Efficient Data Aggregation Scheme in Secure Tree based Wireless Sensor Networks (EDAS)

MS Research Thesis

By

Khalid Khan

Reg: # 567-FBAS/MSCS/F09

Supervised By

Prof. Dr. Muhammad Sher

DEAN FBAS and Chairman DCS & SE

Department of Computer Science and Software Engineering

Faculty of Basic & Applied Sciences

International Islamic University Islamabad

2014

MS
004
KHE

Computer science

Computer system

بسم الله الرحمن الرحيم

DEPARTMENT OF COMPUTER SCIENCE AND SOFTWARE ENGINEERING
FACULTY OF BASIC AND APPLIED SCIENCES
INTERNATIONAL ISLAMIC UNIVERSITY ISLAMABAD

Dated: 08-12-2014

Final Approval

It is certified that we have examined the thesis submitted by **Khalid Khan, Reg: # 567-FBAS/MSCS/F09**. It is our judgment that the thesis is of sufficient standard to warrant its acceptance by the International Islamic University Islamabad for the degree of **Master of Science in Computer Science.**

Committee:

External Examiner

Dr. Abid Khan
Assistant Professor
Department of Computer Science
COMSATS Institute of Information Technology,
Islamabad

Internal Examiner

Mr. Sayed Muhammad Saqlain
Assistant Professor
Department of Computer Science
and Software Engineering
International Islamic University,
Islamabad

Supervisor

Prof. Dr. Muhammad Sher
Dean FBAS / Chairman DCS & SE
International Islamic University,
Islamabad

8.12.14

ABSTRACT

Wireless sensor networks have emerged as an important area of research and have gained interest since the last two decades. Wireless sensor network is a collection of large number of low cost resource constrained sensor nodes that communicate using wireless medium. Sensor nodes are resource constrained in memory, computational capability and battery power. One of the solutions to these problems is to reduce the data transmission in wireless sensor networks. Various data aggregation protocols like ESPDA [20], SRDA [21], have been proposed to reduce the data to be transmitted through the network. But they creates security issues like authentication, data integrity and data freshness. Hop-by-hop secure data aggregation protocols are proposed to provide security along with data aggregation function. Data on the aggregators still need to be secured. In hop-by hop aggregation protocols, data on the aggregators is decrypted to apply aggregation function. Here, aggregators are exposed to node compromises. End-to-end secure data aggregation protocols are proposed. They provide end-to-end security and privacy to data. One of these protocols is SEEDA: Secure End-to-End Data Aggregation Protocol [27], which ensured end-to-end privacy of data. But this protocol contains higher communication cost (data) and computational overheads because it sends extra bits/data regarding non-responding nodes and perform unnecessary computations for non-responding nodes. We proposed new protocol EDAS which reduces 12% to 25% communication and computational overheads as compared to SEEDA with ensuring end-to-end privacy of data. EDAS is also efficient in energy consumption.

DECLERATION

It is hereby declared that this work, neither as a whole nor as a part, has been copied out from any source. It is further declared that I have conducted this research and have accomplished this thesis entirely on the basis of my personal efforts and under the sincere guidance of my supervisor Prof. Dr. Muhammad Sher. If any part of this project/thesis is proved to be copied out from any source or found to be reproduced from some other project, I shall stand by the consequences. No portion of the work presented in this dissertation has been submitted in support of any application for any other degree of this or any other university or institute of learning.

KHALID KHAN

567-FBAS/MSCS/F09

A Dissertation submitted to the

Department of Computer Science and Software Engineering

International Islamic University Islamabad

in partial fulfillment of the requirements

for the degree of

Master of Science in Computer Science

2014

KHALID KHAN

567-FBAS/MSCS/F09

DEDICATION

THIS THESIS

IS DEDICATED TO MY

FATHER WHO TAUGHT ME

THAT HOW DIFFICULT THE TASK MAY

BE, ALWAYS TRY TO SOLVE IT YOURSELF. IT IS

ALSO DEDICATED TO MY MOTHER, WHO

TAUGHT ME THAT EVEN THE LARGEST

TASK CAN BE ACCOMPLISHED

IF IT IS DONE ONE STEP

AT A TIME.

ACKNOWLEDGEMENTS

All praise to **Almighty Allah** Who has all the names, and Who needs no name the most generous, considerate, and compassionate. Allah has blessed mankind with this verdict to think, explore, to learn and discover the hidden secrets of this universe and helped me to broaden the veils of my thought. And Who enabled me to get through the difficulties indulged during this project. Also admiration to our beloved **Prophet Muhammad (PBUH),** who is always a great source of inspiration of divine devotion and dedication to me.

I would cordially pay my special appreciations and whole heartedly considerations to my reverend supervisors **Prof. Dr. Muhammad Sher** for his endless support, guidance and coordination while conducting this project. I owe him a great respect and honor and I am privileged to work under his supervision. It was his effort, courage, moral support and endeavoring attitude that helped me to get through any problem or difficulty during each step of this project.

I would also like to pay my gratitude to all my respected teachers making me capable of what I am today due to their guidance and help. Thanking **Dr. Ali Dawood, Dr. Ayyaz Hussain, Dr. Hasnain Abbas Naqvi, Mr. Shahzad Ashraf, Mr. Syed M Saqlain**, Mr. Anwar Ghani, Mr. Eid Rehman, Mr. Tahir Khattak and Mr. Salabat Khan for their views which helped me in improving my research, also Mr. Bilal Shah for providing the managerial and administrative support.

Finally, my beloved parents and family, who deserve the credit more than I could ever express for always being completely supportive to me. They have been a constant source of advice, love and devotion to me. From moral to financial they have been blessing me with all the support that I needed up till now in my life.

KHALID KHAN

567-FBAS/MSCS/F09

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER NO: 1

INTRODUCTION

1. Introduction

Wireless sensor networks arose as an important new area in wireless technology. It has gained interest of researchers from the last twenty years. Wireless sensor network consists of sensors which are distributed randomly. These sensors are used to detect physical changes or environmental conditions, such as sound, temperature, air pressure and intrusion detection. Sensors send their data in the network to other sensors and finally to a base station also called sink node by using wireless media. Some of the sensor networks also monitor sensor node activity. Wireless sensor network was initially deployed for military applications such as surveillance of battle field. Nowadays wireless sensor networks are used in many industrial and consumer applications such as process control in industries, monitoring of machines etc. The wireless sensor network is composed of sensor nodes which range in number from a few nodes to several hundreds or even thousands of nodes, where each node is connected to one or more than one sensors. The sensors have low power, limited memory and computational resources [1,2].

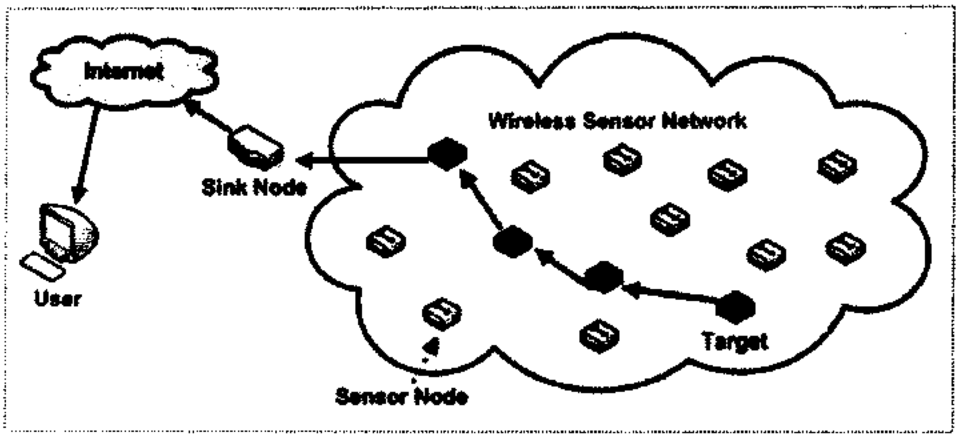


Figure 1.1: Wireless Sensor Network

1.1 Sensor Node Architecture

A sensor node of a wireless sensor network is able to perform some processing, detecting sensory informations and also communicating with other nearby nodes in the network. A sensor node may vary in size from the size of a shoe-box to the size of a grain of dust. Typical sensor node architecture consists of the following parts [3].

1. Application specific sensor
2. Wireless transceiver
3. Simple processor
4. Power source
5. Smaller memory

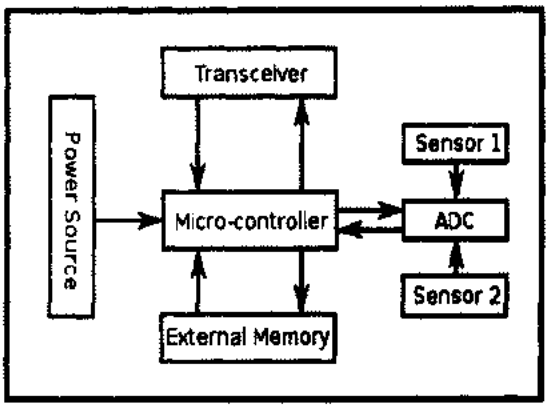


Figure 1.2: Typical architecture of sensor node

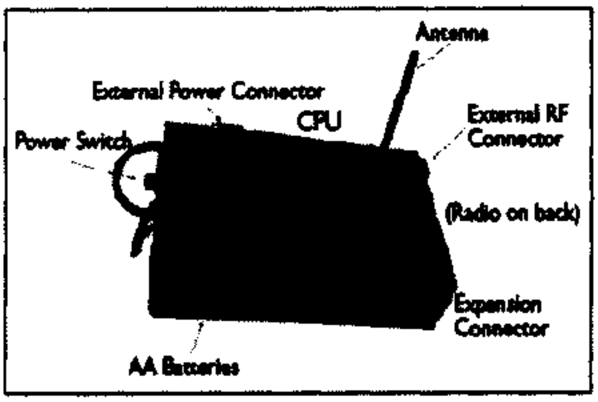


Figure 1.3: Sensor node

1.2 Applications of wireless sensor networks

Wireless sensor networks are becoming popular in many circles of life. They are deployed in wide and vast environment for commercial, civil, and military applications such as surveillance, tracking of vehicles, monitoring of climate and habitats, field of intelligence, medical field and gathering of acoustic data [4]. The deployment of these networks is easy and can be implemented in harsh environments. Some of the well-known applications of wireless sensor networks are as mentioned below.

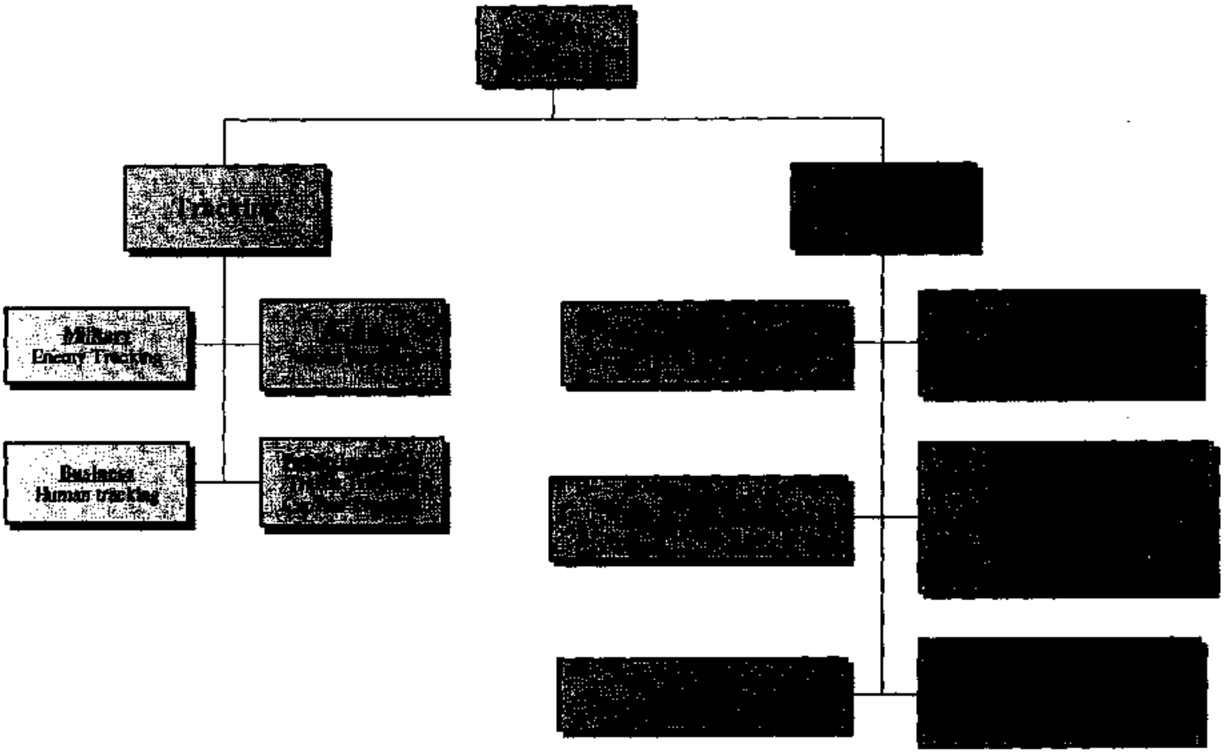


Fig: 1.4 Applications of WSNs

a. Area monitoring

This includes enemy intrusion detection and monitoring of gas and oil pipelines.

b. Environmental/Earth monitoring

This includes monitoring of air quality, measuring pollution in air, detection of forest fire, detection of landslide, monitoring of water quality, detection and prevention of natural disasters.

c. Industrial monitoring

This includes applications like machine performance monitoring, control of industrial process and applications, and monitoring of water wastes.

d. Agriculture monitoring

Applications of wireless sensor networks in agriculture include accurate agriculture, irrigation management and green house effects.

e. Passive localization and tracking

f. Smart home monitoring

In disaster management setup, a large number of sensors can be dropped by helicopter.

1.3 Types of wireless sensor networks

Currently there are five types wireless sensor networks: Terrestrial WSNs, underground WSNs, underwater WSNs, multimedia WSNs, mobile WSNs.

1.3.1 Terrestrial Wireless Sensor Networks

Terrestrial wireless sensor network [5,6] consist of hundreds or thousands of sensor nodes,

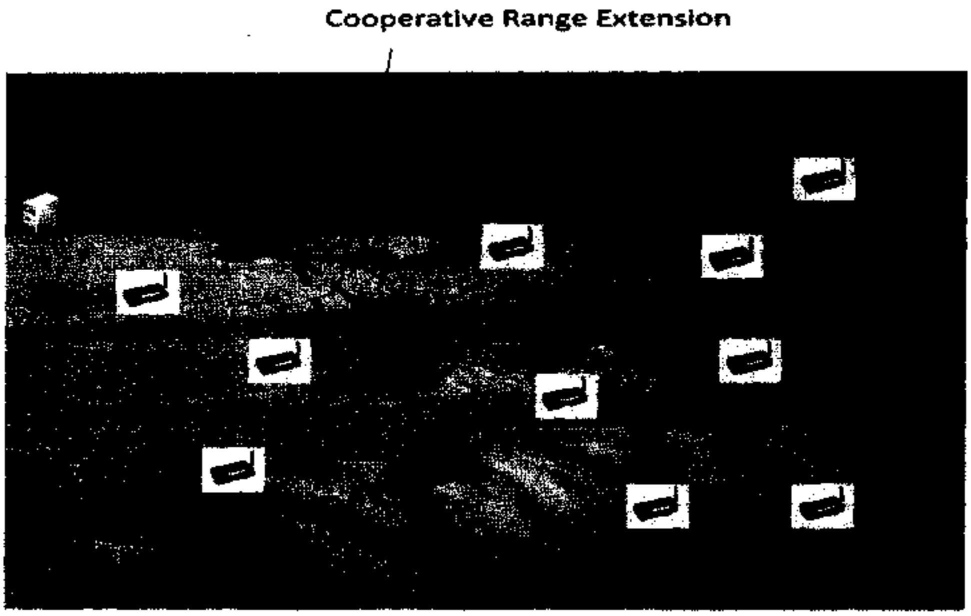


Figure: 1.5 Terrestrial wireless sensor networks may be deployed in an ad-hoc or may be in a fixed manner. Sensor nodes in ad-hoc deployment are dropped from a helicopter and are randomly placed. For a terrestrial wireless sensor network, energy can be saved with multi-hop optimal routing, shorter transmission range, data aggregation inside network and minimizing delays.

1.3.2 Underground Wireless Sensor Networks

Underground WSNs [7] consists of a number of sensors fixed in underground or in a mountain cave which are used to monitor the underground conditions. The base stations are located on upper surface of earth to reliably send data from the sensors to sink node. This type of WSN is more expensive than a terrestrial WSN on the basis of equipment, maintenance and deployment. The underground environment makes excessive signal loss and extra high level of attenuation. Underground WSN requires deep planning for deployment and high energy costs.

1.3.3 Underwater Wireless Sensor Networks

Underwater WSNs [8] consist of number of sensors and underwater vehicles or mobile nodes to collect and carry data to the surface getaways. As compared to terrestrial sensor network the underwater sensor network has sparse deployment, path loss, signal attenuation, harsh environment of water, limited bandwidth, longer delays due to acoustic signals etc. The battery cannot be recharged or replaced.

1.3.4 Mobile Wireless Sensor Networks

Mobile WSNs have sensors which interact with physical environment and can move here and there when their movement is needed. The mobile nodes are able to sense, communicate and compute like a static node. One of the differences is that repositioning of nodes into a network again. One other difference is that of dynamic routing in mobile WSNs while static routing and flooding are used in other sensor networks.

1.3.5 Multimedia WSNs

Multimedia WSNs [9] have been proposed for detection of events in the form of videos, pictures and audios. Multimedia WSN consists of low cost sensor nodes with camera and microphone. In these networks sensors are deployed in a fixed manner for granted communication and coverage. Challenges are demand for high bandwidth, highly energy consumption, quality of service and compressing techniques.

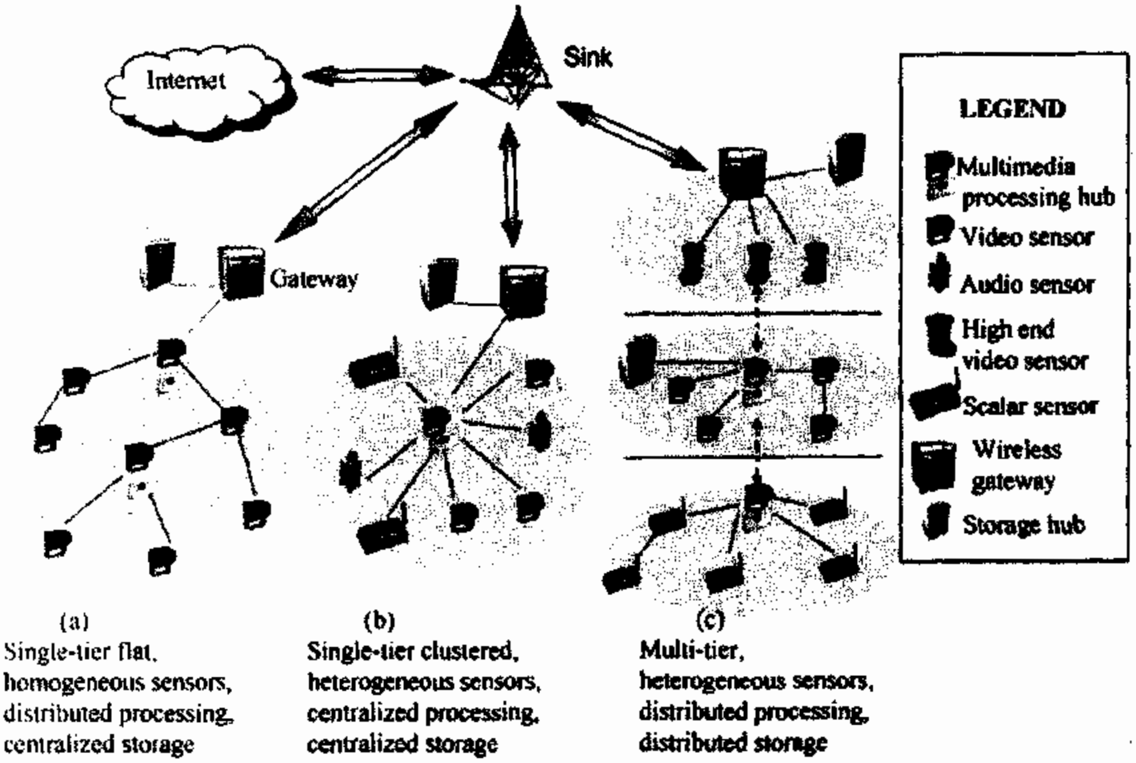

Internet
Sink
Gateway
LEGEND
Multimedia processing hub
Video sensor
Audio sensor
High end video sensor
Scalar sensor
Wireless gateway
Storage hub
(a)
Single-tier flat, homogeneous sensors, distributed processing, centralized storage
(b)
Single-tier clustered, heterogeneous sensors, centralized processing, centralized storage
(c)
Multi-tier, heterogeneous sensors, distributed processing, distributed storage

Figure: 1.6 Wireless Multimedia sensor network

1.4 Protocol Stack for wireless sensor network

The protocol stack for wireless sensor network is similar to the traditional protocol stack. Various layers are: Application layer, Transport layer, Network layer, Data Link layer and Physical layer. Wireless sensor network contains the following management planes i-e Task

Management, Mobility Management and Power Management Planes. The function of Mobility Management Plane is to detect and register movement of the sensor nodes. The function of Task Management Plane is to schedule and balance the tasks related to sensing to some of the nodes and the rest could consume their energy on data aggregation and routing. Power Management Plane has the responsibility of minimization of power consumptions and also has to terminate some of the functionality to preserve energy consumptions. The protocols developed for wireless sensor networks necessarily address all three planes.

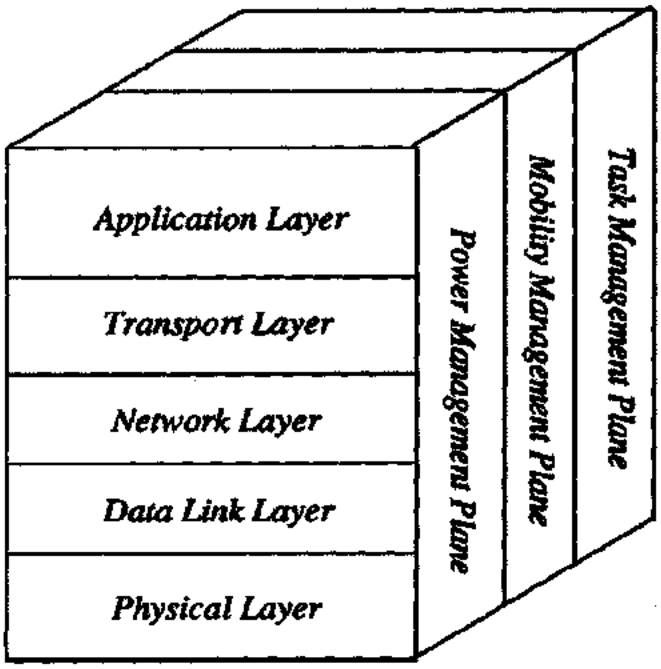


Figure: 1.7 Protocol Stack for wireless sensor network

1.4.1 Physical Layer

This Layer is responsible for functions such as selection of frequency, generation of carrier frequency, modulation, encryption and signal detection. Its main function is to minimize energy consumption and other concerns are same to other wireless networks. The minimum power required to transmit signal over a distance s is directly proportional to a power of p, where p varies from '2' to '4' and is near to '4' when the antennas are nearer to the ground which is common in the wireless sensor networks.

1.4.2 Data Link Layer

This layer is responsible for error control, data frames detection, media access and multiplexing of stream of data. For a wireless sensor network there must be a specialized MAC protocol to minimize power consumption and to minimize data-centric routing. The MAC protocol must have two goals. The most important goal is to create the infrastructure of network, which may include establishment of communication links amongst hundreds or even thousands of the sensor nodes and also providing the network with capability of self-organization. Second goal of Data Link Layer is to efficiently share communication resources amongst all sensor nodes. Existing MAC protocols for wireless network do not meet the above two goals. The reason behind is that power conservation is a secondary concern. Some of the well-known MAC-protocols are sensor-MAC (SMAC) and CSMA for wireless sensor Networks.

1.4.3 Network Layer

The network layer of a wireless sensor network must be designed with keeping in considerations of power efficiency and location awareness. The Data Link layer is related with how two nodes communicate with each other whereas the network layer decides which node has to talk to which node. One of the designs is flooding which is very simple. In flooding each node receive data. All the nodes then broadcast the same data to every other node, until the lifetime of the data expires or the destination node is reached. The major advantage of flooding is the simplicity. It does not require expensive topology maintenance or complexity in route discovery. The first problem in flooding is that of implosion. The second problem is that of overlapping, when two nodes share the same sensing region. The last and most problematic issue is resource blindness.

1.4.4 Transport Layer

The transport layer plays its rule when the system needs communication from one point to other point. Communication from the sink to the user is a problem because the wireless sensor network has not to communicate with outside world. Usually in wireless sensor network the data sink is in direct contact to the network and internet is not used. So in this layer a very little research has been made.

1.4.5 Application Layer

An SMP, Sensor Management Protocol, at the application layer is used to make the software and hardware of other layers transparent to the wireless sensor network management applications. The programmers and system administrators interact with the sensor network by using SMP. The infrastructureless nature and lack of global identification of sensor networks must be brought in consideration while developing applications for these networks. SMP provides the rules for the following to enable interaction between applications and sensor networks:

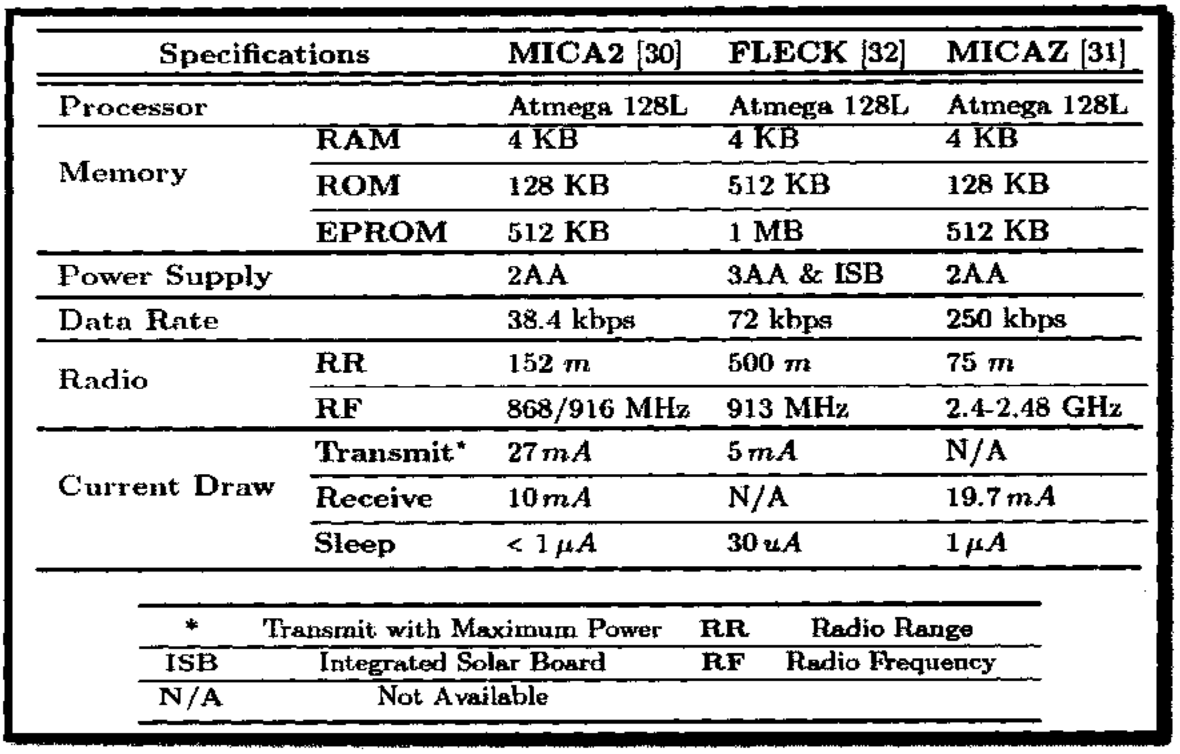
- Clustering, data aggregation and attribute-based naming
- Exchange data related to the location finding algorithms
- Moving sensor nodes
- Time synchronization
- Turning nodes on or off
- Authentication, key distribution, and security
- Querying WSN configuration status, reconfiguring the WSN

1.5 Challenges in wireless sensor networks

The key limitations of wireless sensor networks are the smaller storages, limited energy resources, limited computational capabilities and highly exposed to security threats. All these limitations and due to the architecture of sensor node require energy efficient and secure communication and aggregation protocols. The invention of MEMS (micro electromechanical systems) technology, low power and low cost digital signal processors (DSPs) and radio frequency circuits has contributed much in the wireless sensor network technology [10,11]. The most important challenge in wireless sensor network is to increase the lifetime of sensor nodes and network. The main cause is that it is not possible to change the batteries of thousands of sensor nodes. Therefore, while developing protocols for sensor networks one must keep in mind that how to decrease the computational operations of nodes. Also communication protocols must be made efficient from energy point of view. Among the protocols for wireless sensor networks, the data transmission protocols are of greater importance in terms of energy utilizations. Because 70% of the total energy is consumed in data transmission [12]. Data aggregation techniques can greatly contribute to conserve the limited energy resources of sensor network by removing data redundancy and minimizing the number of bits or data transmissions.

As compared to ad-hoc networks, wireless sensor network have unique specification and constraints due to which simple adaptation of the existing solution design for traditional network is impractical. These unique specification and constraints are named as challenges and are classified into (i) Challenges in the end device (ii) Challenges in network [13].

Table: 1.1 Hardware Specification for three types of nodes



Specifications		MICA2 [30]	FLECK [32]	MICAZ [31]
Processor		Atmega 128L	Atmega 128L	Atmega 128L
Memory	RAM	4 KB	4 KB	4 KB
	ROM	128 KB	512 KB	128 KB
	EPROM	512 KB	1 MB	512 KB
Power Supply		2AA	3AA & ISB	2AA
Data Rate		38.4 kbps	72 kbps	250 kbps
Radio	RR	152 m	500 m	75 m
	RF	868/916 MHz	913 MHz	2.4-2.48 GHz
Current Draw	Transmit*	27 mA	5 mA	N/A
	Receive	10 mA	N/A	19.7 mA
	Sleep	$< 1\,\mu A$	30 uA	1 μA

*	Transmit with Maximum Power	RR	Radio Range
ISB	Integrated Solar Board	RF	Radio Frequency
N/A	Not Available		

1.5.1 Challenges in the End Device

Some of the challenges in the hardware of sensor device are discussed as follow.

a. Limited memory

Sensor node is a small device with a limited memory and storage capacity. To develop one of the best security mechanisms, the code must be small enough to be accommodated in the small memory of sensor. For example the sensor type MECA2 as shown in the table 1.1 has 4KB of RAM, 128KB of program memory and 512KB of flash storage [14]. The total code space for TinyOS (operating system for sensors) is approximately 4KB, while core scheduler takes only about 178 bytes. So the proposed solution must be small enough to accommodate in such smaller memory.

b. Limited Energy Resources

It is the most important challenge in the wireless sensor networks. Once the sensor node is deployed, its battery can neither be recharged nor replaced due to high operating cost. Some of the sensor node MECA2 is powered by 2AA batteries as shown in table. While developing a protocol in a sensor network, the energy resource must be considered.

c. Limited CPU Performance

The CPU architecture in MECA2 is 16-bit and computational power of 8MHz [14]. Complex algorithms as developed for wired networks must be avoided and so simple algorithms should be developed on wireless sensor networks.

d. Tamper Resistant Hardware

Since nodes in the wireless sensor networks exist in remote and hostile areas so they are exposed to hard environmental conditions and also to human and animals hazards. For this, extra strong circuits and body should be used. To run these circuits, powerful energy and extra cost is required.

1.5.2 Challenges in the Network

To perform some specific task sensor nodes are usually scattered in the deployed area. Usually there is no infrastructure for sensor network. To form a network, sensor nodes self-organize into network. Still some network challenges exits. These challenges are as follow.

a. Hostile and Remote Environment

Most of the wireless sensor networks are deployed in hostile and remote areas such as whether for costing and battle fields. Since these nodes are exposed to all types of threats. So nodes can be accessible by adversary physically. He could capture sensor node data and could enter false data into the node.

b. Random Topology

Most of the wireless sensor networks have no specific topology since they are deployed randomly by dropping from a helicopter or any other source. Secondly, since they are deployed in remote and hostile areas the node may fail due to physical damage or due to energy drain off quickly. Due to these reasons the topology keeps on changing with time.

c. Latency

In most of the sensor networks the communication range is too short i-e 100 m to 500 m, because the energy consumption of the network is more in longer range communication than a shorter range communication. Therefore, to move a packet from one end of the network to another end, multi-hop routing is used. Due to this reason, the packet is processed on multiple nodes for some time. So, multi-hop routing and node processing can lead to great delays. In this way the synchronization among nodes is made difficult. The security issues are created from the synchronization problem, and the security mechanisms such as critical event report and cryptographic key distribution should be applied. And these mechanisms are threats to sensors energies.

d. Unreliable communication

This issue is same to other wireless networks. Due to wireless communication the packet may be dropped due to channel error, lack of radio coverage or by highly congested nodes.

1.6 Data Aggregation

In most of the wireless sensor network applications, sensor nodes sense the physical phenomena and send the report/data to Base Station. In order to minimize the energy consumptions of the sensor nodes, the packets sent across the network must be reduced. So these applications may employ an in-network mechanism called 'Data Aggregation'.

Data Aggregation is the collection and combination of data from various sources and reduction of data size by using functions such as suppression (eliminating duplicates), sum, finding minimum or maximum value, averages, variance and standard deviation [15]. Aggregation protocols are used to minimize the amount of data communication in the network. It conserves battery power. As data are detected and sensed by individual sensors, they need to be collected and processed. One of these approaches for sensor network is to send sensor's data to certain special nodes, the Aggregators. Each aggregator, aggregate the data before sending to other nodes. The aggregating nodes may either be special and more powerful nodes or they may be regular sensor nodes. Here, we assumed that all nodes are aggregating nodes except leaf nodes. We also consider that data is aggregated by aggregators before sending.

Consider the example in figure 1.8, here the in the network topology there are sixteen 16 sensor nodes. The aggregators perform SUM function as data aggregation function. The nodes N1, N2, N3, ... N8 are regular sensor nodes and only sense data and send to upper nodes. Nodes N9 to N16 are aggregating nodes that perform both sensing and aggregation functionalities. Here the sensors N1 to N8 will sense data and send to the aggregators N9 to N13. Theses aggregators add their sensed data add it to the received raw data from other sensors, and then apply SUM function. These aggregators send these data to the aggregators N14 and N15 above them. The aggregators N14 and N15 repeat the same action as other aggregators. At the N16 only required data will be received according to the query from the base station. Thus 16 packets are sent across the network due to data aggregation. But if the data aggregation is not applied on this network then a total of 50 packets will be sent across the network. These sixteen packets are the nodes response to base station's query.

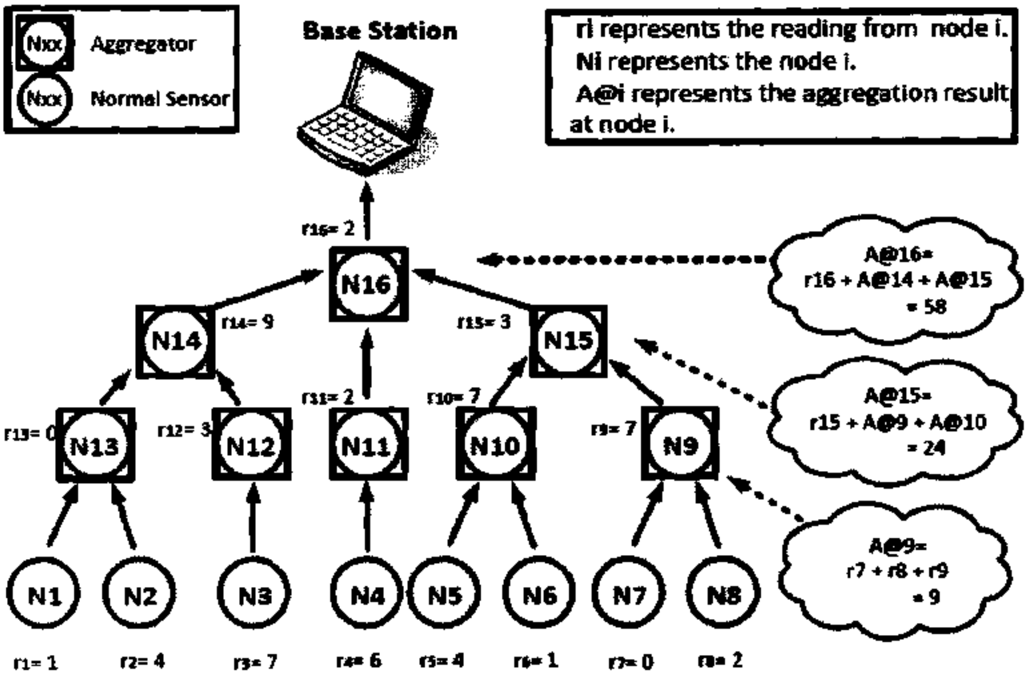


Figure 1.8 An Aggregation function using SUM function

1.7 Advantages and Disadvantages of Data Aggregation

Advantages: Some of the advantages of data aggregation are

- To reduce the number of bits/data transmitted across the network
- To enhance the accuracy and robustness of information received from the network by removing redundant data

- To reduce the traffic load in the network
- And finally to conserve the energies of the sensors and to increase the life of the sensor network

Disadvantages: Some of the disadvantages of data aggregation are

- If the aggregator is compromised then a huge amount of data may lost or changed depending in adversary type
- Delay in data may occur because aggregator may take much time to perform aggregation function
- Part of the network may damage if the aggregator dies
- Extra computation on the aggregators which result quicker loss of energy of these nodes

1.8 Performance Measures of data aggregation

The performance measures are mostly dependent on application concerned. Some of the performance measures of wireless sensor networks are as follow.

a. Energy Efficiency

Using data aggregation, we can increase the energy efficiency of the network by increasing network lifetime and accuracy, and reducing latency and communication overheads.

b. Network Lifetime

It is the data rounds of the fusion function. In data aggregation, these rounds are increased by the intelligent use of sensors in the network. One part of the network is used at a one time and time the other part of the network is used. In this way the network life time is increased.

c. Latency

Latency is evaluated as the time delay between data sense by the sensor, passing data through the network and finally reached to base system. This delay may occur due to data sending or transmission, routing and functions such as data aggregation.

d. Communication Overhead

It is defined as the complexity in data aggregation function. More complex the aggregation function more will be the communication overhead.

e. Data Accuracy

It is the ratio between total number readings received by base station to the total number of readings produced.

1.9 Data Aggregation Architecture

There are three basic architectures for data aggregation function in wireless sensor networks [16] which are explained as follow.

1.9.1 Centralized Architecture

This is the simplest architecture for wireless sensor networks in which aggregation function is applied by a single central node. All the sensors sense data and send to the central node also called cluster-head. The advantage is the simplicity of the network. The disadvantage is the whole working load is on a single node. If this node is damaged or compromised this whole network is affected.

1.9.2 Decentralized Architecture

Here in this network, there is no centralized approach or central node. All the nodes are responsible of data aggregation. The advantages are the reliable network, scalable network and network can be dynamically changed. The disadvantage is the maximum utilization of energy of the energy constrained wireless sensor network.

1.9.3 Hierarchical Architecture

It is one of the important architecture for aggregation of data in wireless sensor network. Here all the nodes are partitioned into hierarchical tree. Sensors sense data and send to nearest aggregator, then from one aggregator to next aggregator, and finally to the sink node. Advantage of this approach is even distribution of work load in the network.

1.10 Data Aggregation approaches in Hierarchical Architecture

In hierarchical approach, the network contains data aggregators which perform in-network aggregation and enroots data to sink. Hierarchical approach can be categorized into four types [17] which are explained as bellow.

1.10.1 Cluster Based approach

Since it is known that wireless sensor network has limited resources therefore, transmitting sensor's data directly to the sink is an energy consuming action. In cluster based approach the data from the sensor is sent to a central node called the Cluster-Head as shown in figure 1.9. This cluster-head aggregate the data from the sensors and send to another cluster-head which is nearer to the base station. This is done through cluster to cluster communication. Cluster head can communicate to base station either directly or through other cluster head. Using this approach the energy of sensors can be saved which results in increased lifetime of the network.

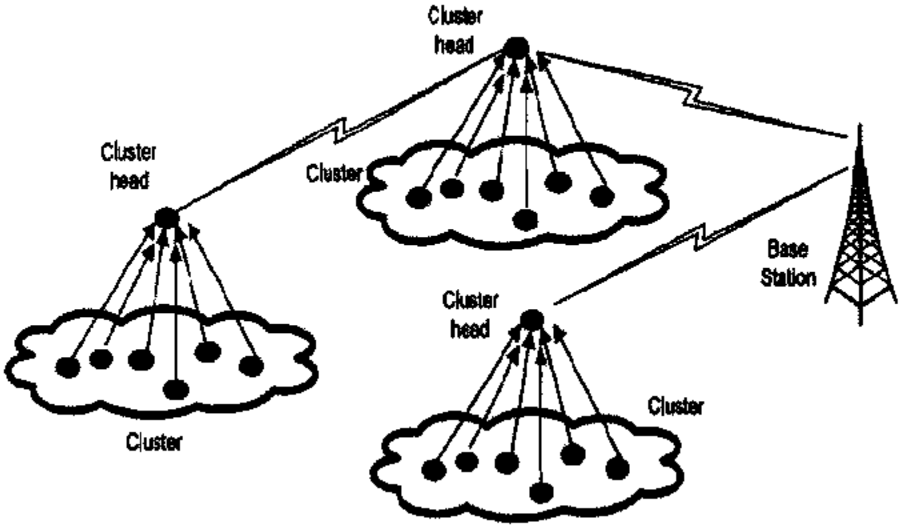


Figure: 1.9 Cluster based approach

1.10.2 Tree Based Approach

In this approach all the nodes are organized in the form of hierarchical tree shown in figure 1.10. The leaf nodes are the sensor and the nodes in other levels are aggregators. Sensors sense data and send it to aggregator. The aggregators send data to the aggregators in upper level. So data reach is routed towards base station in tree form. This approach is suitable for applications which apply in-network data aggregation function. For example monitoring and control of radiation level in nuclear power plants.

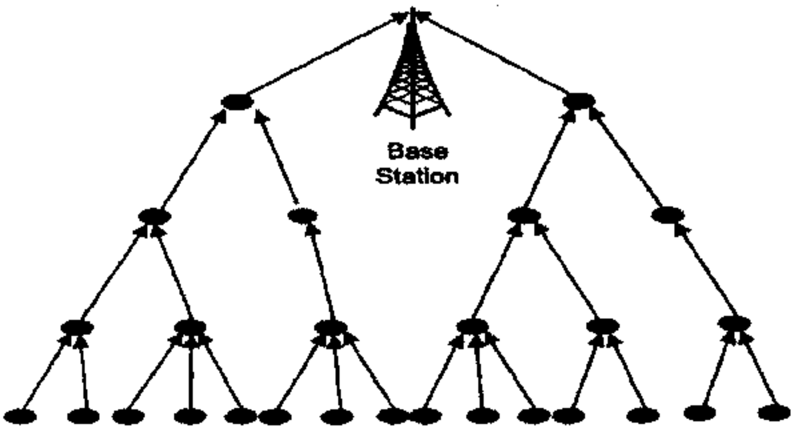


Figure: 1.10 Tree based approach

1.10.3 Multipath based Approach

Wireless sensor network is subject to frequent node and rout failure. So in this manner the tree based approach is limited as the leaf node or aggregator or the rout can easily be damaged or

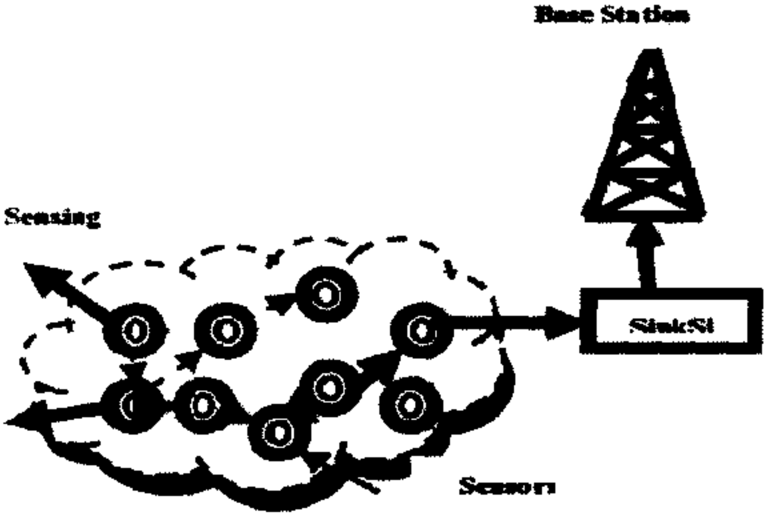


Figure: 1.11 Multipath Approach

compromised. In multipath approach every node can send data to the nearest node. Data aggregation is performed by and every intermediate node found between source and base station sink. By using this approach the system can be made more robust then tree based approach. But the communication overhead is more than the tree approach. Figure 1.11 shows this approach.

1.10.4 Hybrid Approach

This approach is the combination of all the three approaches mentioned above i-e tree based approach, multipath based approach and cluster based approach. Figure 1.12 explains the principal of this approach. In hybrid approach the aggregation function will be performed on the basis of any of the three approaches based on specific network situation and performance statistics.

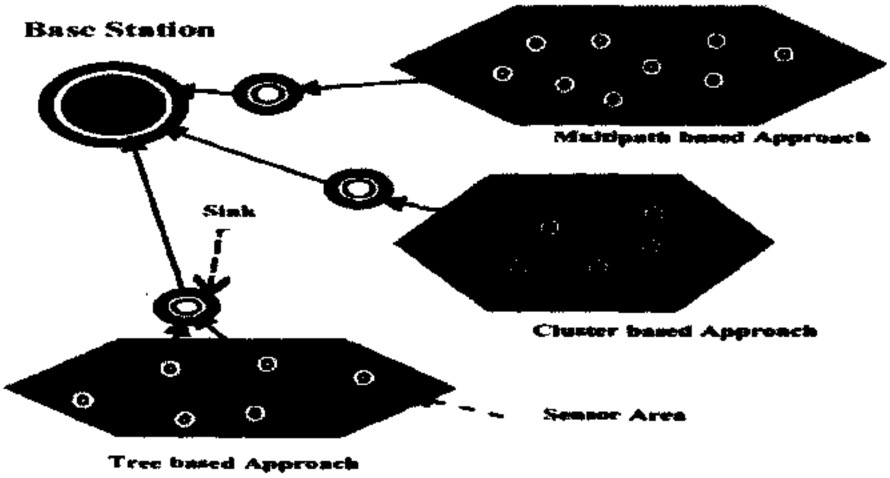


Figure: 1.12 Hybrid Approach

1.11 Secure Data Aggregation

If data aggregated by aggregator is accessed by an adversary then it will be a big damage to the system. There is a need of security for the data as data move upward in the network. One of the best solutions to this problem is cryptography. Cryptography is defined as converting plan text into a format that cannot be read by the attacker directly (encryption) called *Cipher Text* and converting the cipher text back to plan text (decryption) at the sink or destination point.

Three types of secure data aggregation techniques are there in tree networks which are as below.

1.11.1 Hop-by Hop Secure Data Aggregation

In hop-by-hop secure data aggregation scheme the data is sense by the sensors and encrypted. This encrypted data is then sent to the aggregator. The aggregator decrypt the data received from sensors and make the aggregate and then decrypted the data. Now this encrypted aggregated data is sent to the aggregator on next higher level. This process is continued until data reach the base station. The advantage is the security of data. The disadvantage is the increased overhead in terms computation on the aggregators.

1.11.2. End-to-End Secure Data Aggregation

In such type of data aggregation scheme, data sensed by the sensors is encrypted and send to the aggregator. The aggregator make the aggregate of the data received from the sensors without being decrypted. This method of making aggregate from encrypted without decryption is called "Homomorphic Encryption". Advantages of this scheme are removing extra computation from the aggregators and providing end-to-end privacy to data.

1.12 Additive Homomorphic Encryption

Here additively homomorphic encryption technique is introduced in [19, 26] which is used to make the aggregate of encrypted data without being decrypted. Such type of Homomorphic encryption scheme is useful in problem scenarios where there is no decryption key and someone want to perform arithmetic operations on a set of cipher texts. The basic scheme is explained as below.

a. Encryption

Encryption is defined as converting plain text into cipher text, so that data should not be read by unauthorized person.

i. Message m is represented as integer, where m ∈ [0, M −1] and M is the modulus.

ii. Suppose 'k' is randomly generated key stream, and k ∈ [0, M − 1].

iii. Cipher text is compute as $c = Enc\ k\ (m) = (m + k)\ mod\ M$.

b. Decryption

Decryption is defined as converting as cipher text into plain text.

$Dec\ k(c) = (c - k)\ mod\ M$.

c. Addition / Aggregation of Cipher texts

i. Let $c_1 = Enc\ k_1\ (m_1)$ and $c_2 = Enc\ k_2\ (m_2)$.

ii. Aggregated cipher text: $c_l = (c_1 + c_2)\ mod\ M = Enc\ k\ (m_1 + m_2)$ where $k = (k_1 + k_2)\ mod\ M$.

The correctness of aggregation will be assured only if M is sufficiently large. The reason is as because if, $c_1 = m_1 + k_1\ mod\ M$ and $c_2 = m_2 + k_2\ mod\ M$, then $c_l = c_1 + c_2\ mod\ M =$

$m_1 + m_2 \; + \; k_1 + k_2 \; mod \; M$ =$Enc \; k_1 + \; k_2(m_1 + m_2)$. For $k \; = (k_1 \; + \; k_2)$, $Dec \; k(c_l) \; =$ $(c \; - \; k) \; mod \; M = (m_1 + m_2) + (k_1 + k_2) - (k_1 + k_2) mod \; M \; = (m_1 + m_2) \; mode \; M$

It is assumed that 0 ≤ m < M. if *n* different ciphers *ci* are added, M must be larger enough otherwise, correctness of the function does not hold.

1.13 Contribution of the Thesis:

The main emphases in data aggregation protocols for wireless sensor networks is on ensuring security, reducing computational and communication costs. The ultimate result is increased lifetime of the wireless sensor networks. The existing protocol SEEDA [27] ensures end-to-end privacy of data but this protocol creates sufficient communication and computational overheads by computing and sending of cipher text for non-responding nodes. We emphases on these two issues and assume the network and security model to be same as used in SEEDA.

The contribution of the proposed scheme (EDAS) is to transmit less numbers bits/data than existing scheme SEEDA as possible because data transmission data transmission consumes 70% energy of the in wireless sensor networks. Also our goal is to decrease the computational cost created on the aggregators. By decreasing computational and communication cost in wireless sensor network, we can prolong the life time of wireless sensor network.

1.14 Organization of Thesis

The rest of thesis is organized as below.

In chapter 2 a detailed review of the past work done on data aggregation protocols in explained in detail. This chapter contains literature from 2005 up to 2014.

Chapter number 3 is related with the concise overview of the problem statement. It also contains problem scenarios.

Chapter 4 is related with proposed solution, EDAS algorithm and EDAS flow chart.

Chapter 5 contains explanation about simulation tools and environment. This chapter also contains EDAS and SEEDA simulators, and parameters used in research.

Chapter number 6 is related with results of the research in graph form and conclusion of the thesis.

And at the last all the references are mentioned which are used in the thesis.

CHAPTER NO: 2

LITERATURE REVIEW

2. Literature Review

To reduce the data transmission through the network a Pattern Based Energy Efficient Protocol (ESPDA) [20] is proposed by H. Cam and S. Ozdemir. This protocol reduces data transmission through the network by sending patterns of data instead of sending whole data to the base station. The pattern codes reflect characteristics of original data sensed by sensors. ESPDA is cluster based approach where multiple sensors are attached to it. The sensor nodes sense data from the surrounding environment and then make patterns of the sensed data. The patterns are generated in such a way that they represent the sensed data in all respect. Patterns are then sent to cluster head. Cluster head compare the patterns received from various sensors and send only those patterns which are distinct. The redundant patterns are removed here. This protocol also has the capability to sleep those sensor which sense redundant data from the environment. It results in energy conservation of the sensors and increase network life time. For security purpose, ESPDA refreshes the mapping interval of patterns code periodically. ESPDA has some limitations as well along with benefits. This protocol maintains a lookup table for pattern generation in each sensor. For storage constrained sensors, this lookup table is wastage of storage. Second drawback is the recalculation of lookup table time and again after specific interval. It creates computational overheads. The sensor always searches for a critical value in the look up table each time a new data is sensed which again results in computational overheads. The third limitation is that ESPDA do not allow intermediate aggregation. It only allows immediate aggregation.

SRDA, Secure Reference based Data Aggregation protocol [21] is proposed by H. Sanli and S. Ozdemir. In this protocol the base station send reference value to sensor network. Sensors make the deferential data of the original data and reference value. So differential data is then sent to the base station instead of raw data due to which, number of bits transmitted in the network are reduced. For security purpose, SRDA uses an algorithm with consideration of security margin as adjustable parameter. In the first step sensors send raw data to the cluster head. The cluster head compare the raw data with reference value. Now the differential value is sent to higher level. The reference value is refreshed for security reason after session expires. The new reference value is calculated as average of the previous sensor readings. In the conventional data aggregation protocol, raw data was sent by sensors to cluster head which is energy consuming process. Reducing data transmission in wireless sensor networks is essential. To explain SRDA, let us take an example. Let the temperature measured by sensor is 105 F, and reference value is 100. Then differential value will be 5 F. The efficiency of the SRDA will be greater if reference value is nearer to the sensor readings. The drawback of SRDA is same as that of the ESPDA. It also do not allow intermediate node to perform data aggregation. It only allow immediate node to perform aggregation function which significantly reduces the importance of data aggregation.

In paper [22], a Secure Hop-by-Hop Data Aggregation Protocol (SDAP) is proposed by Y. Yang and X. Wang. The authors of the SDAP realized that most of the trust is placed on higher level nodes as compared to low level nodes, which is not a wise decision. So if the compromised node is at higher level, then there will be more chance of false results at base station. This is because higher level nodes carry larger aggregated data then lower level nodes. There is need of method which could reduce trust on the higher level nodes. SDAP has reduced the trust on higher level nodes using divide-and-conquer principal. It uses probabilistic approach and dynamically partitions the topology tree into multiple sub-trees of equal size. Since in this approach, fewer sensors are attached to each higher level node so the security risk of larger data being compromised is reduced. SDAP provides data integrity, source authentication and confidentiality. Along with these benefits, this protocol has also achieved applicability on multiple aggregation functions. The limitations of this protocol include greater amount of energy utilization and transmission overheads. In SDAP data is decrypted at aggregators for aggregation function. At this stage is data is open to intruders and can be captured easily. Moreover extensive encryption and decryption at the aggregators results in computational overheads.

The problem of Hop-by-Hop encryption and decryption is solved by Concealed Data Aggregation (CDA) [23] proposed by D. Westhoff. This protocol is based on symmetric cryptographic key which mean same secret key is used throughout the network. Data sensed by sensor is encrypted using this secret key. Encrypted data is sent to aggregator. Aggregators do not decrypt the encrypted data received from sensors. Due to Privacy Homomorphism (PH), aggregation function is applied on encrypted data. This process leads to get rid of costly decryption and encryption functions at each aggregator. CDA achieved end-to-end privacy of data. The aggregators are required not to store the secret key because they do not decrypt the data. In this manner the memories of aggregators are conserved. The limitation of this approach is the sharing of single secret key throughout the network. If a single node is compromised then intruder can easily capture the data of all the sensors. This technique is also vulnerable to reply attack.

A secure data aggregation protocol called CDAP [24] is proposed by S. Ozdemir which uses asymmetric based privacy homomorphic cryptography for achieving end-to-end to confidentiality along with data aggregation. The author of this paper realized that privacy homomorphism based on asymmetric cryptography carry high computational overheads. For resource constrained sensors cryptography is not affordable. To solve this problem, CDAP employee resource rich nodes called AGGNODEs (aggregator nodes). The function of these AGGNODEs is to carryout privacy homomorphism and aggregation. Each of the neighbors encrypts its data by using encryption: RC5 and sends this encrypted data to the AGGNODE. Now the AGGNODE decrypt all the received data which it get from its neighbors, aggregate them and by using PH the aggregated is encrypted. This encrypted data is then sent to base station. The encrypted data with PH encryption technique can only be decrypted by the private key of base station only. During forwarding process, due to PH property, the intermediate AGGNODE can aggregate the encrypted data. Privacy Homomorphism creates extra computational overheads then hop-by-hop schemes. However data transmission efficiency and aggregation ability of this scheme increase with the increase in number of AGGNODEs.

TH13992

Hierarchical Concealed Data Aggregation protocol (HCDA) [25] proposed by S. Ozdemir and Y. Xiao. This protocol aggregates the data hierarchically using elliptic curve cryptography for security purpose. Analysis shows that this scheme is resistive to node compromise attack. This protocol achieved concealed data aggregation and aggregate data of multiple sensor node groups where each group uses different public key to encrypt data. In this scheme the nodes deployed in groups where sensors are divided into multiple groups before deployment. Each of the groups is deployed over certain location for covering the whole area of interest. Each group is assigned with a separate public key to encrypt data. Based on this public key, the base station would easily determine that to which group this data belongs and also easily classify the data. Elliptic curve cryptography is better than public key cryptography from communicational, key size and memory point of view.

In [26] a new scheme Efficient and Provably Secure Aggregation of Encrypted Data based on homomorphic encryption is proposed. This protocol allows an aggregator to perform the aggregation function and aggregate the encrypted data which it receives from its sensors with no need for decryption or to make the plain text messages at the aggregators. Here in additive homomorphic function the aggregators do not decrypt or recover the original plain text message received from its children. Since the data is in cipher text form so even if an aggregator is compromised still data is safe and cannot be read. Due to this reason much stronger privacy is achieved than a simple aggregation scheme using technique like hop-by-hop encryption. From bandwidth point of this scheme is less efficient than the hop-by-hop scheme. Beside this it provides a stronger level of end to end privacy comparable to that provided by end-to-end encryption with no aggregation. In this scheme the communication load is distributed evenly among all other nodes. This results in longer network lifetime. The limitation in this scheme is the generation of significant overhead if some of the sensors do not respond. Number of bits (data) is increased due to sending of additional information related to non-responding nodes.

To decrease number of bits transmitted as well as to maintain the End-to-End privacy of data in the network, a Secure End-to-End Data Aggregation protocol (SEEDA) is proposed [27]. The SEEDA protocol is applied to tree based wireless sensor networks. The scheme ensures end-to-end data privacy. Less number of bits or data is transmitted from sensors to base station as compared to end-to-end aggregation protocol [26]. SEEDA protocol sends smaller amount of data which the best feature of hop-by-hop. It also adopts the best feature of end-to-end protocols which is end-to-end privacy of data. This scheme in this paper claims that the average data which is transmitted per node is reduced from 30% to 50% as compared to the scheme in [26]. The working of this protocol is explained as below.

- Responding sensor nodes sense data and encrypt using function $c_i = x_i + k_i \ (mod\ M)$ and send this encrypted data to the aggregators.
- Aggregator nodes add the encrypted data received from all responding nodes using Additive Homomorphic encryption i-e $C_{ag\,i} = C_{ag\,i} + c_i$
- For non-responding nodes, the aggregator node encrypts zero data using function $c_j = 0 + k_j\ (mod\ M)$. This cipher text of jth non-responding node is also added with $C_{ag\,i}$ i-e $C_{ag\,i} = C_{ag\,i} + c_j$. Now the ith Aggregator node will construct the message as $m_i = C_{ag\,i} \| q_i$ where q_i is count number of non-responding nodes.
- This message is forwarded to the aggregator at higher level of tree where the same homomorphic encryption is performed. This process is continued until message reach to the base station.

SEEDA protocol encrypts zero data at aggregators for non-responding nodes, $c_j = 0 + k_j\ (mod\ M)$. Since, this protocol computes the average of the data which require data of responding nodes only at the sink node. So there is no need of caring out of the above function and adding it with message. Due to encryption of zero data of non-responding nodes the number of bits transmitted in the network is increased. The computation overhead is also increased on the aggregators due to computing the function $c_j = 0 + k_j\ (mod\ M)$ for non-responding nodes. Extra energy is consumed and the life of the network is decreased.

Concealed data aggregation (CDA) [23] has much more importance in wireless sensor networks. Because the algorithms based on Privacy Homomorphism (PH) are basis for CDA. But the limitation here is that it sends whole encrypted data to the Base Station. This process is energy consuming from communication point of view. An Energy Efficient Recoverable Concealed Data (EERCDA) is proposed by Josna J [28]. This technique reduces the energy burden on the sensors imposed by the recovery based concealed data aggregation by transferring signature and cipher text both. The EERCDA transfer differential data from nodes to the Base Station instead of sending whole encrypted data. In this way the cluster head also has to process small amounts of bits/data. EERCDA gives energy relief to sensor nodes as well as the cluster head. Beside energy befit, it also preserves the security requirement such as source authentication, integrity and confidentiality.

Due to excessive encryption and decryption at the aggregators, the Hop-by-Hop aggregation protocols incur larger delays to data. This operation is not suitable for time crucial and delay sensitive military applications. To overcome delay problem, a protocol PEPPDA: Power Efficient Privacy Preserving Data is proposed by Joyce J [29]. This protocol reduces delay by allowing data aggregation function on encrypted data. It means data is not decrypted on the aggregators. Along with reduced delay PEPPDA also claims confidentiality, data freshness and accuracy. The protocol also claims low communication and computational costs.

A novel protocol called Dynamic Data Aggregation for Energy Optimization in Multi-Hop Wireless Sensor Networks [30] is proposed by Abhilash L N and Devaansh Goenka. This protocol optimizes the energy consumption of sensors in a multi-hop sensor network by using a new data aggregation technique. This technique discarded unnecessary packets to reduce data transmission and energy consumptions. This process is of two steps: in first step Exponential Weighted Moving Average (EWMA) data aggregation technique is used to compare the current data value with all the previous values before deciding whether to forward or drop the packet. The second step optimizes the network even further by considering readings from neighboring sensors into the equation. This protocol claims 20% reduction in data transmission. It also reduces bandwidth requirements of the networks.

CHAPTER NO: 3

PROBLEM STATEMENT

3.1 Problem Background

The main emphases in data aggregation protocols for wireless sensor networks is on ensuring security, reducing computational and communication costs. The ultimate result is to increase lifetime of the wireless sensor networks. SEEDA protocol [27] ensure end-to-end privacy of data and the average number bits (communication cost) is reduced compared to the scheme in [26], but this protocol create extra communication and computational overheads by computing cipher text for non-responding nodes. We will emphasize on these two issues and assume the network and security model to be same as used in [27].

3.2 Problem Statement

Greater amount of energy is consumed in sending more data/number of bits from sensors to the Base Station and higher will be the communication cost. Also extra computation on the aggregators and sensors consume more energy. Computational cost of the network become enormously higher. Higher communication and computational costs results in faster consumption of energy of the sensor network. Reducing data transmission and decreasing computational cost, we can increase the life of nodes and that of wireless sensor network.

3.3 Problem Scenario 1 (Communication Cost):

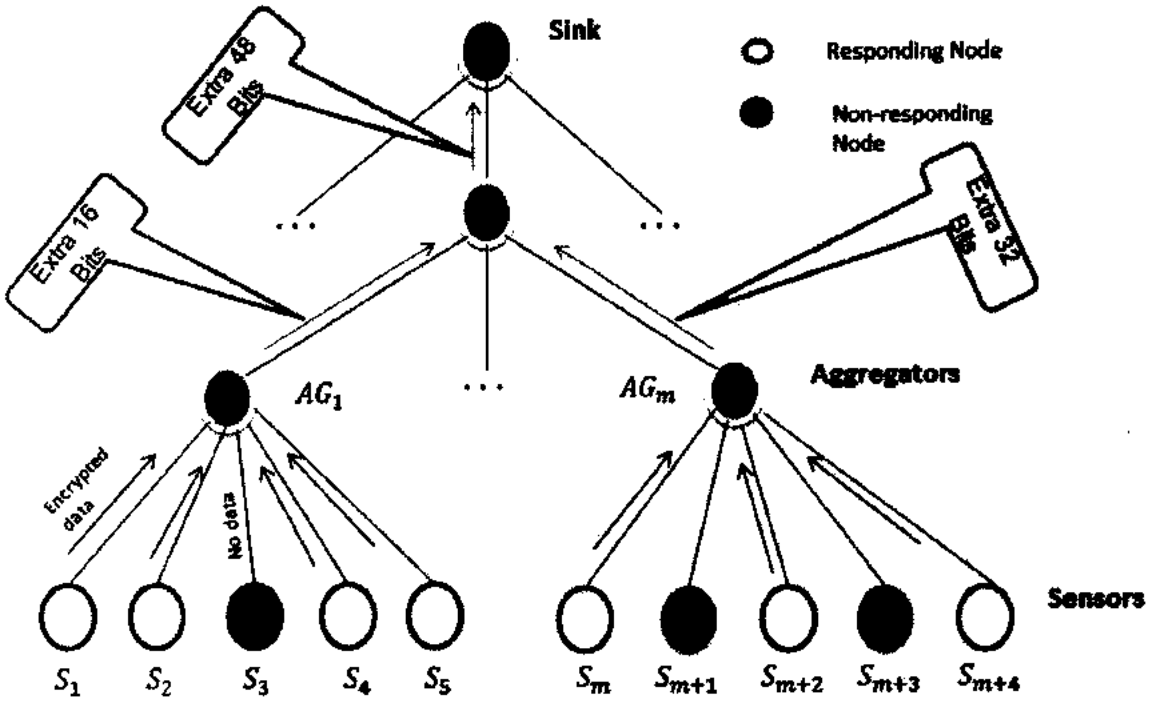


Figure: 3.1 Problem scenario 1 (Communication Cost)

- For non-responding nodes, the aggregator node encrypts zero data using function $c_j = 0 + k_j\ (mod\ M)$. Now this cipher text of jth non-responding node is also added with $C_{ag\,i}$ i-e $C_{ag\,i} = C_{ag\,i} + c_j$. Now the ith Aggregator node will construct the message as $m_i = C_{ag\,i} \| q_i$ where q_i is count number of non-responding nodes. So here it is clear from the above scenario 1 that adding c_j for non-responding with the message increases the message size. This will increase number of bits (communication cost) transmitted in the network.
- For example if cipher text size for one non-responding nodes calculated by the aggregator is 16 bit, then the total size of cipher text for all non-responding nodes at a single

aggregator will be calculated as Total size = number of non-responding nodes x 16 bits. The number of bits will increase as it goes upward through the tree.

3.4 Problem Scenario 2 (Computational Cost):

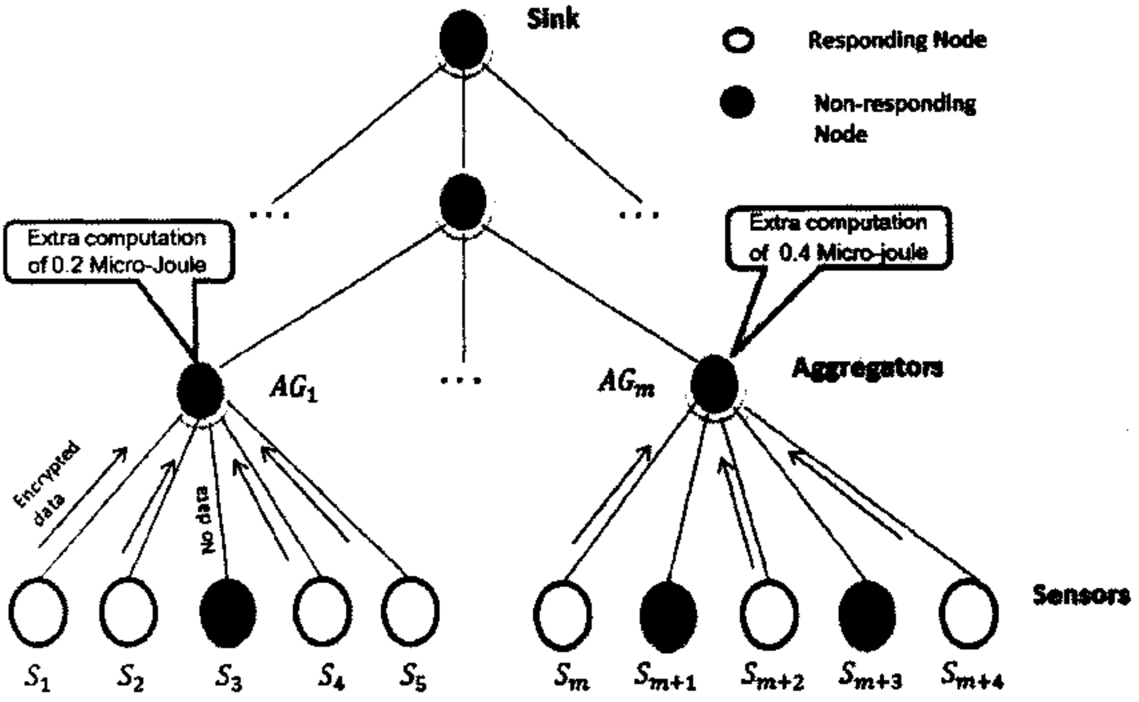


Figure 3.2: Problem scenario 2 (Computational Cost)

- At the aggregators the function is computed for all the non-responding nodes as $c_j = 0 + k_j\ (mod\ M)$ which also increases the computation cost.
- For example in the above scenario 2, if computational energy cost for single non-responding node is 0.2 Micro-Joules. Then the total computational cost for all non-responding nodes at single aggregator will be as Total Energy Cost = number of non-responding nodes x 0.2 Micro-Joules

CHAPTER NO: 4

PROPOSED SOLUTION

4. Proposed Solution

This chapter is related with detailed explanation of assumptions made in the research, network model, algorithms used in the scheme and flow chart of the protocol.

4.1 Assumption in EDAS

In EDAS we have made the following assumptions.

a. The network model is same as that of the SEEDA and tree based fixed network is used.

b. All nodes are potential aggregating nodes except leaf nodes which are sensors.

c. The security model is same as that of the SEEDA.

d. Key management and key distribution is also assumed to be same as SEEDA.

4.2 Network Model

The network model in our scheme is same as that of the SEEDA. A multilevel heterogeneous network tree is used. This tree consists of sensors, aggregators (AGGs) and a base station or sink node as shown in figure 4.1. Sensors are at the leaves i-e level 'h' of the tree. The aggregators are located in all levels of tree except level 'h' and level zero. The sensors have limited energy resources while that of aggregators have greater energy resources. Sensors sense data from the surrounding environment. The aggregators are responsible for forwarding queries and receive data from the sensors. They are also responsible for aggregating data received from various sensors. In the figure 4.1 the nodes $S_{0,}$ $S_{1,}$ S_2...... S_{2m} are sensor nodes and the nodes SG_l to SG_m are aggregating nodes.

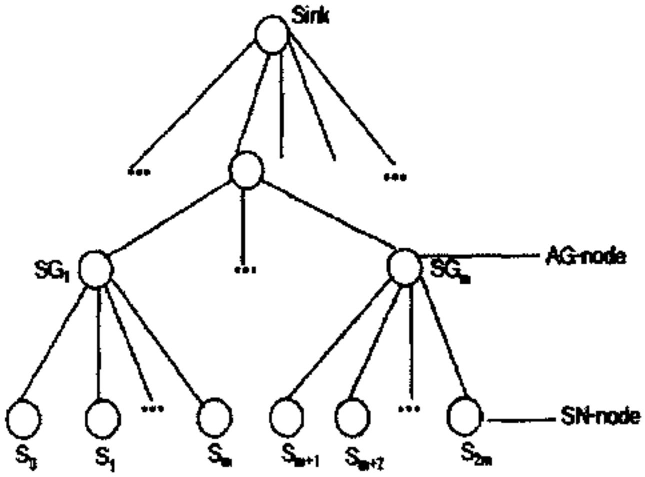


Figure: 4.1 Network Model in SEEDA

4.3 Proposed Model:

4.31 Methodology:

Each sensor node Si will sense data and encrypt using function $ci = xi + ki\ (mod\ M)$. This encrypted data will be sent to the corresponding aggregator node, AGG. The aggregator node will collect encrypted data from all the concerned sensors and will make the aggregate without decrypting data by using additive homomorphic encryption technique. This will ensure security and end-to-end privacy of data. For non-responding nodes the aggregator will append the ID of that node and will make the status of that node to zero which mean that this node will no longer be considered in later rounds. Here SEEDA encrypt zero data for non-responding nodes and add with the message which is more energy consuming process in terms of communication and computation. Now the aggregator will make message by appending IDs of non-responding nodes with aggregated data. The message will be sent to the aggregator at next higher level of the tree. Aggregator at this level collect messages form concerned aggregators and make aggregate and forwards the message to higher level. This process will be continued until message reached to the base station (BS) or Sink.

4.3.2 EDAS Algorithm:

```
Step 1: Sensor Si sense data, encrypt using function $ci = xi + ki\,(mod\,M)$
Step 2: Sensor send data to the concerned AGG
Step 3: AGG checks: if sensor node is responding
        a). Yes: perform aggregation function i-e $C_{ag\,i} = C_{ag\,i} + c_i$
        b). No: Extract ID and make status of node to zero
Step 4: Make message M: aggregated data appending with IDs of non-responding nodes
          i-e $M = C_{agi} || ID_i$
Step 5: Send message to AGG on next level
Step 6: AGG: make aggregate of the messages received from other AGGs
Step 7: Repeat steps 5 and 6 until message reach BS
Step 8: END
```

Explanation: In step 1 of the EDAS algorithm, sensor nodes will sense or detect data from the environment. This detected data will be the encrypted using the function mentioned in the step 1.

In step 2, sensors will send the encrypted data to the concerned aggregator AGG.

In step 3 of the algorithm, the aggregators will check that all sensors are responding or not. For responding sensor, the aggregator will make aggregate of their sent data using Additive Homomorphic Encryption (already mentioned in chapter 1). For non-responding sensors, aggregator will find the ID of that sensor and will make the status of that node to zero. Here zero status means that this node will be marked as permanently non-responding and no processing for that node in future.

In step 4, aggregator will make message by appending aggregated with the IDs of non-responding sensors.

Step 5 states that the aggregators at level (h-1) of the tree will send the message to aggregators at level (h-2). Where 'h' is the height of tree.

In step 6, aggregators at level (h-2) will make aggregate of messages received from aggregators of underlying level.

Step 7 will continue repetition of step 5 and 6 till the message reach the base station.

Step 8 is the end of EDAS algorithm.

The above discussion is also more clear in EDAS flow chart.

4.3.3 Flow Chart for EDAS:

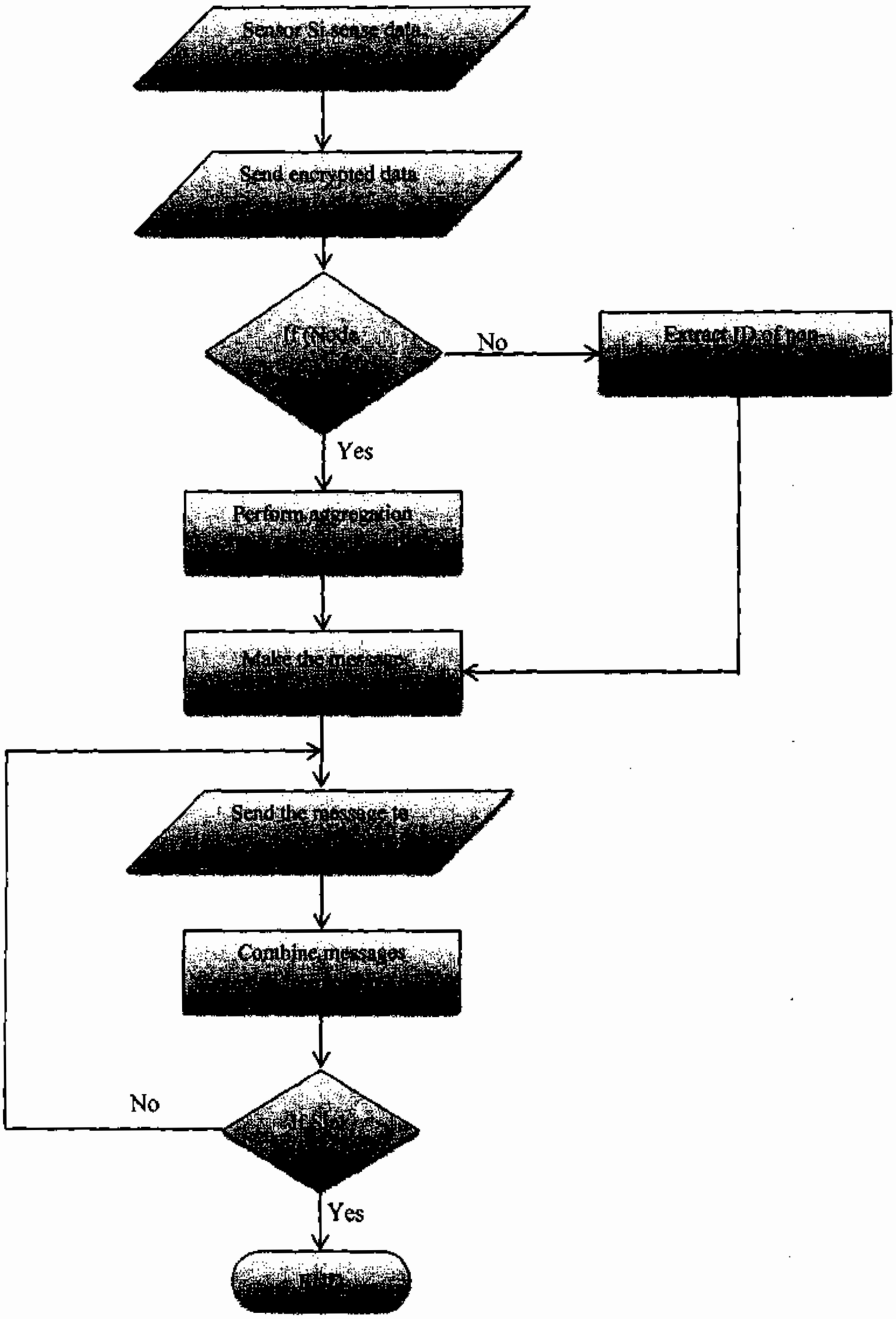


Figure: 4.2 EDAS Flow Chart

CHAPTER NO: 5

IMPLEMENTATION

5. Implementation

This chapter gives a detailed overview of implementation. At the start, implementation of proposed model Efficient Data Aggregation Scheme in Secure Tree Based Wireless Sensor Network (EDAS) is explained. Next to it, the implementation of existing model Secure End-to-End Data Aggregation Protocol (SEEDA) is explained.

5.1 Implementation Environment

The EDAS and SEEDA simulators are developed by using Java JDK 1.7. NetBeans IDE is used for Java JDK 1.7. Java is a programming language. It contains fundamental technology that enables high-technology programs containing utilities, games, mobile applications and business applications.

5.2 Implementation Scenario

Here is the Efficient Data Aggregation Scheme in Secure Tree Based Wireless Sensor Network (EDAS) Simulator shown is figure 5.1. This is just a control form where one can select or call either EDAS simulator or SEEDA simulator. By clicking the button labeled as EDAS, the corresponding simulator is called and displayed on the screen. By clicking the button labeled as SEEDA, the corresponding SEEDA simulator is displayed on the screen. And finally clicking the View Graph button, the Graph Form is displayed. This main EDAS Simulator form can be closed by clicking Exit from the File menu.

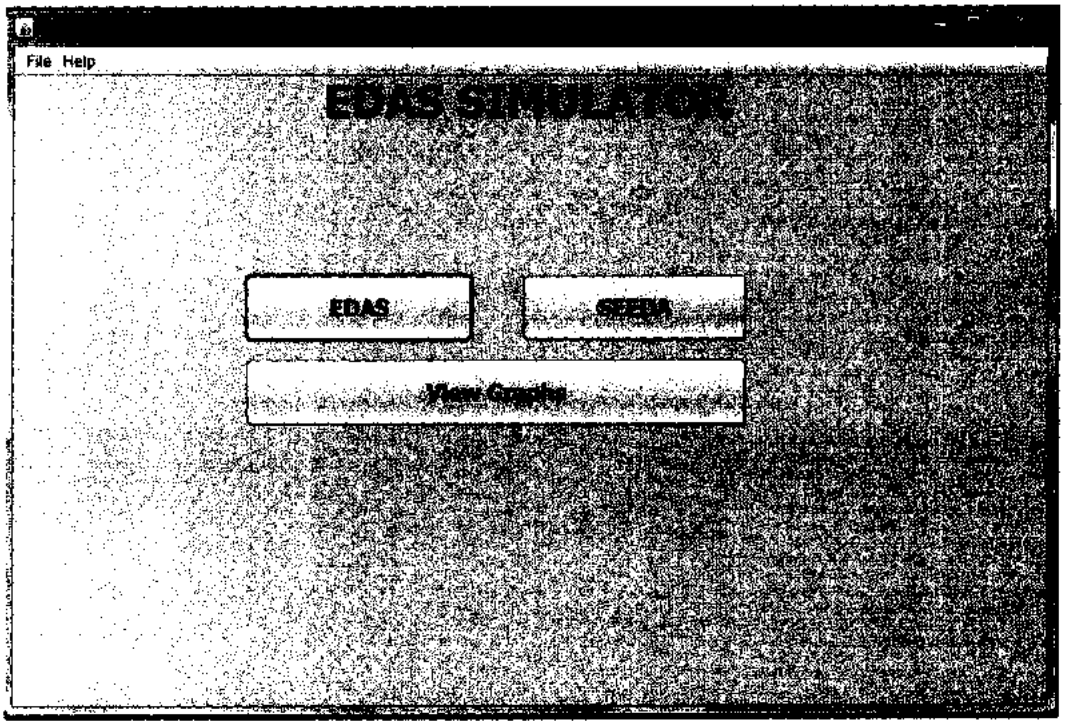


Figure 5.1: Main Form/EDAS Simulator

5.2.1 EDAS Simulator

In figure 5.2 EDAS Simulator is shown. In the panel of the simulator, a tree based network shown. At Level 0 of the tree is a node called Base Station. Aggregators are at Level 1 and Level 2. The leaf nodes are the Sensors. There are two radio buttons at bottom. The one labeled as 10% Non-Responding Nodes, when selected then 90% nodes in the network will be responding nodes and 10% will be non-responding. The radio button labeled as 30% Non-Responding Nodes, when selected then 70% of the nodes will be responding and 30% will be non-responding. There is a control button labeled as Start. The function of this button is to start the simulation.

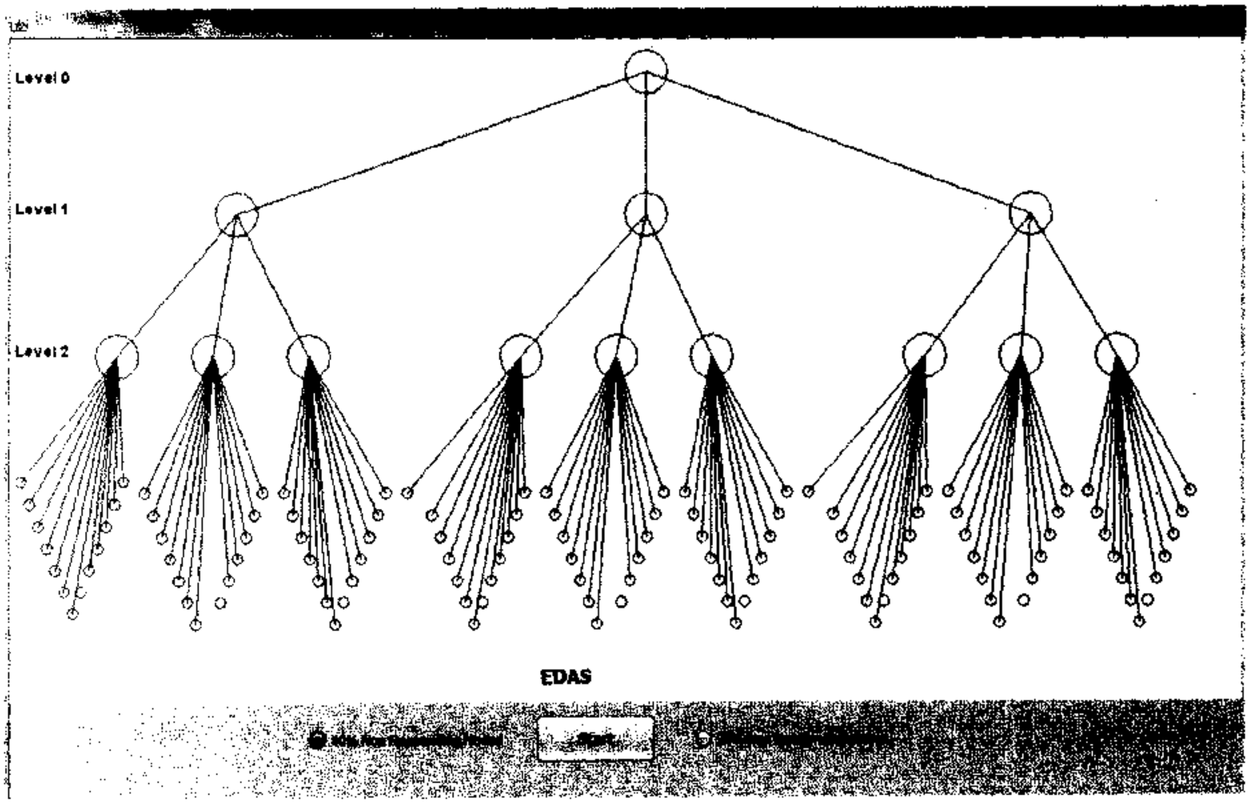


Figure 5.2: EDAS Simulator

5.2.2 SEEDA Simulator

In figure 5.3 SEEDA Simulator is shown. In the panel of the simulator, a tree based network shown. At Level 0 of the tree is a node called Base Station. Aggregators are at Level 1 and Level 2. The leaf nodes are the Sensors. There are two radio buttons at bottom. The one labeled as 10% Non-Responding Nodes, when selected then 90% nodes in the network will be responding nodes and 10% will be non-responding. The radio button labeled as 30% Non-Responding Nodes, when selected then 70% of the nodes will be responding and 30% will be non-responding. There is a control button labeled as Start. The function of this button is to start the simulation.

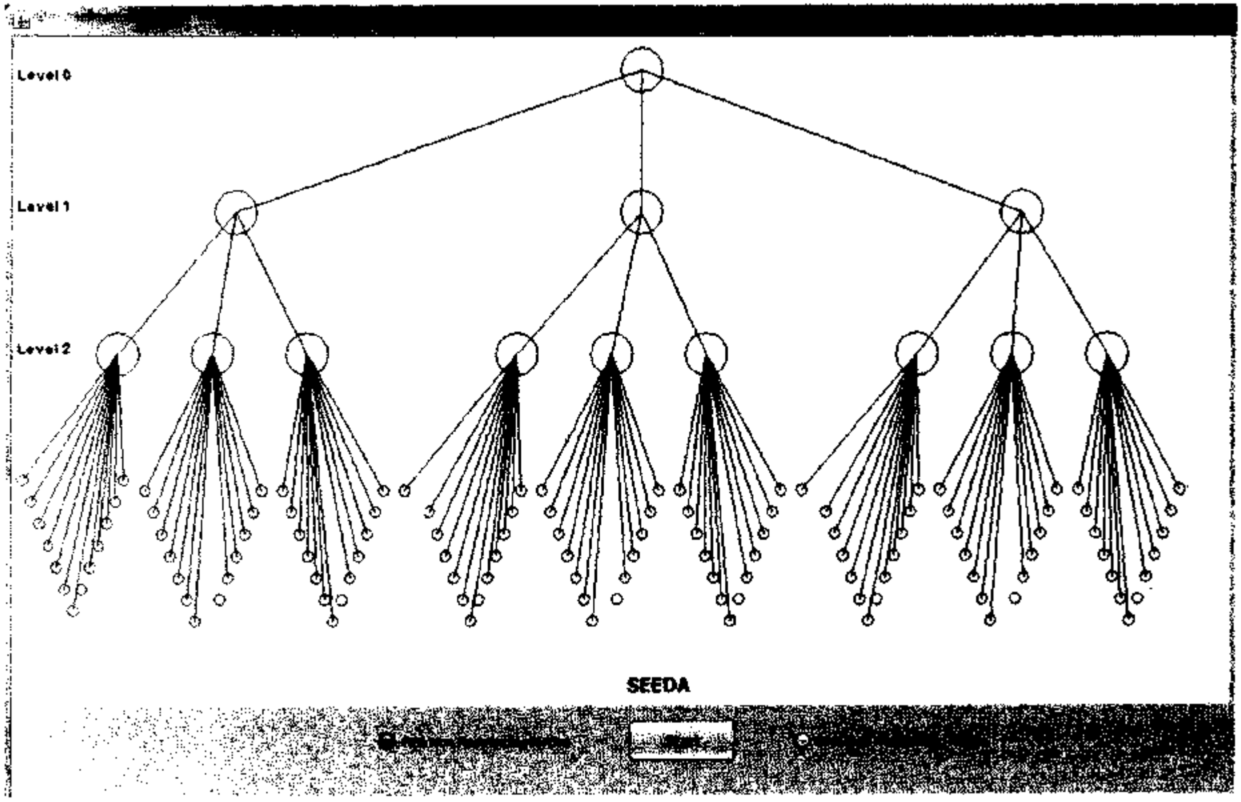


Figure 5.3: SEEDA Simulator

5.3 EDAS Parameters

The proposed EDAS scheme has 108 Sensors and 12 Aggregators 'AGG' and a single Base Station. Network used is heterogeneous i-e having different types of nodes.

The details of parameters used in EDAS scheme are given in the table 5.1.

Table 5.1 EDAS Parameters

S.NO	**Parameter**	**Value**
1	Data packet size	232 bits
2	Header Size	56 bits
3	Data rate	250 k-bits/s
4	Orientation	2D
5	Propagation Environment	Free Space
6	Sensor Nodes	108
7	Aggregators	12
8	Base Station	1
9	Transmission range of sensor node	Up to 300 m
10	Radio frequency	915 MHz
11	Transmit mode Energy consumption	0.5 Milli-Joule
12	Receive mode Energy consumption	0.2 Milli-Joule
13	Computational Energy consumption	0.4 Micro-Joule

CHAPTER NO: 6

RESULTS AND CONCLUSION

6.1 Results

This chapter is related with simulation methodology such as assumption of the simulation, mechanism of the simulation and simulation results. The results are explained with the help of graphs. The results are used to show the efficiency and reliability of the existing technique SEEDA and that of the proposed scheme EDAS in terms of communication, computation, energy, latency and accuracy.

For the simulation of both schemes, we used NetBeans as IDE. For building custom simulator java is used as programming language and JDK 1.7 is used as compiler. Results of the existing solution are studied, observed deeply and thoroughly. Then the results of the proposed solution are compared with the results of the existing technique. The results of the proposed scheme are better as compared to the existing technique.

6.2 Comparison Parameters

We considered the following metrics of performance:

1. Number of bits/data sent to Base Station
2. Computational energy consumed
3. Energy consumed by the sensors nodes in the network
4. Network lifetime
5. Latency/Data Delay
6. Data Accuracy

The first four graphs are plotted when 10% of sensors are non-responding. The last four graphs are plotted when 30% of the sensors are non-responding. Finally the graph of network latency is plotted for various percentages of non-responding nodes.

6.2.1 Number of Bits Sent to the Base Station (10% Non-Responding Nodes)

The graph shown in figure 6.1 is used to represent number of bits sent to the Base Station when 10% of the sensors do not respond and 90% sensors respond. In graph the horizontal axes represents number of rounds while the vertical axes represents number of bits/data sent to the base station. This graph compares EDAS with the SEEDA scheme. It is clear from the plot that in SEEDA scheme, the base station receives more than 150 extra bits than EDAS scheme in each round. Round means data sensed by all the responding sensors and sent to the base station. These extra bits are produced in SEEDA, because the aggregators compute cipher text c_j for all non-responding nodes and add this data with aggregated data. The red line in the plot is for SEEDA scheme while the blue line is for EDAS scheme.

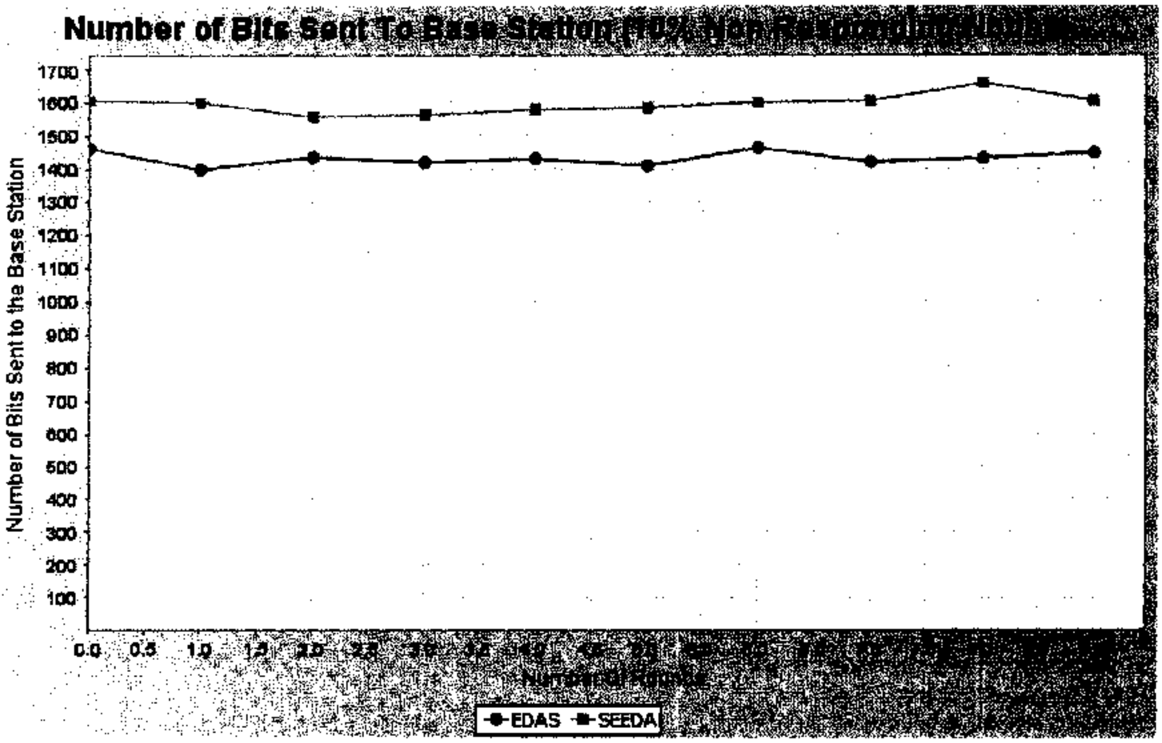


Figure 6.1: Number of Bits sent to Base Station (10% non-responding nodes)

6.2.2 Computational Energy Loss (10% Non-Responding Nodes)

Computational Energy Loss Graph is shown in figure 6.2. This graph is used to represent extra energy loss due to extra computational in SEEDA scheme when 10% of nodes do not respond. In the graph the horizontal axes represents number rounds while the vertical axes represents computational energy loss in micro-joules. The graph shows that EDAS (blue line) scheme is more efficient in energy than the SEEDA scheme (red line). This extra energy loss is the result of extra computation of cipher text c_j for all non-responding nodes at the level-2 Aggregators. This computation is not performed in EDAS. The IDs of non-responding sensors are produced and sent with aggregated result.

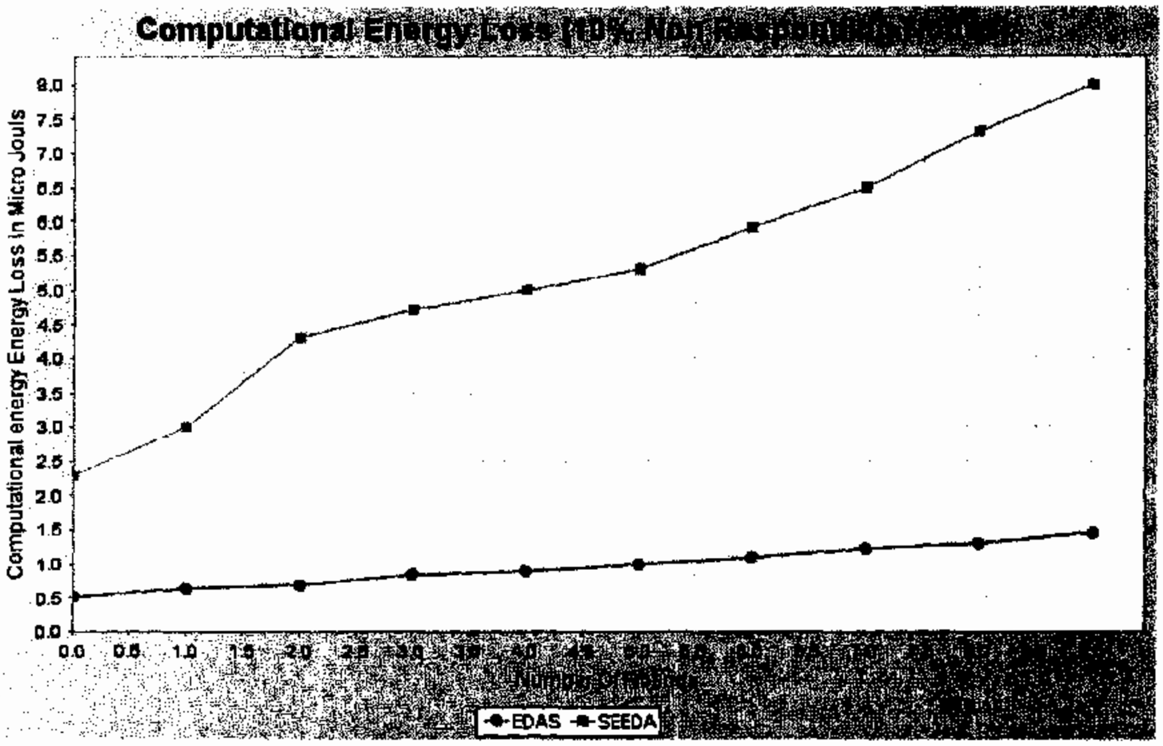


Figure 6.2: Computational Energy Loss (10% non-responding nodes)

6.2.3 Network Energy Consumed (10% Non-Responding Nodes)

This graph is shown in figure 6.3 and is used to represent energy consumption of the overall network when 10% of the sensors do not respond. In the graph the horizontal axes represents number rounds while the vertical axes represents energy consumed in milli-joules. The graph shows that EDAS scheme is more efficient than the SEEDA scheme from energy point of view. This extra energy consumption in SEEDA is because of the extra more than150 bits sent across the network due to non-responding nodes. According to the literature data sending consumes 70% energy of the sensor network. The red line in the plot is for SEEDA scheme while the blue line is for EDAS scheme.

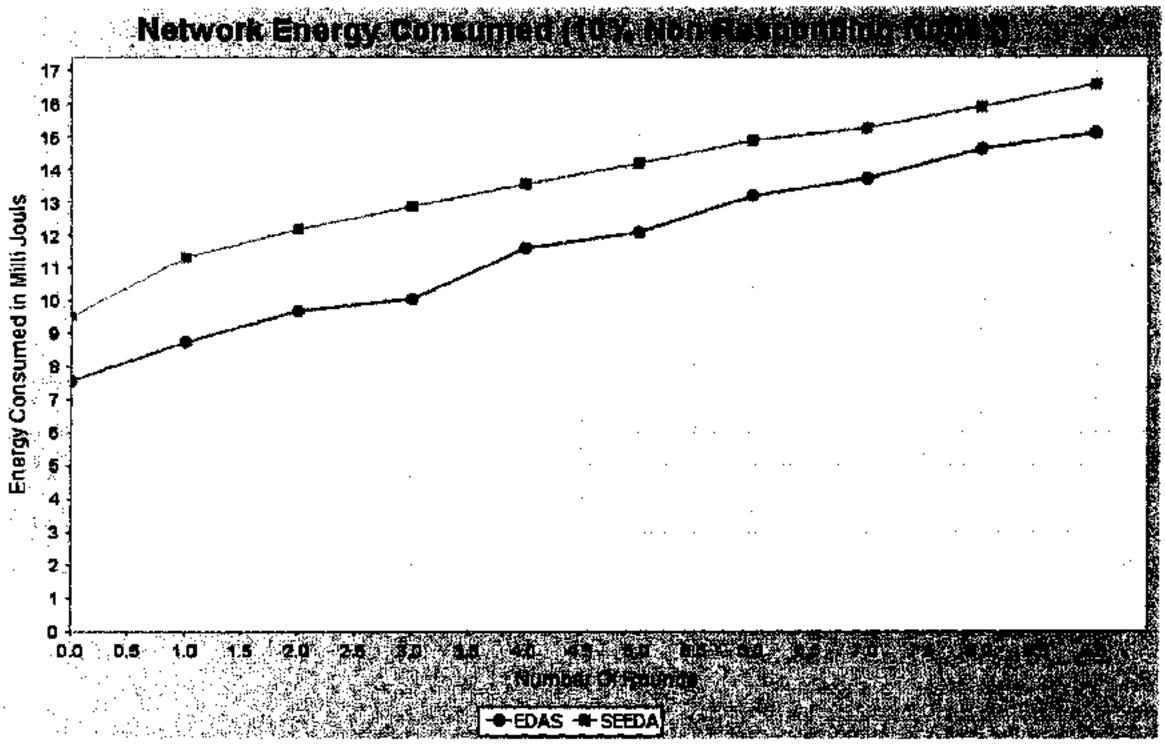


Figure 6.3: Network Energy Consumed (10% non-responding nodes)

6.2.4 Network Lifetime (10% Non-Responding Nodes)

This graph is used to represent overall lifetime of the network in seconds when 10% of sensors are non-responding and 90% are responding. The difference between lifetimes of EDAS (blue line) and SEEDA (red line) clearly shows that EDAS network lifetime is more by 90 seconds then corresponding SEEDA network. The vertical axes represents energy of the network consumed in milli-joule and the horizontal axes represents lifetime of the network in seconds. This time efficiency of EDAS is due to the fact that SEEDA has more computational and communication costs. This results in faster energy consumption of the SEEDA network and it dies earlier. The graph is shown in figure 6.4.

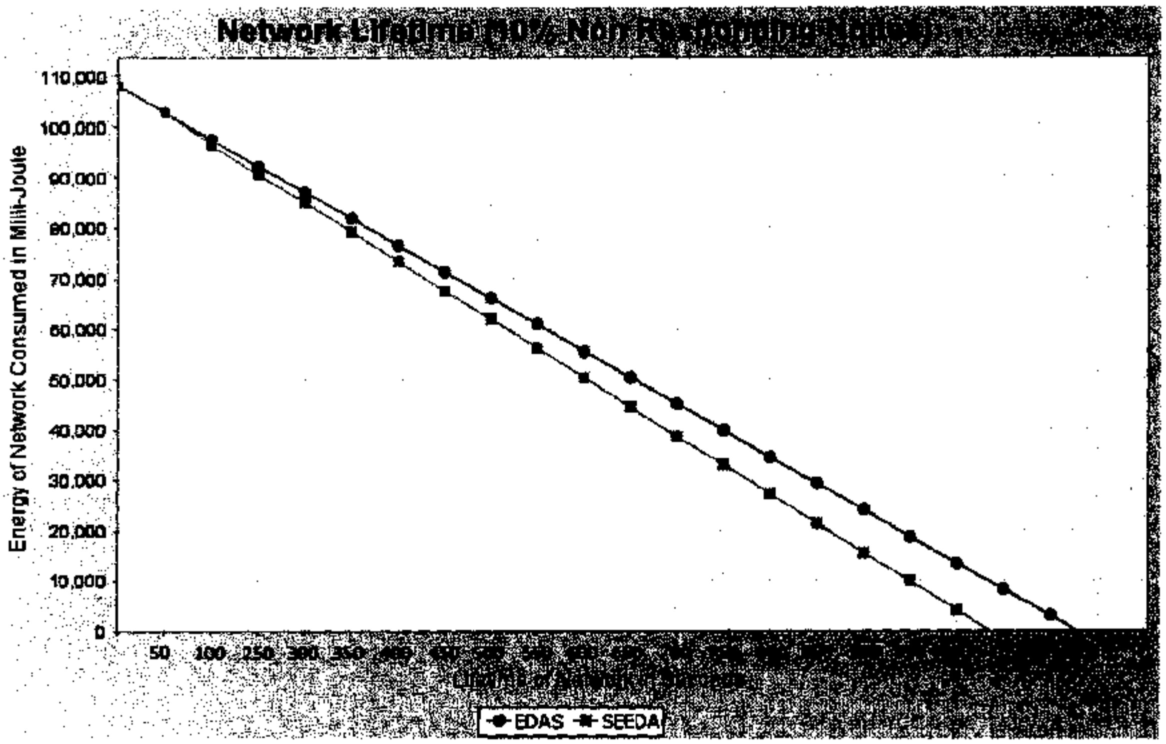


Figure 6.4: Network Lifetime (10% non-responding nodes)

6.2.5 Number of Bits Sent to the Base Station (30% Non-Responding Nodes)

This graph is shown in figure 6.5 and is used to represent number of bits sent to the base station when 30% of the sensors do not respond. In the graph the horizontal axes represents number rounds while the vertical axes represents number of bits/data sent to the base station. The graph shows that EDAS scheme is more than 400 bits efficient than the SEEDA scheme. These extra bits are generated in SEEDA scheme due to computation of cipher text for non-responding nodes. These extra bits/data will increase when the number of non-responding nodes exceed than 30%. The graph is shown in figure 6.5.

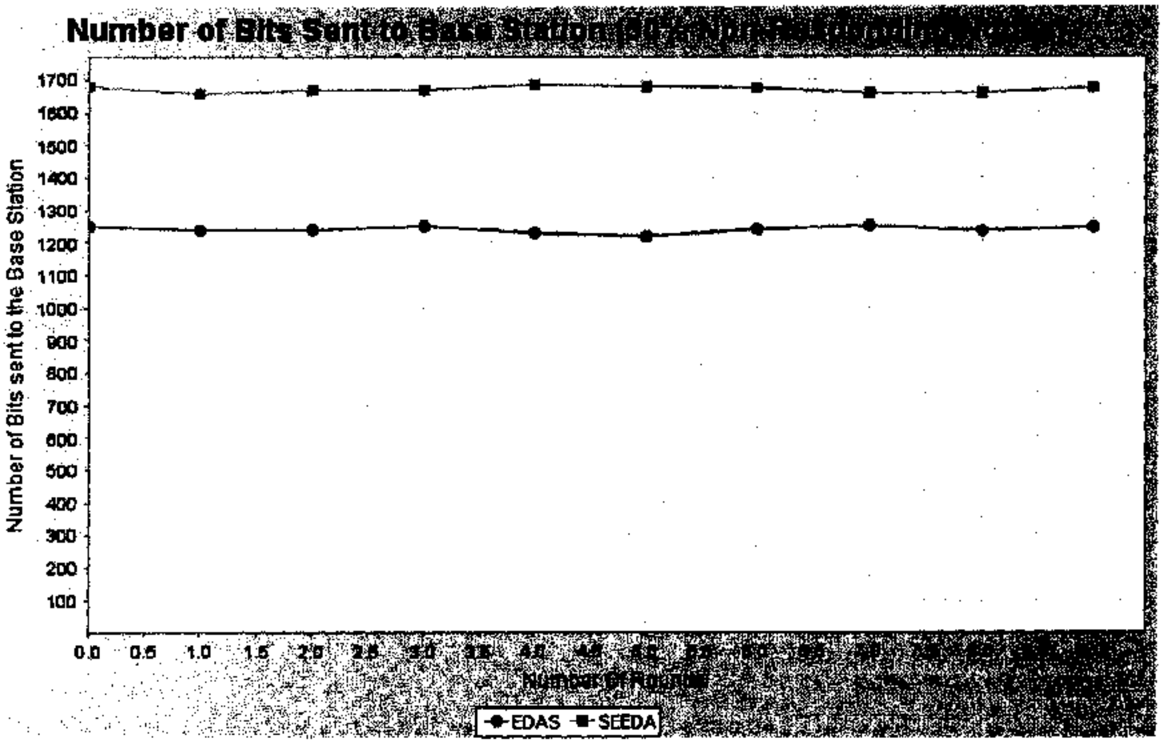


Figure 6.5: Number of bits sent to Base Station (30% non-responding nodes)

6.2.6 Computational Energy Loss (30% Non-Responding Nodes)

This graph is shown in figure 6.6 and is used to represent computational energy loss when 30% of the sensors are non-responding and 70% are responding sensors. In the plot, the horizontal axes represents number of rounds while the vertical axes represents computational energy loss in micro-joules. The graph shows that EDAS scheme is more efficient in computational energy than the SEEDA scheme. This extra energy loss is the result of extra computation for non-responding nodes at the level 2 Aggregators. The red line in the plot is for SEEDA scheme while the blue line is for EDAS scheme.

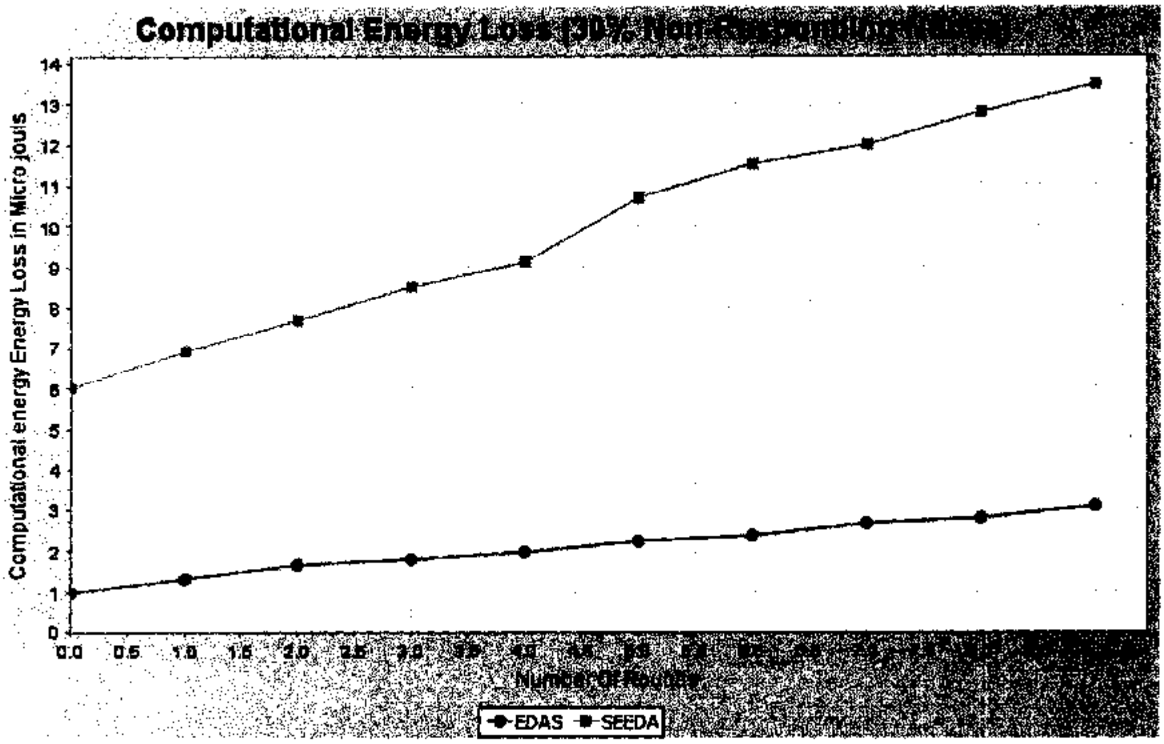


Figure 6.6: Computational Energy Loss Graph (30% non-responding nodes)

6.2.7 Network Energy Consumed (30% Non-Responding Nodes)

This graph is shown in figure 6.7 and is used to represent energy consumption of the overall network when 30% of the sensors do not respond. In the graph the horizontal axes represents number rounds while the vertical axes represents energy consumed in milli-joules. The graph shows that EDAS scheme is more efficient than the SEEDA scheme from energy point of view. This extra energy consumption in SEEDA scheme is due to the fact that more than 400 bits are sent across the network for non-responding nodes. Data transmission consumes 70% energy of the sensor network. The red line in the plot is for SEEDA scheme while the blue line is for EDAS scheme.

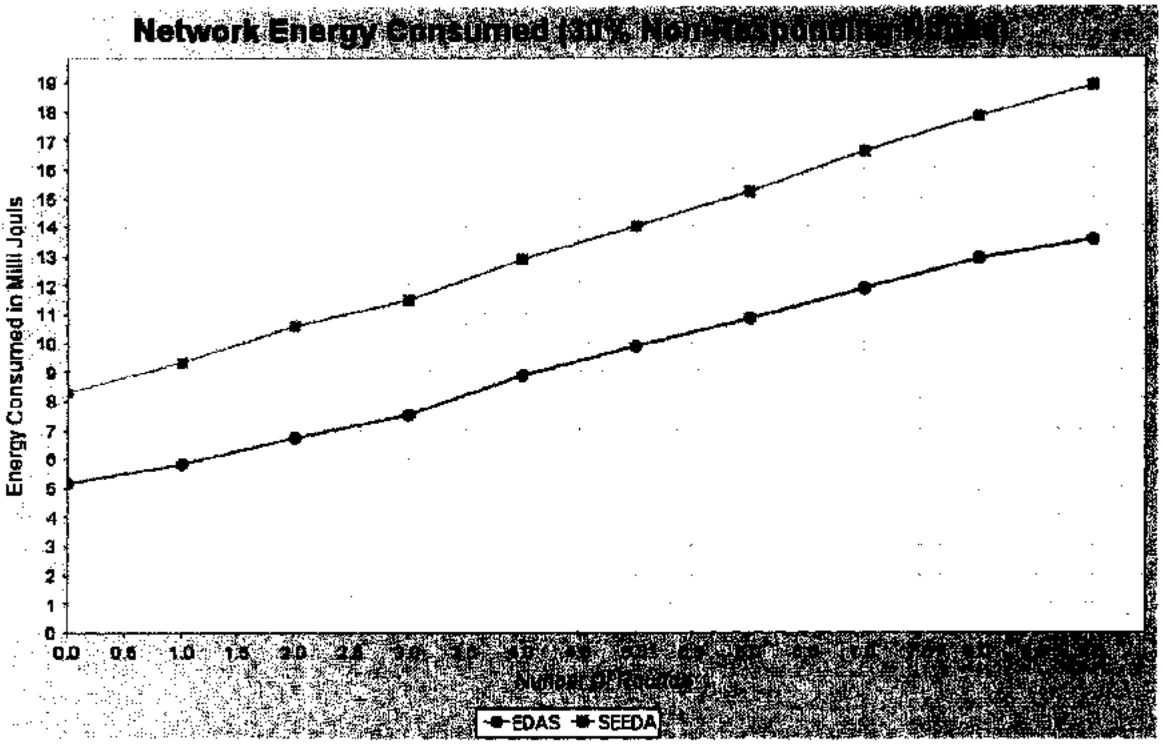


Figure 6.7: Network Energy Consumed (30% non-responding nodes)

6.2.8 Network Lifetime (30% Non-Responding Nodes)

This graph is used to represent overall lifetime of the network in seconds when 30% of sensors are non-responding. The difference between lifetime of EDAS (blue line) and SEEDA (red line) clearly shows that the lifetime of EDAS scheme is more by 270 seconds than the SEEDA scheme. The vertical axes represents energy of the network consumed in milli-joule and the horizontal axes represents lifetime of the network in seconds. This time efficiency of EDAS is due to the fact that SEEDA has more computational and communication costs. This results in faster energy consumption of the SEEDA network and it dies earlier.

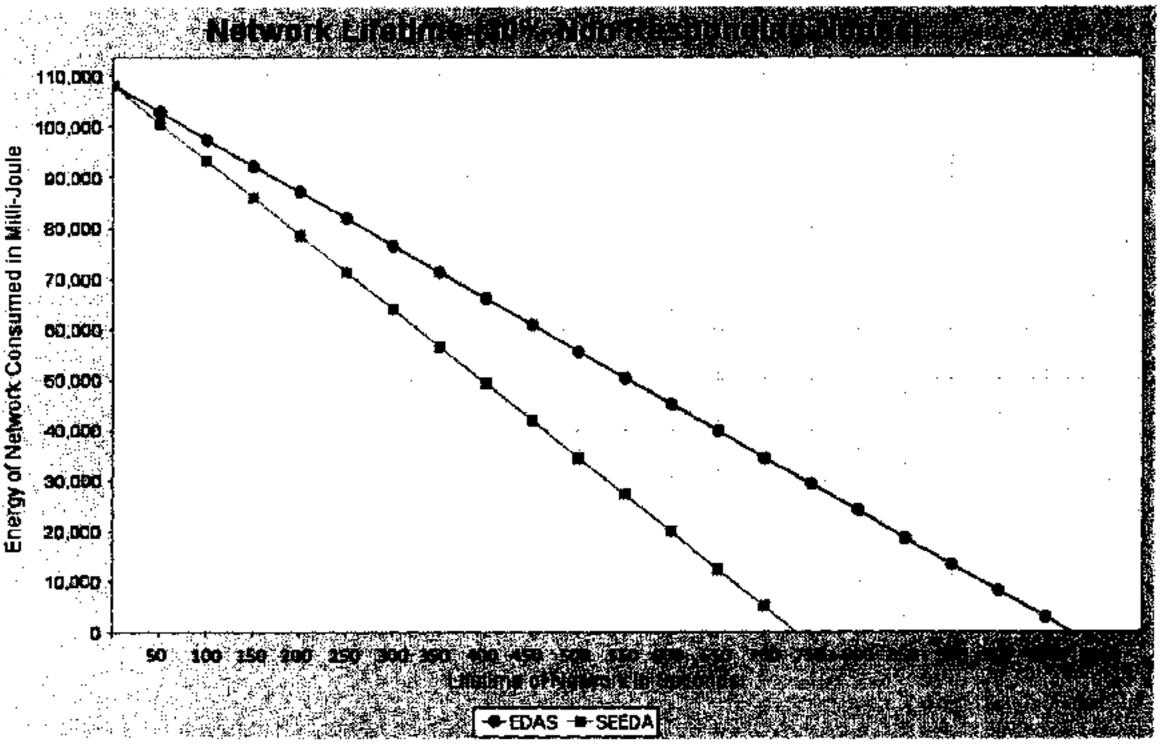


Figure: 6.8 Network Lifetime (30% non-responding nodes)

6.2.9 Latency / Data Delay (10% - 70%)

Data takes time when sent from sensor node to the base station. This delay of data from sensors to base station is called Latency. The latency of SEEDA and EDAS are shown in figure 6.9. Vertical axes represents time delay in milli-seconds per round. The horizontal axes represents number of non-responding nodes in percentage (from 10% to 70%). The red line in the plot is for SEEDA scheme while the blue line is for EDAS scheme. The graph clearly shows that SEEDA protocol has lager data delay than EDAS. This is due to the fact that SEEDA has higher computational cost. Higher computational cost results in lager delays to data.

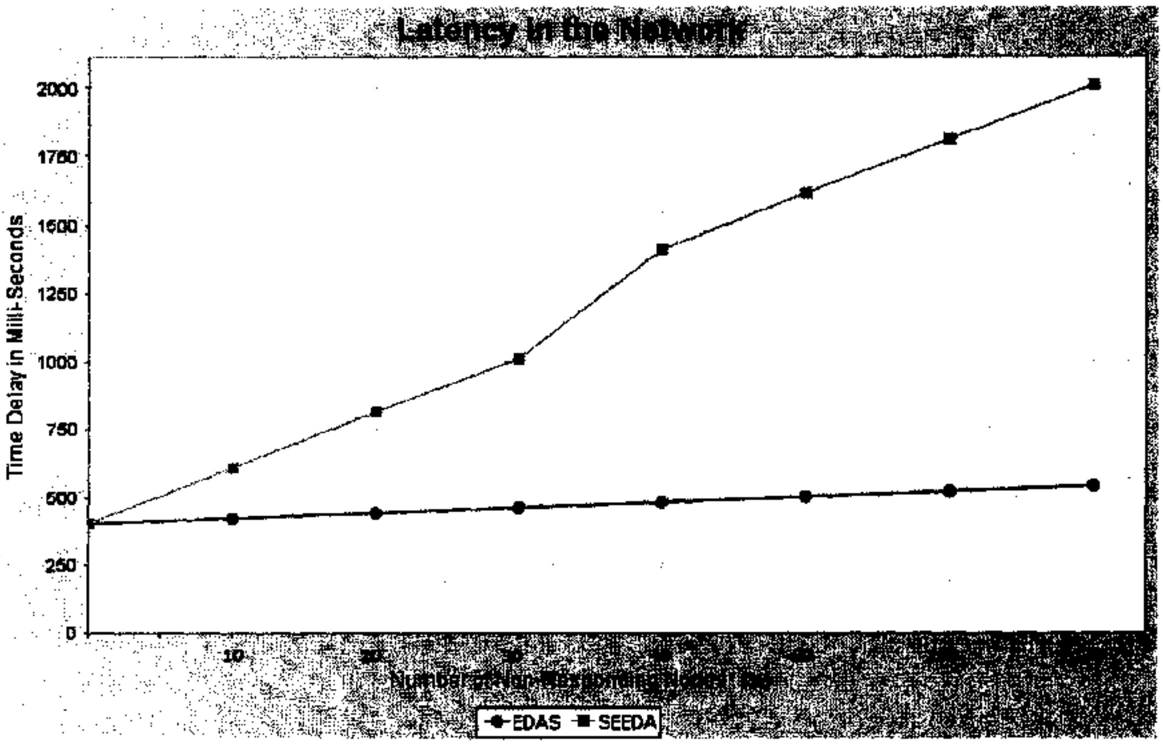


Figure: 6.9 Latency / Data Delay (10% - 70% Non-Responding Nodes)

6.2.10 Data Accuracy (10% – 70% Non-Responding Nodes)

Data Accuracy of the EDAS and SEEDA schemes are shown in figure 6.10. In the graph the vertical axes represents data accuracy which has standard value of 1. The horizontal axes represents number of non-responding nodes in percentage (10% to 70%). The blue line in the plot represents EDAS scheme while the red represents SEEDA scheme. It clear from the graph that as numbers as of non-responding increase the accuracy of both the schemes decreases. But the accuracy of SEEDA is less than that of EDAS scheme.

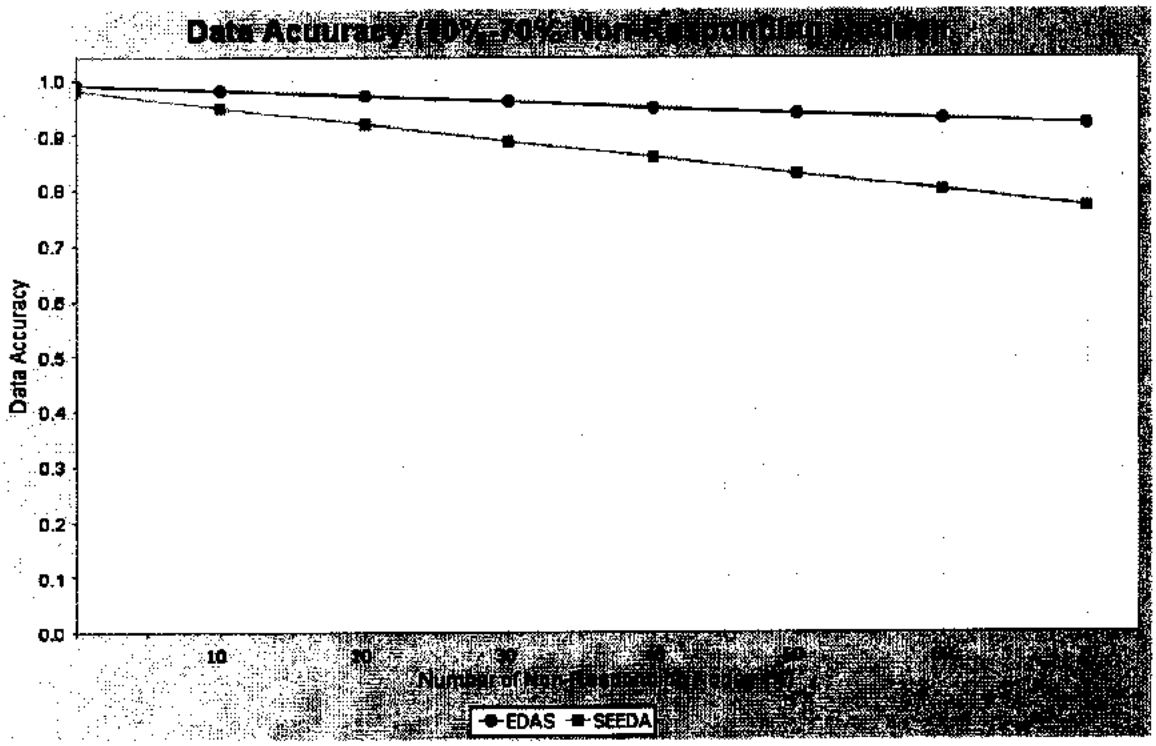


Figure 6.10 Data Accuracy (10% - 70%)

6.3 Conclusion

In this thesis new scheme, Efficient Data Aggregation Scheme in Secure Tree Based Wireless Sensor Network (EDAS) is proposed to achieve secure but end-to-end data aggregation. The proposed scheme ensures end-to-end privacy of data and less number of bits are transmitted from sensors to base station as compared to Secure End-to-End Data Aggregation (SEEDA) scheme. The EDAS protocol adopts the best feature of hop-by-hop data aggregation i-e transmission of less number of bits/data from sensors to base station. EDAS also adopts the best characteristic of end-to-end protocols that is end-to-end data privacy. EDAS protocol is also more efficient in energy consumption than SEEDA. Again EDAS is also more efficient in computational energy cost than that of corresponding SEEDA scheme. One unique feature is that the efficiency of EDAS increases as the number of non-responding nodes in network increases. Simulation results shows that the proposed scheme is more efficient in terms data accuracy. Our proposed scheme has smaller latency/data delay than the existing scheme. The lifetime of the EDAS network is more than that of SEEDA scheme.

References:

[1] K. Akkaya, M.Younis, "A survey on routing protocols for wireless sensor networks" International Journal on Ad Hoc Networks 3, pp. 325–349, 2005.

[2] J. Yick, B. Mukherjee, D. Ghosal, "Wireless Networks Survey", The International Journal of Computer and Telecommunications Networking, Vol. 52, pp. 2292-2330, 2008.

[3] S. Peter, D. Westhoff, C. Castelluccia, "A Survey on the Encryption of Convergecast Traffic with In-Network Processing", IEEE Transactions on Dependable and Secure Computing, vol. 7, pp. 20-34, March 2010.

[4] K. Maraiya, K. Kant, N. Gupta, "Architectural Based Data Aggregation Techniques in Wireless Sensor Network: A Comparative Study" International Journal on Computer Science and Engineering (IJCSE), Vol. 3, pp. 1131-1138, March 2011.

[5] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, "A survey on sensor networks", IEEE Communications Magazine 40 (8), pp. 104–112, 2002.

[6] http://users.ece.gatech.edu/~sriram/research.htm

[7] I.F. Akyildiz, E.P. Stuntebeck, "Wireless underground sensor networks: research challenges", Ad-Hoc Networks 4, pp. 669–686, 2006.

[8] I.F. Akyildiz, D. Pompili, T. Melodia, "Challenges for efficient communication in underwater acoustic sensor networks", ACM Sigbed Review 1 (2), pp. 3–8, 2004.

[9] I.F. Akyildiz, T. Melodia, K.R. Chowdhury, "A survey on wireless multimedia sensor networks", Computer Networks Elsevier 51, pp. 921–960 2007.

[10] A. Sinha, A. Chandrakasan, "Dynamic power management in wireless sensor networks", IEEE Design and Test of Computers 18 (2), pp. 62–74, 2001.

[11] A. Chandrakasan, R. Amirtharajah, S. Cho, J. Goodman, "Design considerations for distributed micro sensor systems", Proceedings of IEEE Customs Integrated Circuits Conference (CICC), pp. 279–286, May 1999.

[12] A. Perrig, R. Szewczyk, J.D. Tygar, V. Wen, D.E. Culler, "SPINS: Security protocols for sensor network", Journal of Wireless Networks 8 (5), pp. 521–534, 2002.

[13] Tanya R, Shiuphyng S, and Shankar S, "Taxonomy of Security in Sensor Networks", In proceedings of the first IEEE International conference on System Integration and Reliability Improvement, SIRI' 06, pp. 43-48, December 2006.

[14] Macros A, Claudionor N, Diogenes C, Jose M, "Survey on Wireless Sensor Network Devices," Proceedings of the 9th IEEE conference on Emerging Technology and Factory Automation, ETFA' 03, vol. 1, pp. 537-544, September 2003.

[15] B. Krishnamachari, D. Estrin, S. Wicker, "The impact of data aggregation in wireless sensor networks", Proceedings of the 22nd International Conference on Distributed Computing Systems Workshops, pp. 1-4, 2002.

[16] Kiran M, Kamal K, Nitin G, "Architecture based data aggregation Techniques in wireless sensor network: A comprehensive study", International Journal on Computer Science and Engineering (IJCSE), vol. 3, pp. 1131-1138, March 2011.

[17] P.N. Renjith, E. Baburaj, "An Analysis on Data Aggregation in wireless sensor network", International Conference on Radar, Communication and Computing (ICRCC), pp. 62-71, December 2012.

[18] Priyanka K, Shah V, Kajal V, Shukla, "Secure Data Issues in Wireless Sensor Network: A Survey", Journal of Information and Communication Technologies, vol. 2, pp. 10-17, Jan. 2012.

[19] C. Castelluccia, E. Mykletun, G. Tsudik, "Efficient Aggregation of encrypted data in wireless sensor networks", Transactions of Mobi Quitous, pp. 1-9, 2005.

[20] H. Cam, S. Ozdemir, P. Nair, and D. Muthuavinashiappan, "ESPDA: energy-efficient and secure pattern-based data aggregation for wireless sensor networks", IEEE Sensors–The Second IEEE Conference on Sensors, pp. 446-455, Oct. 2003.

[21] H. Sanli, S. Ozdemir, and H. Cam, "SRDA: Secure Reference-Based Data Aggregation Protocol for Wireless Sensor Networks", Proc. IEEE 60th Int'l Conf. Vehicular Technology (VTC 04-Fall), vol. 7, pp. 4650-4654, Sept. 2004.

[22] Y. Yang, X. Wang, S. Zhu, and G. Cao, "SDAP: A Secure Hop-by- Hop Data Aggregation Protocol for Sensor Networks", ACM Trans. Information and System Security (TISSEC), vol. 11, pp. 1-43, 2008.

[23] D. Westhoff, J. Girao, and M. Acharya, "Concealed Data Aggregation for Reverse Multicast Traffic in Sensor Networks: Encryption, Key Distribution, and Routing Adaptation", IEEE Trans. Mobile Computing, vol. 5, pp. 1417-1431, Oct. 2006.

[24] S. Ozdemir, "Concealed Data Aggregation in Heterogeneous Sensor Networks Using Privacy Homomorphism", Proc. IEEE Int'l Conf. Pervasive Services, pp. 165-168, July 2007.

[25] S. Ozdemir and Y. Xiao "Hierarchical concealed data aggregation for wireless sensor networks", In: Proceedings of the Embedded Systems and Communications Security Workshop in conjunction with IEEE (SRDS), pp. 1-5, 2009.

[26] C. Castelluccia, C-F Chan, E. Mykletun, G. Tsudik, "Efficient Provably secure Aggregation of encrypted data in wireless sensor networks", ACM Transactions on Sensor Networks, vol. 5, pp. 1-36, May 2009.

[27] A.S. Poornima, B.B. Amberker, "SEEDA: Secure end-to-end data aggregation in Wireless Sensor Networks", In Proceeding of the 7th International Conference of Wireless and Optical Communications Networks (WOCN), pp. 1-5, 2010.

[28] Josna J, Manoj Kumar S, Joyce J, "Energy Efficient Recoverable Concealed Data Aggregation in Wireless Sensor Networks", IEEE International Conference on Emerging Trends in Computing, Communication and Nanotechnology (ICECCN), pp.322-329, 2013.

[29] Joyce J, M .Princy, Josna J, "PEPPDA: Power Efficient Privacy Preserving Data Aggregation for Wireless Sensor Networks", IEEE International Conference on Emerging Trends in Computing, Communication and Nanotechnology (ICECCN), pp.330-336, 2013.

[30] Abhilash L N, D. Goenka, C. Kumar, "Dynamic Data Aggregation for Energy Optimization in Multi-Hop Wireless Sensor Networks", IEEE International Advance Computing Conference (IACC), pp. 143-148, 2014.

[31] S. Ozdemir, Y. Xiao, "Secure Data Aggregation in Wireless Sensor Network: A Comprehensive Overview", Computer Networks, vol. 52, pp. 2022-2037, 2009.