher level CHs when compared to low level CHs.

[19] While discussing SRDA in their research paper [19] stated that the protocol designers argued about the intercepting messages which are transmitted at higher levels of clustering hierarchy will produce a summary of large number of transmission at lower levels. From this observation, designers came to a conclusion that the security level must be gradually increased whenever the messages are transmitted to higher levels. So, due to this reasons a cryptographic algorithm is used by [56] to change the security strength.

4.9 Encryption

Basically, for secure data aggregation in WSN two methods are used which are illustrated below [72]:

- Hop-by-Hop encrypted data aggregation
- end-to-end encrypted data aggregation

In hop-by-hop encrypted data aggregation the encryption of the information is carried out by the sensing nodes and the decryption is carried out by the aggregator nodes. Here, the aggregator nodes aggregate the information among and again it will encrypt the aggregation result. Finally, the sink node which acquires the last encrypted aggregation result will decrypts the information. On other hand, in case of end-to-end encrypted data aggregation the aggregator nodes in between will not include any decryption keys and can only perform the aggregation on encrypted data.

In earlier cases, the information is encrypted by sensing nodes and is decrypted by aggregator nodes. The aggregator nodes then aggregate the information and again it encrypt the result of aggregation. Finally, the sink node obtains the ultimate encryption aggregation result and decrypts it. In later versions, the intermediate aggregator nodes doesn't have any decryption keys and can carry out only aggregations on encrypted data. While carrying out research on this aspect [15] identified the two practical issues that are involved in implementing the data encryption at sensors. The first issue is that the size of the encrypted message and the second issue is about the execution time for encryption at sensors. PH (Privacy Homomorphisms) is an encryption function that allows a set of operations that need to execute on encrypted data without having knowledge on decryption functions. Basically, PH is used to assess the feasibility of the implementation of security in sensors.

4.10 Concealed Data Aggregation (CDA)

[16] Proposed a scheme called as CDA (Concealed Data Aggregation) based on PH (Privacy Homomorphism) that is proposed in [13]. This proposed scheme can be utilized to compute SUM and AVERAGE in a hierarchical WSN. To compute AVERAGE the aggregator requires the number of sensor nodes '*n*'. CDA provides

End-to-end encrypted data aggregation by making use of privacy homomorphism encryption. In WSN, the sensor nodes share a secret key from sink and this is not

visible to any intermediate aggregator. This proposed scheme supports the end-to-end encrypted data aggregation for reverse multicast traffic among the sensors and sinks. Further, in this scheme the aggregator execute the aggregation without decrypting the information and hence reduces the delay that is caused by decryption/encryption process [16]. CDA make use of additive and multiplicative homomorphic encryption scheme which permits the aggregator to aggregate the encrypted information. Here the author argues that the security level is still reasonable and PH helps to implement the encryption in WSN, even though the [67] prove that PH is unsecure against selected plain text attacks. However, author stated that the encryption in CDA is very expensive and further adds among 0% to 22% additional information overhead when compare to RC5 that raises the consumption of power of a sending node. Thus, CDA ensures only data confidentiality.

4.11 Secure Hierarchical in-network Data Aggregation (SHDA)

This scheme was proposed by Chan et al. [76] and was published in 2006. SHDA enhances SIA by expanding into a distributed network that usually includes more than one aggregator. The aim of this scheme is that to prevent a battery exhaustion of heaviest-loaded nodes which are most probably close to BS and further regards all the network nodes as sensor nodes. [9] Explains that this scheme provides exactly what SIA used to do that is data integrity, authentication, and confidentiality. Here in the proposed scheme each of the parent sensor carry out an aggregation function whenever it gets a message from its child nodes. Additionally, it should generate a commitment to the input set which is used to calculate the aggregated result by making use of merkle hash tree. Now, it forwards the aggregated information and the commitment to its parent node until it reaches its BS. Once BS receives the final commitment values, further it shows them to rest of network in an authenticated broadcast. Here every node is responsible for verifying whether its contribution was added to aggregation information or not. After adding its reading, it sends an authentication code for node. In order to provide better communication efficiency, the authentication codes are aggregated along the way to BS. So, in case if any one authentication code is missing then simply BS will reject the aggregated result.

4.12 Privacy-preserving Data Aggregation (PDA) in Wireless Sensor Networks

Now-a-days, one of the major problems in WSN is providing efficient data aggregation while preserving the data privacy. [69] Proposed two privacy-preserving data aggregation schemes for providing additive aggregation functions. The first scheme proposed is Cluster-based Private Data Aggregation (CPDA) and second scheme is Slice-Mix-AggRegaTe (SMART). The main aim of the [69] research is to bridge the gap between the collaborative data collection by WSN and data privacy. Whenever there is no packet loss in both the schemes then the sensor network can acquire a precise aggregation result while assuring that no private sensor reading is released to other sensors. Now let us discuss the two proposed scheme briefly:

Cluster-based Private Data Aggregation (CPDA)

In this scheme, CPDA includes three phases which are discussed below:

- ***Cluster formation*** - to perform intermediate aggregations CPDA initially construct clusters. A distributed protocol have been used by [69] for attaining this purpose. Firstly, the query server triggers the query by sending the message. After receiving the message the sensor node itself choose a cluster leader with some probability that is preselected parameter for all nodes. In case if a node becomes cluster leader then it forward its message to its neighbour nodes or else it will wait for the particular period of time to attain the message from its neighbours and then it make a decision to join any one of the clusters by broadcasting join message. In this way multiple cluster are constructed in this phase.
- ***Computing aggregate results within clusters*** -the aim of this phase is to make intermediate aggregations within clusters.
- ***Cluster data aggregation*** - one of the common techniques for data aggregation is to build the routing tree. [69] Implemented CPDA on top of TAG protocol. Here each of the cluster leader routes the derived sum within the cluster back towards the query server via TAG routing tree which is rooted at the server.

Slice-Mix-AggRegaTe (SMART)

One of the main limitations of cluster based protocol is computational overhead of data aggregation within the clusters. So, [69] proposed a new scheme known as SMART that decreases the computational overhead at cost of slightly raised

communication bandwidth consumption. This scheme includes three steps which are slicing, mixing and aggregation. In this scheme, each of the nodeshides its private information by slicing the data and further sending the encrypted information slices into different aggregators. Now the a-ggregators gather and forward the information to a query server. Once the server receives the aggregated information and further it computes the final aggregation result.

Hence, both of the proposed schemes results are assessed by data aggregation accuracy, privacy-preservation efficacy and communication overhead. For future work, [69] included designing the private preserving data aggregation schemes for general aggregation functions. Further [69] even investigating on robust private-preserving data aggregation schemes under the malicious attacks.

4.13 Dynamic and Scalable Routing to Perform Efficient Data Aggregation in WSN

In WSN, data aggregation is considered as one of the main methods to conserve energy. The aggregation process will reduce the redundant data that is being passed through intermediate nodes by decreasing the number of messages exchanged and further reduces the communication cost almost all the data aggregation protocols are proposed based on static routing scheme even though these protocols save energy by removing the data redundancy. However, in dynamic scenarios these protocols can incur high overhead to further reconstruct the routing tree. So, [41] research consider the problem of constructing a dynamic and scalable structure for data aggregation in wireless sensor networks. To overcome from this issue [41] proposed a novel routing protocol known as Dynamic and Scalable Tree (DST). This proposed protocol can be easily adapted in different scenarios without incurring the overhead of other schemes. One of the main reason behind introducing DST is that in event-based applications the source nodes are not known in advance and frequently change, hence the data aggregation approaches supported by static routing schemes cannot efficiently aggregate the information. Here [41] proposed three different variations of DST algorithm where each variation defines a new approach for coordinator nodes for creating straight lines are illustrated below [41]:

- ***DST-BC (Dynamic and Scalable Tree-Best Combination)*** - here firstly it verifies all possible combinations of straight lines and selects the

combination which offers the shortest Euclidean distance for creating the routing tree. This approach is optimal as it identifies the routing tree at low cost.

- ***DST-CF (Dynamic and Scalable Tree - Closest First)*** - here the closest coordinator to sink node is initially created the straight lines then the second closest coordinator create a straight line. This process continues until all the coordinator create their straight lines.
- ***DST-FF (Dynamic and Scalable Tree-Farthest First)*** - in this case the coordinator which is farthest to sink is initially created the straight line. Then the second farthest create straight line and these steps are repeated until all coordinators create their straight line segments.

The main aim of creating the routing tree is for coordinator, initially among the three approaches only one scheme is being selected to create its straight line segment to sink. Now the second coordinator selected by DST approach creates its straight line segment to its nearest point which already exists. All these steps are repeated until all coordinator create their straight line segments. Whenever the coordinator nodes wants to compute its straight line segment, then it make use of data acquired in phases 1-3. Since the order in which the coordinator create their straight line depends on DST approach employed. If a coordinator node knows that the straight line segments are created by other coordinator then it will create its own straight line segment to its nearest point of straight line segment of another coordinator. Whenever the coordinator node transmits the message then the closest nodes to both its straight line segment computed before and to endpoint of the straight line will be selected in order to forward the message.

Thus, DST improves the number of overlapping routes and chooses the routes by considering the highest aggregation rate. Simulation results of [41] proved that the routing tree built by DST offers the best efficiency when compared to other algorithms and outperformed them for all the scenarios.

4.14 An Efficient Data Aggregation Scheme Using Degree of Dependence on Clusters in WSNs

[63] Stated that one of the most significant challenges in WSN is to decrease the consumption of node's energy as these nodes run on scarce battery resources. In past,

various authors explained that by exploiting the sink node's mobility to decrease the energy consumption however even this scheme has some challenges to sink node routing and information aggregation. Based on this aspect, [63] proposed a new routing and data aggregation scheme based on clustering.

In the proposed scheme, [63] assumed that the nodes in sensing area are grouped into K clusters by making use of EM algorithm and the mobile sink traces the trajectory of TSP via these cluster centroids. In order to aggregate the data efficiently the mobile sink and nodes utilizes cluster adapted Directed Diffusion. The proposed data aggregation scheme based on clustering is summarized as follows:

Clustering Algorithm - in WSN to decrease the energy consumption one of the important issue is transmission distance as the required power of wireless transmission is proportional to square of transmission distance. The EM algorithm enables to decrease the sum of squares of distances among nodes and cluster centroids. Additionally, the EM algorithm is constructed under less-strict assumptions when compare to K-means algorithm. Thus, [63] make use of EM algorithm to group WSN nodes into K clusters to decrease the consumption of energy. The two-dimensional EM algorithm is only based on an assumption in which the nodes are distributed according to two-dimensional Gaussian distribution.

Trajectory of Mobile Sink - [63] identify the actual trajectory of mobile sink node after clustering the WSN nodes. The mobile sink goes across the clusters and aggregates the information from numerous nodes. Since it is possible to raise the efficiency by shortening the traveling time, it is preferable that the mobile sink traces the shortest path between cluster centroids. Thus, [63] make use of TSP solution as the trajectory.

Cluster adapted Directed Diffusion - after dividing the WSN nodes into K clusters based on GMM and identifies the trajectory of mobile sink. Now [63] addresses the problems of aggregating the sensed information gathered from WSN nodes. For this purpose, [63] considers the directed diffusion approach under the already stated assumption pertaining to the nodes distribution.

Thus, from the [63] simulation results it can be stated that the proposed scheme provide better energy efficient data aggregation than KAT mobility.

5. Conclusion and Future Work

5.1 Challenges in Data Aggregation

MAC layer- Data aggregation in wireless sensor networks will transmit less data by eliminating redundant data flows and reduce the total energy consumption.

Collisions occur when sensor nodes that are within each other's transmission range try to transmit simultaneously. When collision occurs, retransmissionis required to ensure the data is successfully received.

However, additional data collisions incur extra data retransmissions. These data retransmissions not only increase the energy consumption, but also increase the delays. The decision of when and where to aggregate data depends on the trade-off between data aggregation and data retransmission. The challenges of this problem need to address the routing (layer 3) and the MAC layer retransmissions (layer 2) at the same time to identify energy-efficient data-aggregation routing assignments. [83]

Topology changes–After deployment of nodes, topology changes occur due tochange in sensor nodes

• Position,

• Reachability (due to jamming, noise, moving obstacles,etc.),

• Available energy and

• Malfunctioning

Device failure is a regular or common event due to energy depletion or destruction. It isalso possible to have sensor networks with highly mobile nodes.

Algorithms for generating collision-free schedules are centralized which require the sink to compute the schedule and disseminate it to the sensors. Once all the sensor nodes receive the schedule, they work according to the schedule. Since topology changes often occur in sensor networks such as node failures, the sink has to gather new topology information from the network, re compute a schedule and disseminate it frequently. These processes consume lots of energy, which makes centralized algorithms inefficient.

Distributed aggregation scheduling algorithm is proposed by [84] which is an adaptive method suitable for network with topology changes.

5.2 Future Work

Industrial standards for wireless sensor networks are discussed in the introduction chapter; which are Wireless Hart, ISA-100.11A and WIA-PA.

Wireless-Hart didn't allow data aggregation because of security. The other two industrial standards allow data aggregation in the industrial and process automation applications.

Future work can be carried out on implementing packet aggregation in wireless- Hart and thus prolonging the wireless-Hart network lifetime.

Most of the existing research literatures construct the aggregation tree by only taking into consideration of data aggregation aspect. However, there is one moreissue that is important to the construction of data aggregationtree, *MAC layer retransmission* issue, which is discussed in the challenges in data aggregation above. So, Future work can be carried out on improving the aggregation protocols by considering the MAC layer retransmission issue.

Centralized data aggregation algorithms are not efficient in a network with higher topology changes as explained in the challenges above. Future work can be carried out by proposing new protocols to address higher topology changesand higher scalability.

5.3 Conclusion

WSN includes large number of sensor nodes which transfer the data from one system to another system without making use of any wires. All these sensor nodes in the network are resource constraint, so because of this reason the lifetime of the network is limited. Thus, various researchers proposed numerous protocols or approaches for increasing the lifetime of the wireless sensor networks. In this means, the data aggregation concept has been introduced in this report as it is one of the important techniques that enhances the network lifetime. To get a clear idea on all these techniques, a theoretical survey is conducted in this research by referring numerous secondary resources like journals, authorized pdf's, etc. The data was collected by categorizing the research into two parts. In the first part the data was collected on the

protocols that are being proposed for data aggregation technique. And in the second part, the data was gathered based on the approaches proposed for securing the WSN data aggregation. In this report, various data aggregation algorithms in WSN are discussed. Further a comprehensive study of different data aggregation protocols are presented under the network architecture. The data aggregation algorithms discussed in this report mainly focuses on three concepts which are efficient routing, organization and data aggregation tree construction. This report described the main features, benefits and limitations of different data aggregation algorithm. However after discussing all the data aggregation protocols it can be concluded that the performance of data aggregation protocol is strongly coupled with network infrastructure. Even though many of the data aggregation techniques which are discussed looks promising but still there is significant scope for further research. So, this research further extended the work by making study on secure data aggregation since none of the researcher discussed the approaches which provide security to the data aggregation. It's very essential to provide security to the WSN data aggregation since there are numerous effects on network due to lack of security and moreover one cannot increase the lifetime of network without providing security to WSN. Further of this research clearly discusses the various approaches proposed for securing the data aggregation techniques in WSN. Security can be provided to WSN in various means like false detection, etc. From this study one can easily understand how important to secure the network and to successfully carry out the data aggregation techniques.

6. References

[1] Akyildiz, I. F. et al.(2002) *Wireless Sensor Networks: a survey*. Elsevier Science.

[2] Anthony D. Wood and John A. Stankovic (n.d) http://pdf.aminer.org/000/258/247/a_taxonomy_for_trusted_services.pdf (Acc. 13 December 2012).

[3] AzamAslam and Razwan Ahmed (2009)http://home-networking.wikidot.com/wireless-vs-wired(Acc.18 October 2012).

[4] BhavnaArora Makin and DevAnandPadha (2010) http://www.iupindia.in/910/IJIT_A_Trust_Based_Secure_Data7.pdf (Acc. 13 December 2012).

[5] Boulis, A. et al. (2003) *Aggregation in Sensor Networks: An energy-accuracy trade-off.* Proceedings of the First IEEE International Workshop on Sensor Network Protocols and Applications.

[6] Cam, H. et al. (2004) *Secure Differential Data Aggregation for Wireless Sensor Networks*, Sensor Network Operations. Wiley.

[7] Cam, H., Muthuavinashiappan, D. and Nair, P. (2003) *ESPDA: Energy-Efficient and Secure Pattern-Based Data Aggregation for Wireless Sensor Networks*, Proc. IEEE Sensors. Toronto, Canada.

[8] Castelluccia, C., Mykletun, E. and Tsudik, G. (2005) *Efficient Aggregation of Encrypted Data Wireless Sensor Network*, Proc. ACM/IEEE Mobiquitous, San Diego, CA.

[9] Chan, H., Perrig, A. and Song, D. (2006) *Secure Hierarchical in-network aggregation in sensor networks*, in A. Juels, R. N. Wright and S. D. C. di Vimercati, eds, ACM Conference on Computer and Communications Security, ACM.

[10] Curt Schurgers and Mani B. Srivastava(2001) *Energy Efficient Routing in Wireless Sensor Networks*.MILCOM'01.

[11] Deng, J., Han, R. and Mishra, S. (2003) *Security Support for in_network Processing Wireless Sensor Networks, SASN'03:Proc. 1st ACM Wksp. Security of ad Hoc and Sensor Networks.* New York: ACM Press.

[12] De-Shuang Huang, Kang Li and George William Irwin (2006) *International Conference on Intelligent Computing: Computational*

*Intelligence.*Germany: Springer-VerlagBerlin,ISBN 3540372741, 9783540372745.

[13] Domingo-Ferrer, J. (2002) *A Provably Secure Additive and Multiplicative Privacy Homomorphism*, lecture Notes Comp. Sci., Vol. 2433.

[14] Du, W. et al. (2003) *A Witness-based Approach for Data Fusion Assurance in Wireless Sensor Networks*, in GLOBECOM'03: Proceedings of IEEE Global Telecommunications Conference, San Fransciso.

[15] Duarte-Melo, E. J. and Liu, M. (2003) *Computer Networks:the Int'l. J. Computer and Telecommunication Net.*, Vol. 43.

[16] Girao, J., Westhoff, D. and Schneider, M. (2005) *CDA: Concealed Data Aggregation for Reverse Multicast Traffic Wireless Sensor Networks, ICC'05: Proc.IEEE Int'l. Conf.Commun.* Seoul, Korea.

[17] Gurpree Sigh Chhabra and Dipesh Sharma (2011) *Cluster-Tree based Data Gathering in Wireless Sensor Network.* International Journal of Soft Computing and Engineering, Vol. 1, Issue 1.

[18] Hani Alzaid, Ernest Foo and Juan Gonzalez Neito and DongGook Park (2011) *Secure Data Aggregation in Wireless Sensor Networks*, Emerging Communications for Wireless Sensor Networks, InTech.

[19] Hani Alzaid, Ernest Foo and Juan Gonzalez Nieto (2007) *Secure Data Aggregation in Wireless Sensor Networks: a Survey.* Australian Computer Society.

[20] Haowen Chan, Adrian Perrig and Dawn Song (2006) *Secure Hierarchical In-Network Aggregation in Sensor Networks, CCS'06.* Alexandria, Virginia, USA

[21] HART (2010) http://www.hartcomm.org/protocol/training/resources/wiHART_resources/Security_Overview_LIT114.pdf (Acc.23 April 2012).

[22] Heinzelman, W. R., Kulik, J. and Balakrishnan, H.(1999) http://www.cs.huji.ac.il/labs/danss/sensor/sensors/kulik_99adaptiveprotocols.pdf (Acc.24 April 2012).

[23] Heinzelman, W., Chandrakasan, A. and Balakrishnan, H. (2000) *An Energy-Efficient Communication Protocol for Wireless Microsensor*

Networks. Proceedings of the 33rd Hawaii International Conference on System Sciences. IEEE

[24] Hevin Rajesh, D. and Paramasivan, B. (2012) *Fuzzy Based Secure Data Aggregation Technique in Wireless Sensor Networks*. Journal of Computer Science, Vol.8.

[25] Hu, L. and Evans, D. (2003) *Secure Aggregation for Wireless Networks.*Wksp. Security and Assurance in Ad Hoc Networks.

[26] Ignacio Solis and Katia Obraczka (n.d) citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.75.2779&rep=rep1&type=ps (Acc.18 June 2012).

[27] Information Systems (1996) *Computer World*. The Newspaper of Information Systems Management, Vol. 30, No. 4.

[28] Jason Lester Hill (2003) *System Architecture for Wireless Sensor Networks*.Spring.

[29] JaydipSen (2012) *Secure and Energy-efficient Data Aggregation in Wireless Sensor Networks*, In Proceedings of 2nd National Conference on Computational Intelligence and Signal Processing, IEEE.

[30] JianShu et al. (2010) *Study on Aggregation Tree Construction Based on Grid*. Journal of Computers, Vol. 5, No. 6.

[31] John A. Stankovic (2006) http://www.cs.virginia.edu/~stankovic/psfiles/wsn.pdf (Acc.20 April 2012).

[32] Jorge Tavares, Fernando J. Velez and Joao M. Ferro (2008) *Application of Wireless Sensor Networks to Automobiles*. Measurement Science Review, Volume 8, No. 3.

[33] Jun Zheng and Abbas Jamalipour (2009) *Wireless Sensor Networks: A Networking Perspective*. USA: John Wiley & Sons, ISBN 0470167637, 9780470167632.

[34] Jung, S. M., Han, Y. J. and Chung, T. M. (2007) *The Concentric Clustering Scheme for Efficient Energy Consumption in the PEGASIS*. Proceedings of 9th International Conference on Advanced Communication Technology, Vol 1. IEEE Xplore

[35] KiranMaraiya, kamal Kant and Nitin Gupta (2011) *Wireless Sensor Network: A Review on Data Aggregation*. International Journal of Scientific and Engineering Research, Vol. 2, Issue 4.

[36] Kiranmaraiya, Kamal Kant and Nitin Gupta (2011) *Application based Study on Wireless Sensor Network.* International Journal of Computer Applications, Vol. 21, No.8.

[37] KonstantinosKalpaks, KoustuvDasgupta and ParagNamjoshi (2002) *Maximum Lifetime Data Gathering and Aggregation in Wireless Sensor Networks*. TR CS-02-12.

[38] Kui Wu et al. (2006) *Secure Data Aggregation without Persistent Cryptographic Operations in Wireless Sensor Networks.* IEEE computer Society.

[39] Kumar, D., Aseri, T. C. and Patel, R. B. (2011) *EECDA: Energy Efficient Clustering and Data Aggregation Protocol for Heterogeneous Wireless Sensor Networks.* Int. J. of Computers, Communications and Control, ISSN 1841-9836, Vol. VI, No. 1.

[40] Kuong-Ho Chen, Jyh-Ming Huang and Chieh-Chuan Hsiao (2009) *CHIRON: An Energy Efficient Chain-Based Hierarchical Routing Protocol in Wireless Sensor Networks*. IEEE.

[41] Leandro A. Villas et al. (2011) *Dynamic and Scalable Routing to Perform Efficient Data Aggregation in WSN.* IEEE Communications Society.

[42] Liyang Yu, Neng Wang, Wei Zhang and ChunleiZheng (2006) *GROUP: a Grid-clustering Routing Protocol for Wireless Sensor Networks*. In proceedings of 2006 IEEE International Conference on Wireless Communciations, Networking and Mobile Computing.

[43] Messina, D., Ortolani, M. and Lo Re, G. (2007) *A Network Protocol to Enhance Robustness in Tree-Based WSN's using Data Aggregation*. IEEE International Conference on Mobile Adhoc and Sensor Systems.

[44] Mukesh Kumar Jha and Sharma, T. P. (2011) *Secure Data aggregation in Wireless Sensor Network: A Survey.* International Journal of Engineering Science and Technology, Vol.3.

[45] NahdiaTabassum, QuaziEhsanulKabirMamun and YoshiyoriUrano (2006)*COSEN: A Chain Oriented Sensor Network for Efficient Data Collection.* Proceedings of the Third International Conference on Information Technology: New Generations.

[46] Neander, J.(2011) *Prolonging Wireless HART Network Lifetime Using Packet Aggregation*. Industrial Electronics
(ISIE) IEEE International Symposium.

[47] NisheethShrivastava, ChiranjeebBuragohain and DivyakantAgrawal (2004) *Medians and Beyond: New Aggregation Techniques for Sensor Networks*. USA: SenSys.

[48] Postech (n.d) http://monet.postech.ac.kr/research.html (Acc.20 April 2012).

[49] PrashantKrishnamustry (2009) *Wireless Sensor Networks*. John Wiley & Sons, Ltd

[50] Priyanka S. Fulare and Nikita Chavhan (2011) *False Data Detection in Wireless Sensor Network with Secure Communication.* International Journal of Smart Sensors and Ad Hoc Networks (IJSSAN), Vol.1.

[51] Przydatek, B., Song, D. and Perrig, A. (2003) *SIA: Secure Information Aggregation in Sensor Networks*, SenSys'03: Proc. 1st Int'l. Conf. Embedded Networked Sensor Systems. New York: ACM Press.

[52] Raghavendra, C. S. (2006) *Wireless Sensor Networks.* USA: Springer, ISBN 0-387-5269-4.

[53] Ramesh Rajagopalan and Pramod K. Varshney (2006) *Data Aggregation Techniques in Sensor Networks: A Survey*. Electrical Engineering and Computer Science.

[54] Rocky Dunlap (n.d) *In-Network Aggregation in Wireless Sensor Networks*.

[55] Roosta, T. et al. (2006) *Taxonomy of Security Attacks in Sensor Networks.* IEEE International, Washington, Dc, USA.

[56] Sanli et al. (2004) *SRDA: Secure reference-based data aggregation protocol for wireless sensor networks*, in Vehicular Technology Conference.

[57] Satapathy S. S. and Sarma, N. (2006) *TREEPSI: tree based energy efficient protocol for sensor information*. IFP International Conference.

[58] Sharmila, S. and Umamaheshwari, G. (2012) *Detection of Sybil Attack in Mobile Wireless Sensor Networks.* International Journal of Engineering Science and Advanced Technology, Vol.2.

[59] Shih-I Huang, ShiuhpyngShieh and Tygar, J. D. (2009) *Secure Encrypted-Data Aggregation.* Springer Science+Business Media, LLC.

[60] Steffen Peter et al. (n.d) http://www.ist-ubisecsens.org/publications/cda-ihp.pdf (Acc. 13 December 2012).

[61] Stephanie Lindsey and Cauligi S. Raghavendra (2002) *PEGASIS Power-efficient Gathering in Sensor Information Systems*. In Proceeding of IEEE Aerospace Conference. Montana IEEE Aerospace and Electronic Systems Society.

[62] SunHee Yoon and Cyrus Shahabi (2007) *The Clustered AGgregation (CAG) Technique Leveraging Spatial and Temporal Correlations in Wireless Sensor Networks*. ACM Transactions on Sensor Networks (TOSN), Vol. 3, Issue 1, No. 3.

[63] TetsushiFukabori et al. (2010) *An Efficient Data Aggregation Scheme Using Degree of Dependence on Clusters in WSNs.* IEEE Communication Society.

[64] Thangaraj, M. and PunithaPonmalar, P. (2011) *A Survey on Data Aggregation Techniques in Wireless Sensor Networks*. International Journal of Research and Reviews in Wireless Sensor Networks, Vol. 1, No. 3.

[65] VaibhavPandey, AmarjeetKaur and Narottam Chand (2010) *A Review on Data Aggregation Techniques in Wireless Sensor Network*. Journal of Electronic and Electrical Engineering, Vol.1, Issue 2.

[66] Vijay Erramilli, Ibrahim Matta and AzerBestavros (2004) *On the Interaction between Data Aggregation and Topology Control in Wireless Sensor Networks*.NSF grants ANI-0095988.

[67] Wagner, D. (2003) *Cryptanalysis of an algebraic privacy homomorphism.* Springer.

[68] Wagner, D. (2004) *Resilient Aggregation Sensor Networks, SASN'04: Proc.2nd ACM W ksp. Security of Ad Hoc and Sensor Networks.* New York: ACM Press.

[69] Wenbo He et al. (n.d) http://cairo.cs.uiuc.edu/publications/papers/wenbohe_PDA.pdf (Acc.14 December 2012).

[70] Xiuming Zhu, Wei Dong, Aloysius K. Mok, Deji Chen and Mark Nixon (n.d) http://www.cs.utexas.edu/~wdong86/rtcsa09.pdf (Acc.23 April 2012).

[71] Yi Yang et al. (2006) *SDAP: A Secure Hop-by-Hop Data Aggregation Protocol for Sensor Networks, MobiHoc'06.* ACM, Italy.

[72] Yingpeng Sang et. (2006) *Secure Data Aggregation in Wireless Sensor Networks: A Survey*, A Seventh International Conference on Parallel and Distributed Computing, Applications and Technologies.

[73] Yingshu Li, My T. Thai and Weili Wu (2008) *Wireless Sensor Networks and Applications*. USA: Springer, ISBN 978-0-387-49591-0.

[74] Yong Wang, GarhanAttebury and Byrav Ramamurthy (2006) *A Survey of Security Issues in Wireless Sensor Networks.* CSE Journal Articles.

[75] Yu-CheeTseng et al. (2004) *Location Tracking in a Wireless Sensor Network by Mobile Agents and Its Data Fusion Strategies*. The Computer Journal, Vo. 47, No. 4.

[76] Zybnek Ondrak (2010) http://is.muni.cz/th/139513/fi_m/dp_merged.pdf (Acc. 14 December 2012).

[77] ZHANG Yi-ying, LI Xiang-zhen and LIU Yuan-an (2012) *The Detection and Defence of DoS Attack for Wireless Sensor Network*. Science Direct, the Journal of China Universities of Posts and Telecommunications.

[78] Sophia Kaplantzis et al. (2007) *Detecting Selective Forwarding Attacks in Wireless Sensor Networks using Support Vector Machines.* IEEE.

[79] Zhi-Ting Lin et al. (2006) *Compromised Nodes in Wireless sensor Network*.

[80] Special Report CEA (2011) http://www.field-wireless.com/en/technology/doc/%28Yokogawa%29_Control_Engineering_Asia_Apr_2011.pdf

[81] Kallol Das and Paul Havinga (n.d) http://eprints.eemcs.utwente.nl/22421/01/IoT2012ieeechecked.pdf (Acc. 11 March 2014)

[82] Tang Zhong, (2010) Real time communication in WIA-PA industrial wireless networks, IEEE International Conference.

[83] FrankYeong, Hong-Hsu, Shu-Ping and Yean-Fu (2006)MAC Aware Energy-Efficient Data-Centric Routing in Wireless Sensor Networks. IEEE publication.

[84] Bo Yu, Jianzhong Li and Yingshu Li (2009) Distributed Data Aggregation Scheduling in Wireless Sensor Networks. INFOCOM 2009 IEEE publication.