ENERGY-EFFICIENT PROTOCOLS FOR SELECTIVE DATA AGGREGATION IN WIRELESS SENSOR NETWORKS

by

Suresh Lalwani

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

ARIZONA STATE UNIVERSITY

December 2002

ENERGY-EFFICIENT PROTOCOLS FOR SELECTIVE DATA AGGREGATION IN WIRELESS SENSOR NETWORKS

by

Suresh Lalwani

has been approved

December 2002

APPROVED:

, Chair

Supervisory Committee

ACCEPTED:

Department Chair

Dean, Graduate College

ABSTRACT

Sensor networks are easy to deploy and are used to gather information for a wide variety of applications. These networks are energy constraints and, thus, it is important for them to gather and transmit data in an energy-efficient manner. Gathering data in an energy-efficient way is also critical for the operation of sensor networks for a long duration of time. With the recent developments in sensor technology, it is possible to use sensors for a wide range of applications. One such application emphasized in this research work is a bio-medical application. In this application, smart bio-compatible sensors are implanted within the human body for various health-monitoring purposes. This research work proposes five protocols for data gathering. These protocols are combinations of schemes based on contention, time division multiple access (TDMA), and code division multiple access (CDMA). Some of them are adopted from the protocols available for multicasting in wireless networks. In the first two protocols, sensors respond directly to the base-station, while in the remaining protocols, sensors are grouped into clusters and they respond to the base-station through their cluster-heads. Simulation results show that the protocol using a combination of TDMA-CDMA along with clustering (i.e., protocol 5) performs better than the other protocols in terms of energy and delay as metrics. This protocol performs 1.5 to 4 times better than other protocols in terms of energy consumption and 1.5 to 6 times better in terms of delay. All TDMA-based protocols perform better than contention-based protocols in terms of communication overhead and session efficiency. The TDMA-based protocols have negligible or no communication overhead and 100% session efficiency, whereas the contention-based protocols have a communication overhead between 300% to 400% and a session efficiency between 90% to 94%. All these results were obtained from a wide range of simulation scenarios and node configurations. The simulation results were further validated by using mathematical analysis. The analytical results support the simulation results. The best performing protocol is, finally, compared with Power-Efficient Gathering in Sensor Information Systems (PEGASIS). Comparison results show that this protocol outperforms PEGASIS, and as the number of sensors increases, the relative performance of PEGASIS deteriorates.

ACKNOWLEDGMENTS

I thank Dr. Sandeep Gupta for guiding me and providing valuable suggestions during the period of this research work. He has always been eager to discuss the technical problems and took valuable time from his busy schedule to further the progress of this research work.

I am thankful to Dr. Arunabha Sen and Dr. Goran Konjevod for agreeing to be part of my thesis committee and providing their valuable time and suggestions. I thank National Science Foundation for funding this research work in parts under the grants number ANI-0086020 and ANI-0196156. I would also like to thank Valliappan Annamalai, Chetan Reddy, Sriram Cherukuri and Yashwanth Prakash for helping me with the Network Simulator, Perl scripts and Lower layer networking details. I wish to thank my friends for proof-reading this research document.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

Data aggregation is a process of collecting data from a set of nodes to a central node. Collecting data from a set of nodes can be viewed as the reverse process of multicasting. Multicasting is defined as a process of data dissemination to a select group that subscribes for the service. In other words, multicasting is “one to many” and collecting data from a set of nodes is “many to one”. Data aggregation involves both multicasting as for requesting data from the set of nodes and reverse of multicasting as for collecting data from the set of nodes. Thus, it can be said that data aggregation involves both “one to many” and “many to one” communications though “many to one” communication is more prominent. In this research work, the set of nodes is referred as a group. Multicasting is an efficient paradigm for sending information from one sender to a group of receivers as compared to unicast, where the sender has to send information individually to each receiver. For this reason, this research work uses multicasting for sending request to the group for selective data aggregation. Data aggregation is widely used in sensor networks. It has been identified as an essential paradigm for wireless routing in sensor networks [1]. Some of the applications of data aggregation in sensor networks are military surveillance, environmental monitoring and medical applications. The application of data aggregation that is given emphasis in this research work is data aggregation in sensors implanted within the human body.

1. Motivation

Data aggregation in sensor networks implanted within the human body is a future project. So as of now, A person can only think of it at the research level and not at the commercial level. Though still in its nascent stages, the recent development in the chip technology and fabrication of bio-compatible materials have made feasible the implantation of sensors within the human body citemarc. Sensors, in general, have the potential to bring a revolution in the fields of health, military and environment applications. Imagine a scenario where the human body has implanted sensors for various applications. Some of the applications may be artificial retinal prosthesis, glucose level monitors, organ level monitors, cancer detectors and general health monitors as described in paper [3]. All these are distinct applications. Sensors responsible for one application have information different from the sensors responsible for some other application. Now consider a scenario in which the base-station wants to send a request to sensors that are responsible for glucose level monitoring, for gathering glucose level information. In this case, the base-station use multicast for data aggregation rather than broadcast that gather data from all the sensors or unicast that require us to gather data from each sensor individually.

2. Constraints and Ways to Improve Energy Efficiency

Sensors have a number of constraints that force the problem statement to be defined in a different manner from the normal wireless applications. The main constraints are the limited source of energy, limited computation power and limited memory. This requires us to develop the protocol that is energy-efficient and use less computation power and memory. Improved energy-efficiency can be achieved by the following steps:-

1. By eliminating or reducing the number of retransmissions required for reliable communications, which, inturn, can be achieved by developing a contention free protocol.
2. By efficiently reusing the channel bandwidth (i.e. by using TDMA or CDMA) [8] [23]. Note that this step will lead to first step i.e. a contention free protocol.
3. By using some data compression technique to compress the data before transmitting from one node to another [7] [8].
4. By exploiting the broadcasting nature of the wireless medium i.e. using single transmission to send data to all receivers that are in the vicinity of the transmitter.
5. By reducing the number of direct communications with the central node. This is achieved by data fusion. Data fusion is a process in which two or more data packets are joined together to form one data packet.

3. System Model

- The system consist of several sensors and a central device known as base-station that is placed at a larger distance as compared to the distance between the sensors.
- Base-station has unlimited power supply where as sensors have limited but consistent supply of power.
- All sensors gather suitable data and send it to the base-station on request.
- The base-station gathers data from a select group of sensors not from all the sensors. The groups are defined based on the requirements of the applications.
- The base-station initiates the sessions for data aggregation at any random time i.e. the data aggregation sessions are sporadic.

4. Assumptions and Definitions

- Groups are already defined and each group has a group-id.
- Group is a select set of sensors.
- Group-id is a unique ID of the group.
- Group-member is a member of the group.
- Base-station has the address information of all sensors i.e. it can contact any sensor directly.
- Base-station does not maintain the group information i.e. which sensor belongs to which group? It only maintains the group-id for a particular group. This assumption is made due to the following reasons:-
    1. In long run, sensors can die and base-station may not know about it, which could give rise to wrong information about the group.
    2. Sensors can be temporarily down during a session.
    3. Maintaining this information is an overhead and also not required since sensors know the group they belong to. The protocols, proposed in this research work, can work well without requiring sensors to maintain this information.
- There is no mobility of sensors.
- For protocol 3, 4, and 5, sensors are assumed to be already divided into clusters and clusters are separated by CDMA codes.

5. Thesis Overview

Data aggregation is an important paradigm for wireless routing in sensor networks. It is a process of collecting data from a set of sensors to a central base-station. With the development in the sensor technology, sensor networks are, now, very easy to deploy and are used for a wide range of applications. This research work proposes several protocols based on contention, CDMA and TDMA schemes. The performance of these protocols is evaluated with the help of simulation and analytical results.

The next chapter describes the related work. Chapter 3 describes the protocols along with the state chart diagrams. This chapter also defines the performance metrics that are used in evaluating the protocols using simulation results. Chapter 4 describes the performance evaluation and comparison of the protocols. In this chapter, the protocols are evaluated using the simulation results and then they are validated by using mathematical analysis. Analysis results supports the simulation results. Finally the best protocol is compared with the "PEGASIS" protocol available in the scientific literature. The "energy * delay" performance metric is used for mathematical analysis of protocols and also for comparison with PEGASIS. The last chapter concludes this research work and gives the direction for the future research.

CHAPTER 2

Related Research Work

1. Review of Sensor Networks

Sensor networks are used for specific purposes. They, by nature, are energy constraint and are usually discarded after their energy is consumed. Thus, in order to increase the expected life of sensors, gathering of data need to be achieved in an energy-efficient manner. In paper [2], the authors have given the upper bound on the lifetime of sensor networks. They have shown with the help of theory and extensive simulations that these upper bounds are tight or near-tight. In case of this research work, the sensor network consists of several sensors. It is assumed that they are rechargeable i.e. they have limited but consistent source of energy [3]. Paper [3] has presented the research challenges that sensor networks need to overcome in order to be commercially viable. This paper has also identified many applications for sensor networks.

Paper [4] and [5] present an overview of wireless sensor networks and review of factors and issues influencing the design of sensor networks. Both papers have explored the potential applications of sensor networks and the communication architecture with the algorithms and protocols developed for each layer. Paper [5] also discussed the open research issues involved in sensor networks. A theoretical bound on lifetime of sensor networks is also described by paper [4].

In paper [6], Bhagwat et al. have identified the important issues and principles involved in designing short range wireless networking support for handheld, portable devices. The issues discussed are network topology, range of communication and selection of wireless link capacity. According to them, there is a trade off between cost, power, quality and utilization in selecting and designing network components.

Heinzelman et al. [7] have proposed Low Energy Adaptive Clustering Hierarchy (LEACH) protocol that involves the randomized rotation of cluster-heads in order to evenly distribute the energy load among the sensors in the network. They also proposed the idea of data-fusion into the routing protocol to reduce the amount of information that must be transmitted to the base-station. Shankar et al. [8] have presented two protocols; cluster based protocol and tree based protocol. They showed analytically that the cluster based protocol performs more efficiently than the tree based protocol. They also stressed on the compression of data before transmission and allowing only a small subset of nodes to make a long distance transmission in order to conserve energy.

Lindsey et al. [9] have proposed Power-Efficient GAthering in Sensor Information Systems (PEGASIS). It is a chain based protocol in which each node communicates with a close neighbor and takes turns for transmitting data to the base-station. They have shown through the simulation results that PEGASIS performs better than LEACH by about 100% to 300% for different network sizes and topologies.

In paper [10] the problem of data collection from a sensor web is considered. The objective of the paper is to find data gathering scheme that balances the energy and delay cost in sensor networks, as measured by energy*delay. They conducted extensive simulation experiments and proved that a binary scheme performs best in terms of energy*delay in case of CDMA capable sensor nodes. If the nodes are not CDMA capable, a chain based 3 level hierarchy scheme performs better.

Ye et al. [11] proposed a MAC protocol (S-MAC) for wireless sensor networks. They emphasized on the need of a new MAC protocol over IEEE802.11 due to special characteristics of sensor networks. They used three techniques to reduce energy consumption and support self configuration in their MAC protocol; periodical sleep, auto-synchronization on sleep schedules, and message passing. The experiment results show that S-MAC performs 2-6 times better than 802.11 in terms of energy consumption.

Krishnamachari et al. [12] have examined the impact of source-destination placement and communication network density on the energy costs and delay associated with data aggregation. They showed that data centric routing offers better performance gains over the traditional address centric routing schemes.

2. Review of Multicasting Protocols

As discussed in chapter 1, data aggregation involves both “one-to-many” and “many-to-one” communication. The scientific work available for multicast communication in wireless networks gives a good background for understanding the issues involved in data aggregation.

There are several multicast protocols available for wireless networks. Some of them are [13] [14] [15] [16] [17] [18] [19]. The major focus of these protocols has been maximizing throughput, system capacity or reliability. The conservation of energy is not the main issue as devices are assumed to have ample energy though it is one of the issues.

There is not much research done in the area of multicasting in sensor networks. However, there are examples of energy-efficient broadcasting and multicasting in ad-hoc wireless networks. Wieselthier et al. [20] [21] [22] have addressed the multicasting problem for energy limited ad-hoc networks. They presented and evaluated several algorithms for the solution of this problem. Their study shows that improved performance can be obtained by exploiting the broadcasting property of the wireless medium i.e. a single transmission by transmitter for delivery of message to multiple receivers rather than simply transmitting individually to all receivers as in wired networks. They have also suggested that the improved performance can be obtained by jointly considering physical layer issues and network layer issues.

Chlamtac and Kutten [23] have utilized two basic properties of multihop radio networks to efficiently implement broadcasting in multihop radio networks. One is the broadcast nature of the radio that allows every single transmission to reach all nodes that are in the vicinity of the transmitting node. The other, spatial reuse of the wireless channel, which due to the multihop nature of the network allows multiple simultaneous transmissions to be received correctly. Their proposed protocol uses these properties to obtain a collision free forwarding of the broadcasted message on a tree. They presented centralized and distributed schemes to construct the broadcast tree. In centralized scheme, the algorithm construct broadcast tree directed from the source node. It assigns the time-slots to all other nodes in the network based on the topology knowledge. In distributed scheme, the source node routes a token in the network to assign time-slots, for message transmission, to individual nodes.

Stojmenovic et. al. [24] have used the concept of internal nodes to significantly reduce or eliminate the communication overhead involved in broadcasting. A node A is said to be an internal node if there exist two neighboring nodes B and C of A such that they are not direct neighbors to each other [25]. They propose the idea to allow only the internal nodes to retransmit for reliable broadcasting, since any other node is directly linked to an internal node. This saves the communication cost with significant reduction in the number of re-broadcasting messages, resulting in reduced contention and collision problems in the network and thus improve the energy-efficiency. The broadcasting algorithms presented are cluster based algorithms.

All these broadcast and multicast protocols suggest three ways to follow in order to construct energy-efficient multicast protocols. Broadcast property of the wireless communication should be exploited, contention should be eliminated by spatial reuse of resources or by use of internal nodes concept, and vertical integration of physical layer issues and network layer issues.

This research work has done a brief study of the performance metrics of various multicast protocols and the ways by which they tried to achieve the energy-efficiency. Tables [1, 2, 3] show the comparison of performance metrics and the ways the energy-efficiency is achieved by different multicast protocols. In table 1, following denominations are also used:-

1. n_i = number of intended destinations by i^{th} multicast arrival.
2. m_i = number of destinations reached by i^{th} multicast session.
3. d_i = duration of i^{th} multicast session.
4. p_i = sum of the transmitter powers used by all nodes in i^{th} multicast session.
5. v_i = multicast value of i^{th} multicast session.
6. X = total number of multicast sessions.

Table 1. **Comparison of performance metrics of multicast protocols**

Paper name	**Performance metrics used**	**How energy efficiency is achieved**
Multicasting in Energy-Limiteds Ad-Hoc Wireless Networks	1. Multicast Value per Energy unit, $\frac{V}{E}$ 2. Multicast Efficiency Value per Energy Unit, $e\frac{V}{E}$ Where, $V = \sum_{i=1}^{X} v_i$ $e = \frac{\sum_{i=1}^{X} m_i}{\sum_{i=1}^{X} n_i}$ $= \frac{total\ number\ of\ destinations\ reached}{total\ number\ of\ intended\ destinations}$ $E = \sum_{i=1}^{X} E_i$ and $E_i = p_i d_i$	Energy is saved by using the broadcast nature of the medium i.e. one transmission to all receivers that lie within the transmission range instead of using unicast to send message to each receiver individually
On the Construction of Energy-Efficient Broadcast and Multicast Trees in Wireless Networks	1. Normalized Tree Power, $Q_i^{'}(m) = \frac{Q_i(m)}{Q_{best}(m)}$ where, $Q_i(m)$ = total power of multicast tree for network m, generated by algorithm i, and $Q_{best}(m) = min\{Q_i(m), i \in I\}$ where I is the set of algorithms.	This paper utilizes the broadcast nature of the wireless medium, CDMA code to avoid contention for the channel, and joining of physical and network layer issues to achieve energy efficiency
Algorithms for Energy-Efficient Multicasting in Ad Hoc Wireless Networks	1. Multicast Efficiency: of the i^{th} algorithm is defined as to be the fraction of desired destinations of the i^{th} multicast service request that are actually reached, $e = \frac{1}{X} \sum_{i=1}^{X} (\frac{m_i}{n_i})$ 2. Yardstick metric $Y = \frac{1}{X} \sum_{i=1}^{X} (\frac{m_i}{p_i})(\frac{m_i}{n_i})$ Note:To take into consideration the conflicting objectives of reaching as many destinations as possible and of maximizing the number of destinations reached per unit energy, they defined yardstick metric	They utilize the broadcast nature of wireless medium, CDMA code to avoid contention for the channel, and join consideration of physical layer and network layer issues to achieve energy efficiency
Bandwidth-Efficient Multicast Routing for Multihop Ad-Hoc Wireless Networks	1. Communication overhead: defined as the total number of control packets transmitted during the simulation. 2. Multicast efficiency: defined as the ratio of the total number of multicast packets received by all receivers to the total number of multicast packets transmitted during the simulation.	Energy efficiency is not an objective

Table 2. **Comparison of performance metrics of multicast protocols (cont.)**

Paper name	**Performance metrics used**	**How energy efficiency is achieved**
Tree-Based Broadcasting in Multihop Radio Networks	1. Number of time-slots in a cycle, c 2. Average bandwidth consumption, $C - avg$ 3. Average memory requirements, M_{avg} 4. Average throughput, T. 5. Maximum delay, $D - max$ 6. Number of received copies 7. Number of transmissions	This paper utilizes two basic properties of wireless networks to achieve energy efficiency; first, the broadcast nature of the medium that allows every single transmission to reach all nodes that are in the vicinity of the transmitting node and second, spatial reuse of the wireless channel, which due to the multihop nature of the network allow the multiple simultaneous transmissions to be received correctly
Internal Nodes Based Broadcasting Algorithms in Wireless Networks	1. Reachability (RE): the number of nodes receiving the broadcast message divided by total number of nodes that are reachable from the source 2. Saved ReBroadcast (SRB): $\frac{(r-t)}{r}$, where r is the number of nodes receiving the broadcast message, and t is the number of nodes that actually transmitted the message.	This paper use the concept of internal nodes to significantly reduce or eliminate the communication overhead involved in broadcasting. A node A is said to be an internal node if there exist two neighboring nodes B and C of A that are not direct neighbors themselves. They restrict broadcasting to internal nodes only, since any other node is directly linked to an internal node. This saves the communication cost with significant reduction in the number of re-broadcasting messages, resulting in reduced contention and collision problems in the network and thus improve the energy efficiency.

Table 3. **Comparison of performance metrics of multicast protocols (cont.)**

Paper name	**Performance metrics used**	**How energy efficiency is achieved**
Reliable Multicast in Multi-access Wireless LANs	1. Mean channel holding time.	Energy efficiency is not an objective.
CBT-FR	1. Network bandwidth 2. Probability of packet loss	Energy efficiency is not an objective.
A Performance Comparison Study of Ad Hoc Wireless Multicast Protocols	1. Packet delivery ratio: the ratio of the number of data packets actually delivered to the destinations versus the number of data packets supposed to be received. 2. Number of data packets transmitted per data packet delivered. 3. Number of control bytes transmitted per data bytes delivered. 4. Number of control and data packets transmitted per data packet delivered.	Energy efficiency is not an objective. The protocols compared are AMRoute, ODMRP AMRIS, CAMP, and flooding

CHAPTER 3

Protocols Description and Performance Metrics

1. Protocols Description

1.1. Protocol 1. One simple approach is that the base-station broadcasts the message with the group-id. All sensors check the group-id. Sensors that are part of the group will respond directly to the base-station and sensors that are not part of the group will ignore the message. The “many-to-one” communications from the base-station to group-members is carried out using leader based channel contention protocol [13] and the “one-to-many” communications from group-members to the base-station is carried out using MAC 802.11.

In leader based protocol, one of the sensor is selected as a leader. The base-station broadcasts an RTS packet and if the leader receives RTS correctly then it sends CTS to base-station otherwise it remains quiet. If a group-member receives RTS correctly then it remains quiet otherwise it sends NCTS(Negative CTS) to base-station. The base-station will proceed only when it receives a CTS otherwise it will retransmit RTS. Similarly, when the base-station broadcasts a data packet and if the leader receives data packet correctly then it sends ACK to base-station otherwise it remains quiet. If a group-member receives data packet correctly then it remains quiet otherwise it sends NAK to base-station. The base-station will proceed only when it receives an ACK otherwise it will retransmit data

Table 4. **Description of events and actions used in state chart diagrams**

Event/Action name	**Description**
$start\_data\_aggregation()$	This event is called when the upper layer request for starting data aggregation
$rcv\_RTS()$	This event is called when RTS is received
$rcv\_CTS()$	This event is called when CTS is received
$rcv\_RES()$	This event is called when RES is received
$rcv\_REQ()$	This event is called when REQ is received
$rcv\_NAK()$	This event is called when NAK is received
$rcv\_NCTS()$	This event is called when NCTS is received
$send\_RTS()$	This action is taken to send RTS
$send\_CTS()$	This action is taken to send CTS
$send\_RES()$	This action is taken to send RES
$send\_REQ()$	This action is taken to send REQ
$send\_NCTS()$	This action is taken to send NCTS
$send\_NAK()$	This action is taken to send NAK
$start\_timer()$	This action is taken to start timer
$stop\_timer()$	This action is taken to stop timer
$timeout()$	This event is called when timer is expired
$RES\_ALL()$	This event indicates that response is received from all members
$fuse\_data()$	This action is taken to fuse the member data with its data before transmitting to the base-station

packet.

The advantage of protocol 1 is that it does not need clustering algorithm at lower layer for grouping the sensors into clusters (used in Approach 3 and 4). Also, this protocol is easy to implement. This protocol is suitable for the wireless networks where energy efficiency is not the primary goal and collisions of packet is not a big issue, which can be over come by using retransmissions. In this research case, the author is interested in developing energy-efficient protocols for the sensor networks where retransmission of the packets may not be a feasible solution. Figures 1, 2, and 3 show the state chart diagram of base-station, leader, and member respectively. The events and actions used in the state chart diagram are explained in table 4

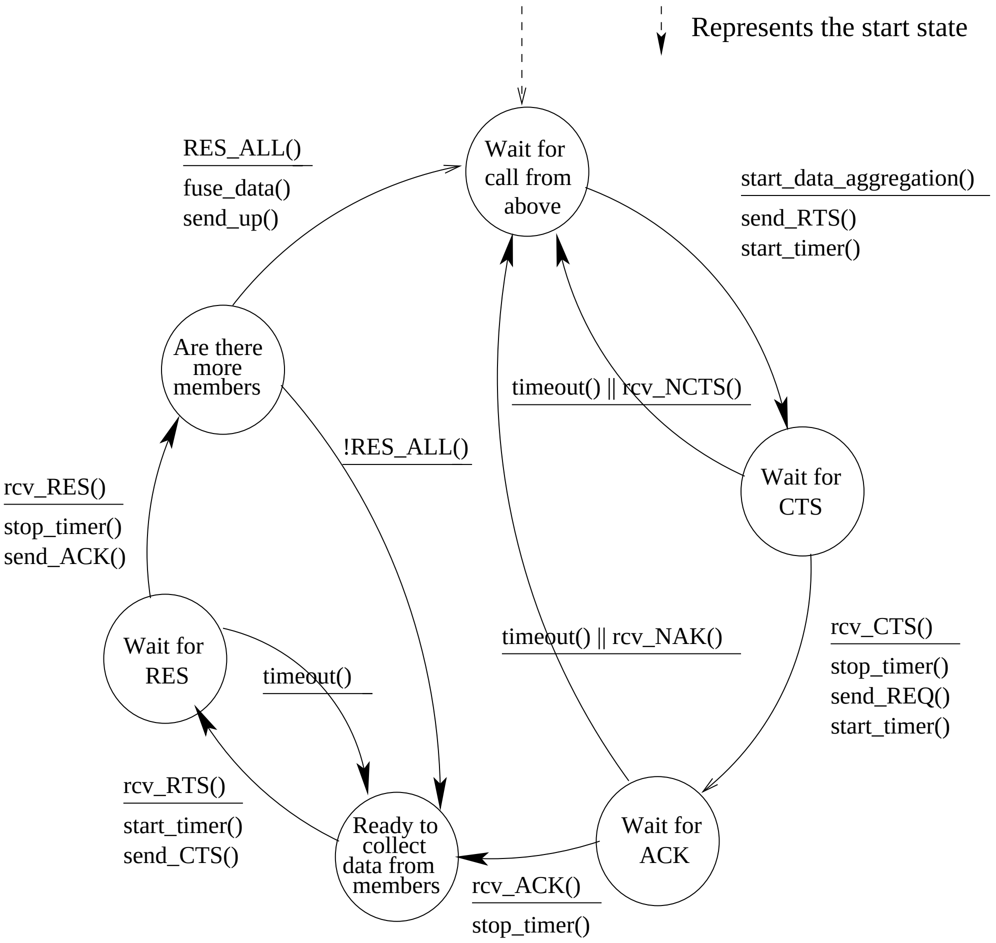


Figure 1. **State chart diagram of base-station in protocol 1**

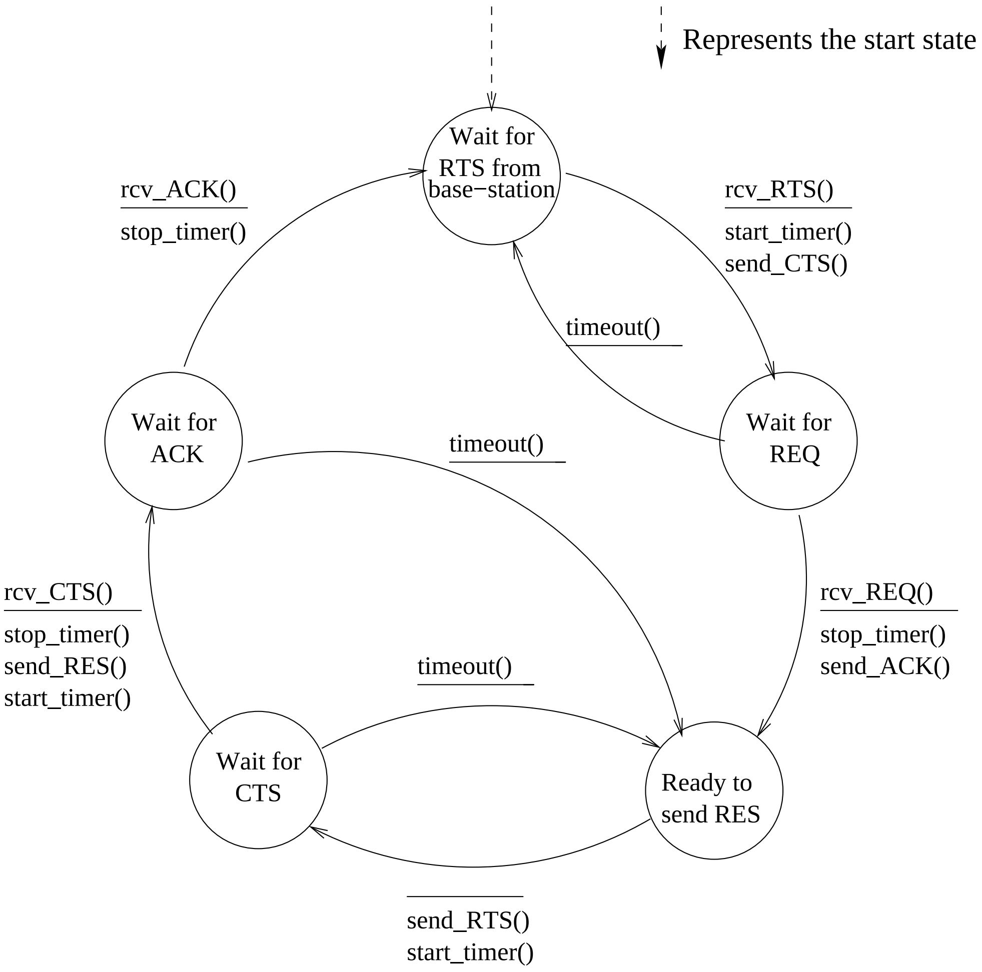


Figure 2. **State chart diagram of leader in protocol 1**

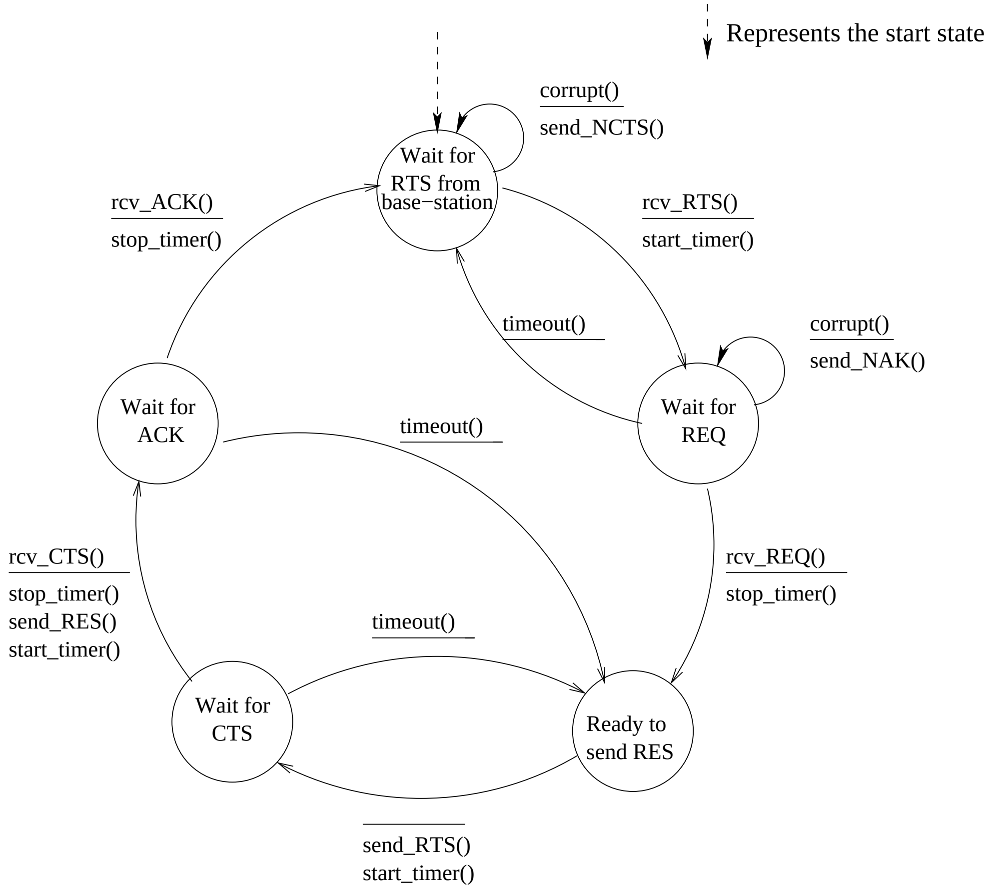


Figure 3. **State chart diagram of member in protocol 1**

1.2. Protocol 2. This protocol is similar to the first protocol except that it uses a TDMA-based scheme instead of contention-based scheme for the communication. The base-station and the sensors form a time-frame for communications. Each node including base-station has a specific time-slot in the time-frame for transmitting data. A node transmits only in its time-slot. All sensors have to listen to channel when base-station transmits. When a sensor is neither transmitting nor receiving, its radio during that time is switched off. When the base-station broadcasts a request for data in its time-slot, the group members respond to it with the data that they have.

This protocol, like protocol 1, does not need clustering protocol at lower layer for grouping the nodes into clusters. This protocol works well as long as the number of sensors is small. As the number of sensors increases, the frame size increases that result in more sleep time and thus reduce the overall data communication efficiency. Also if the number of group-members is small compared to the total number of sensors in the network, then a large number of slots will be empty within a frame during the communication between the base-station and group-members. This is due to the fact that the sensors, which are not part of group will not transmit any data in their time-slot. Figure 4 shows the state chart diagram of base-station and members in protocol 2. The events and actions used in the state chart diagram are explained in table 4

1.3. Protocol 3. In this protocol, sensors inside the body are considered to be grouped into clusters. Each cluster has a separate CDMA channel for communication and consists of a cluster-head and cluster-members. Cluster-head is the leader or representative of the cluster and cluster-member is a non cluster-head member of the cluster. It is assumed that the clustering of the node is already done at a lower layer and the base-station has the

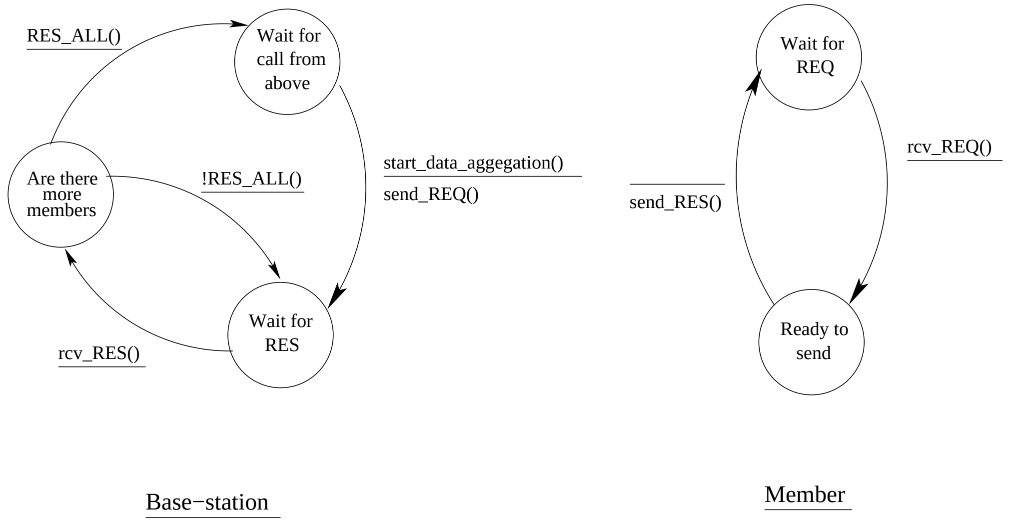


Figure 4. **State chart diagrams for protocol 2**

clustering information i.e. the information regarding the members of each cluster.

The communication from sensors to base-station is carried out using atmost two hops. When a group-member, which is also a cluster-member wants to send a message to the base-station, it sends the message to the cluster-head using the CDMA channel allotted to that cluster. The cluster-head sends that message to the base-station. In other words, the cluster-members can only send messages to the base-station through the cluster-head. When a group-member, which is also a cluster-head wants to send a message to the base-station, it directly sends it to the base-station. The idea is to reduce the number of long distance communications and reduce the number of collisions by separating the clusters using CDMA channel.

The cluster-heads can collect the data from the cluster-members, diffuses them together and then send them as one message to the base-station. In such a case, the cluster-head should have some buffer to store data. The collected data can also be compressed at the cluster-head before sending it to the base-station. This research work is only performing data diffusion but not data compression. Data compression is left for the future research topics.

The communication from base-station to sensors is carried out using leader based protocol as explained in Approach 1. Since the base-station sends message with enough power and changing the CDMA code at sensors does not involve much overhead, therefore the cluster-members can receive the message sent by the base-station directly and thus avoiding the need to route the message through cluster-heads.

This protocol results in less number of collisions as compared to protocol 1 but this still result in collisions that hampers the energy efficiency. Also, this protocol is more complex to implement than the previous protocols. Figures 5, 6, 7, and 8 show the state chart diagram of base-station, leader, cluster-head and cluster-member respectively. The events and actions used in the state chart diagram are explained in table 4

1.4. Protocol 4. This protocol uses the similar architecture as in protocol 3. The sensors are grouped into clusters with each cluster having different CDMA code. Here, the communication between the base-station and the sensors is carried out using two layer TDMA scheme insist of using contention based protocols. The cluster-heads communicate with the base-station in one TDMA frame and each cluster-head communicates with its cluster-members in another TDMA frame. In other words, there is one more TDMA frame than the total number of clusters. One TDMA frame for each cluster and one for base-station and cluster-heads communication. The base-station-cluster-heads TDMA frame is synchronized with each clusters TDMA frame. The synchronization is described below:-

The first time slot of all time frames is allotted to the base-station. The consecutive time slots in the base-station-cluster-heads time frame are allotted to cluster-heads. These

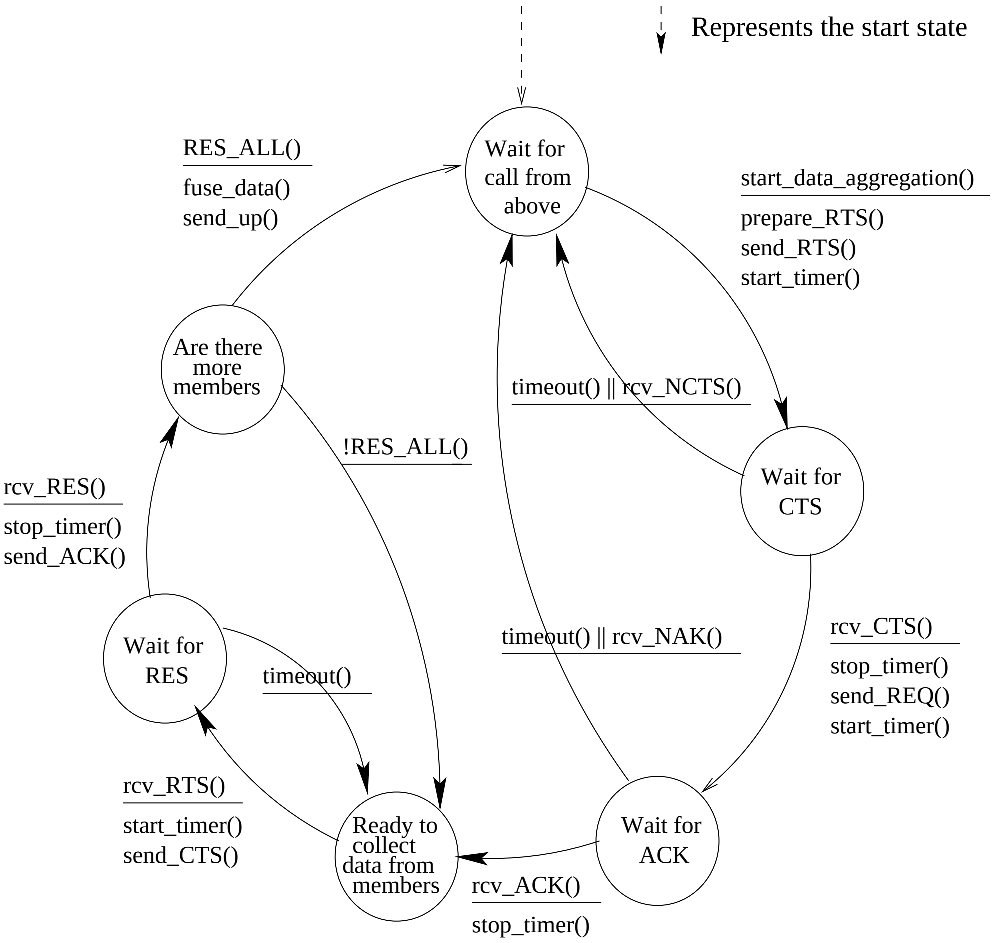


Figure 5. **State chart diagram of base-station in protocol 3**

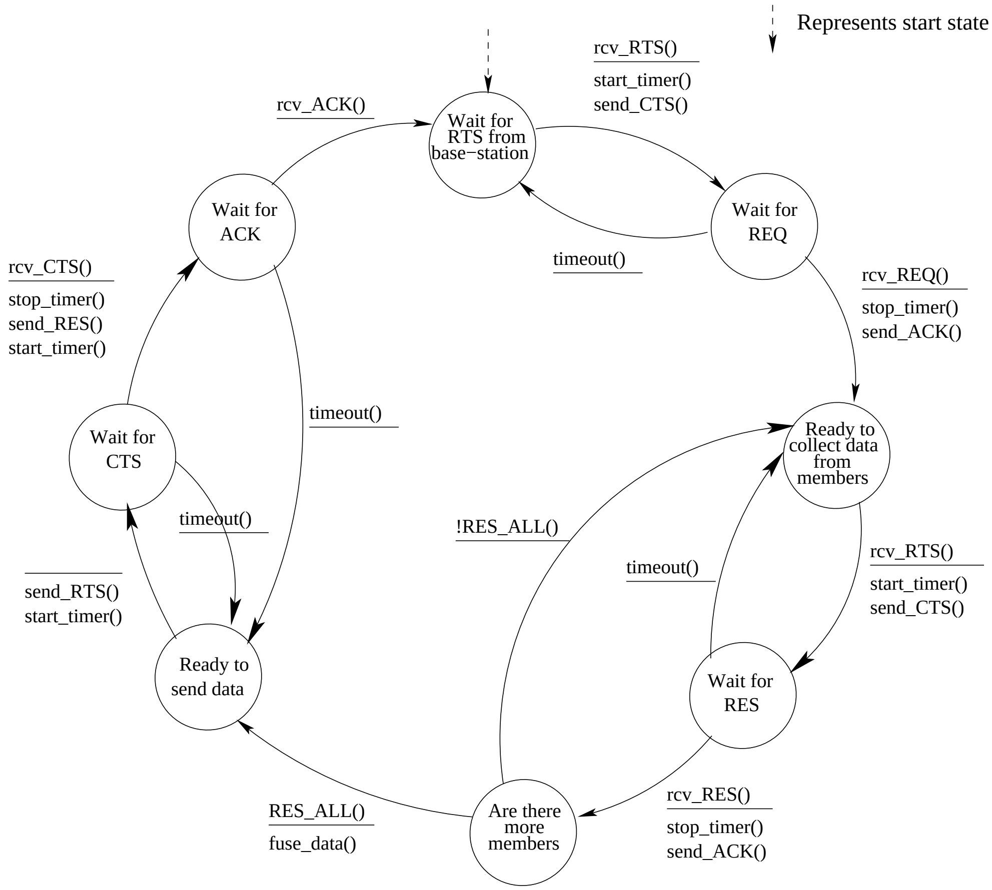


Figure 6. **State chart diagram of leader in protocol 3**

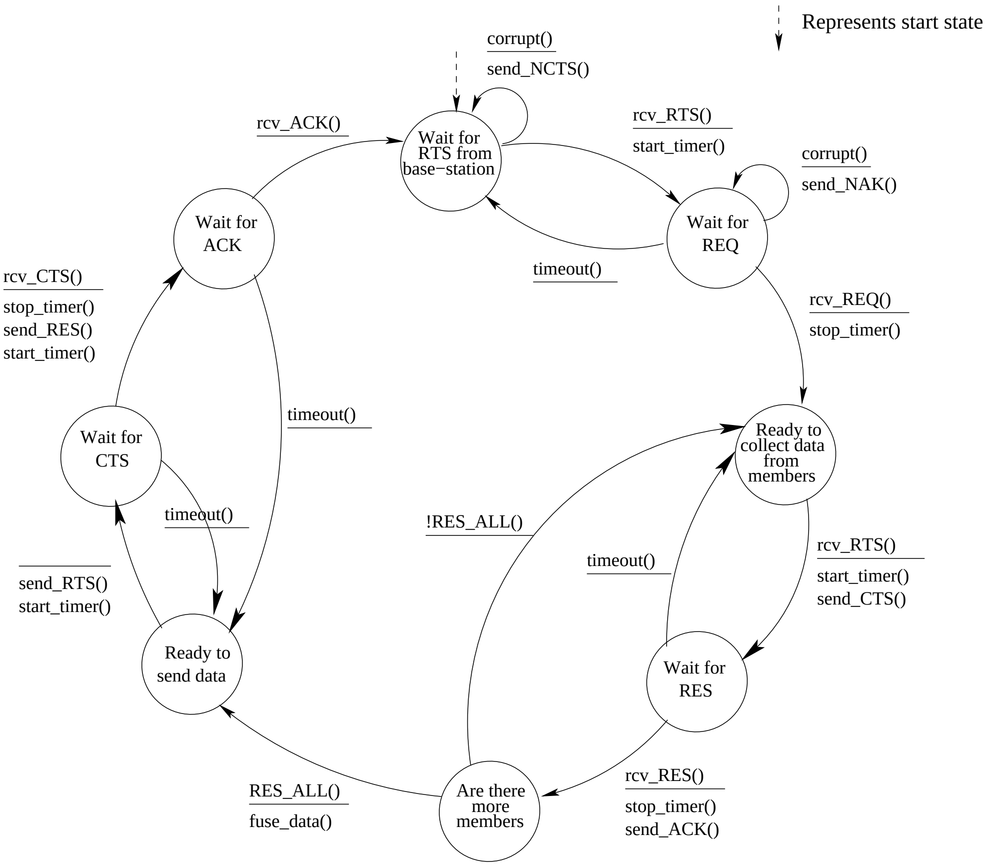


Figure 7. **State chart diagram of cluster-head in protocol 3**

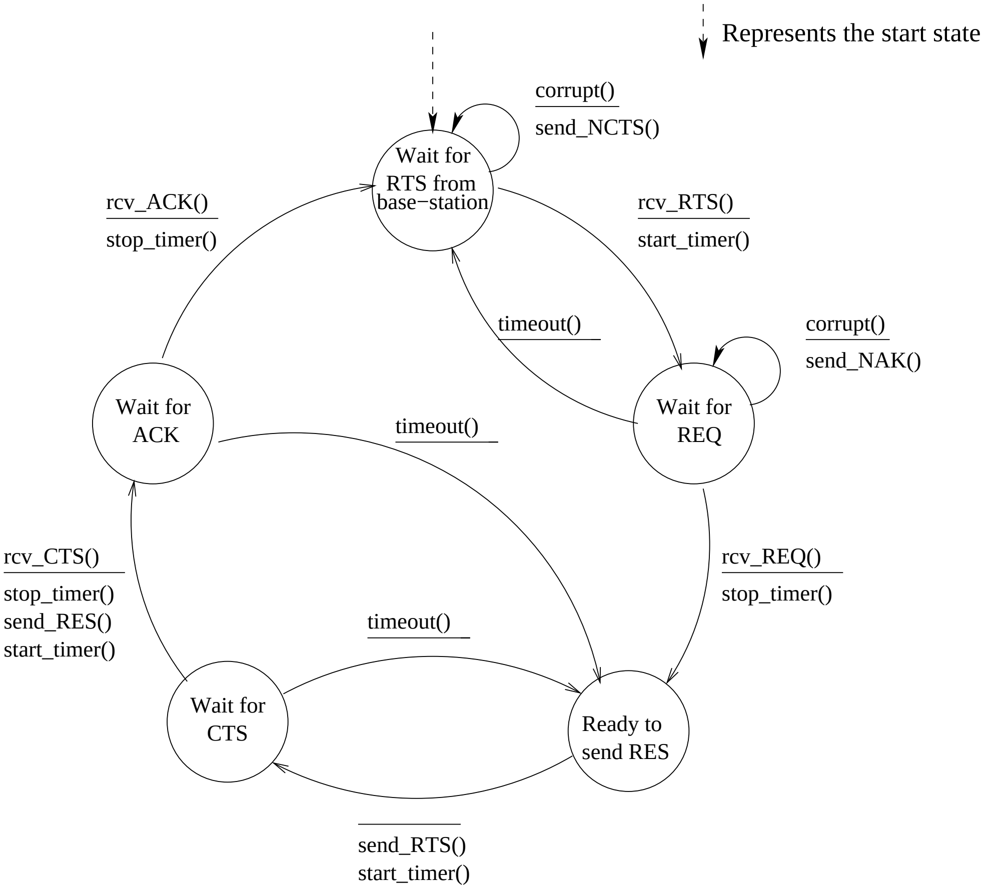


Figure 8. **State chart diagram of cluster-member in protocol 3**

consecutive time slot starts after a time gap as shown in figure 9. The figure 9 is drawn for the network topology shown in figure 10. The time slots for the cluster-members in each cluster time frame starts after the base-station's time slot. After the cluster-members time slot in each cluster, the cluster-head time slot starts, which may be consecutive or after some time gap depending on its time slot in the base-station-cluster-head time frame. The slot size is equal for all sensors in this scheme.

The base-station transmits a message containing group-id during its time slot. All sensors listen to the message. Cluster-members that are part of the group send data in their time slot to the cluster-head. The cluster-head then sends it to the base-station in its time slot. If the cluster-head is also part of the group then it also append its data in to the message that it sends to the base-station. Cluster-members that are not part of group just ignore the message. It is to be noted that even if the cluster-head is not the member of the group, it has to transmit data if some of its cluster-members are group-members.

This protocol still results in empty slots though less in number than protocol 2. Figure 11 shows the state diagrams for protocol 4. The events and actions used in the state chart diagram are explained in table 4

1.5. Protocol 5. In this protocol, It is assume that there already exist the architecture of protocol 4 for communications other than data aggregation between the base-station and the sensors. This protocol builds a new architecture over this architecture for providing efficient data aggregation by the base-station before the start of data aggregation session. This new architecture will only consist of the group-members.

Before the start of data aggregation session, the base-station broadcasts a message with the group-id for setting up the new architecture for communication between the base-

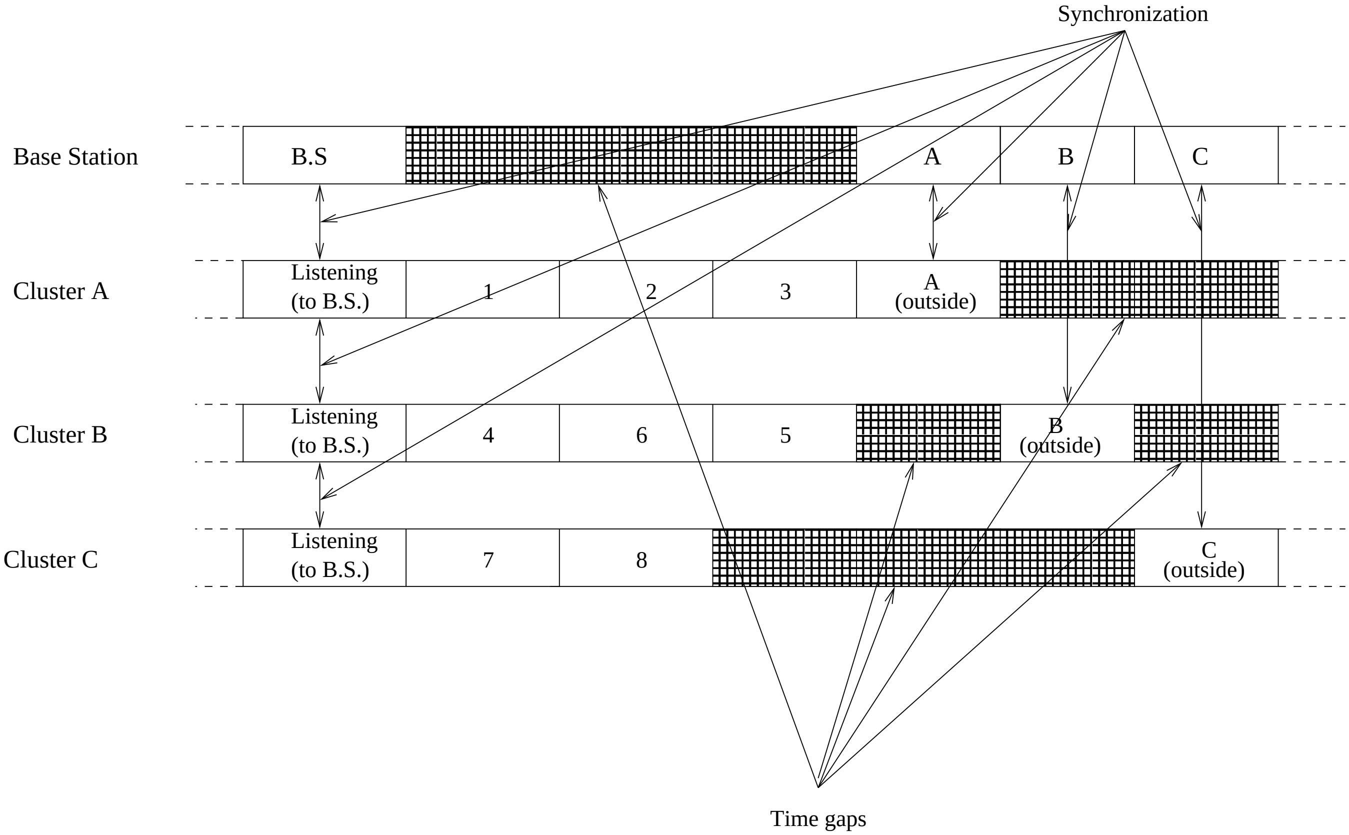


Figure 9. **Loose Time Frame Design**

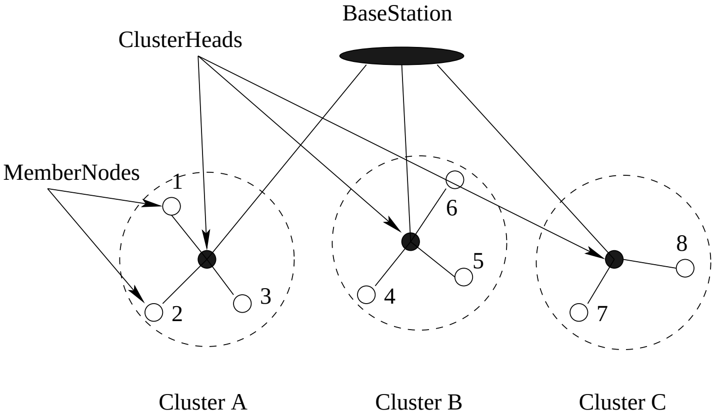
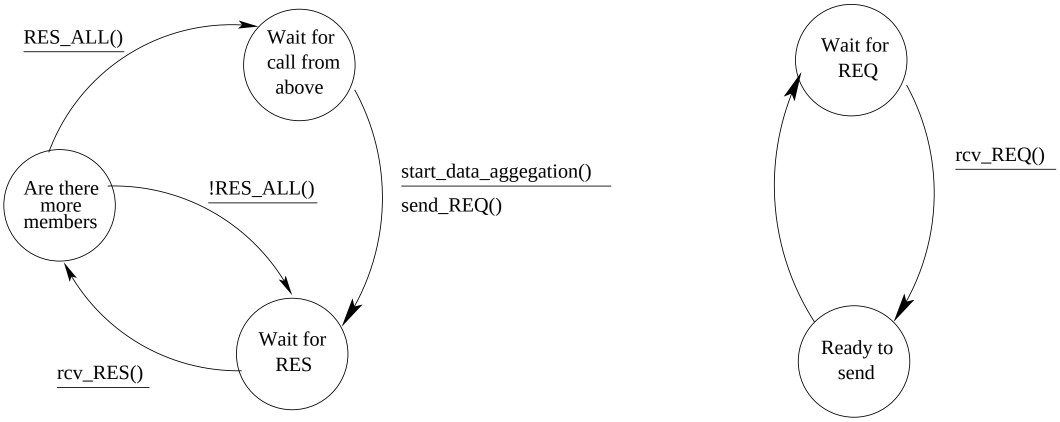
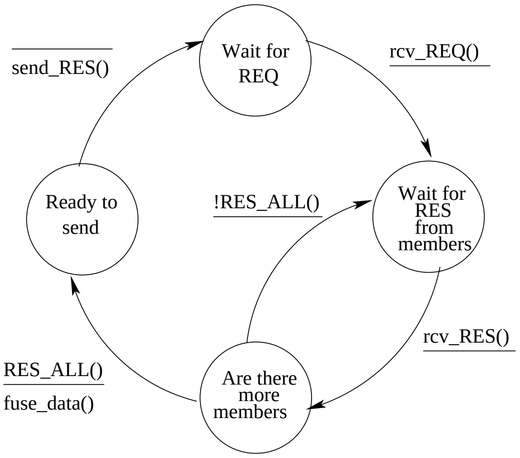


Figure 10. **An example of network layout**



Base−station

Member



Cluster−head

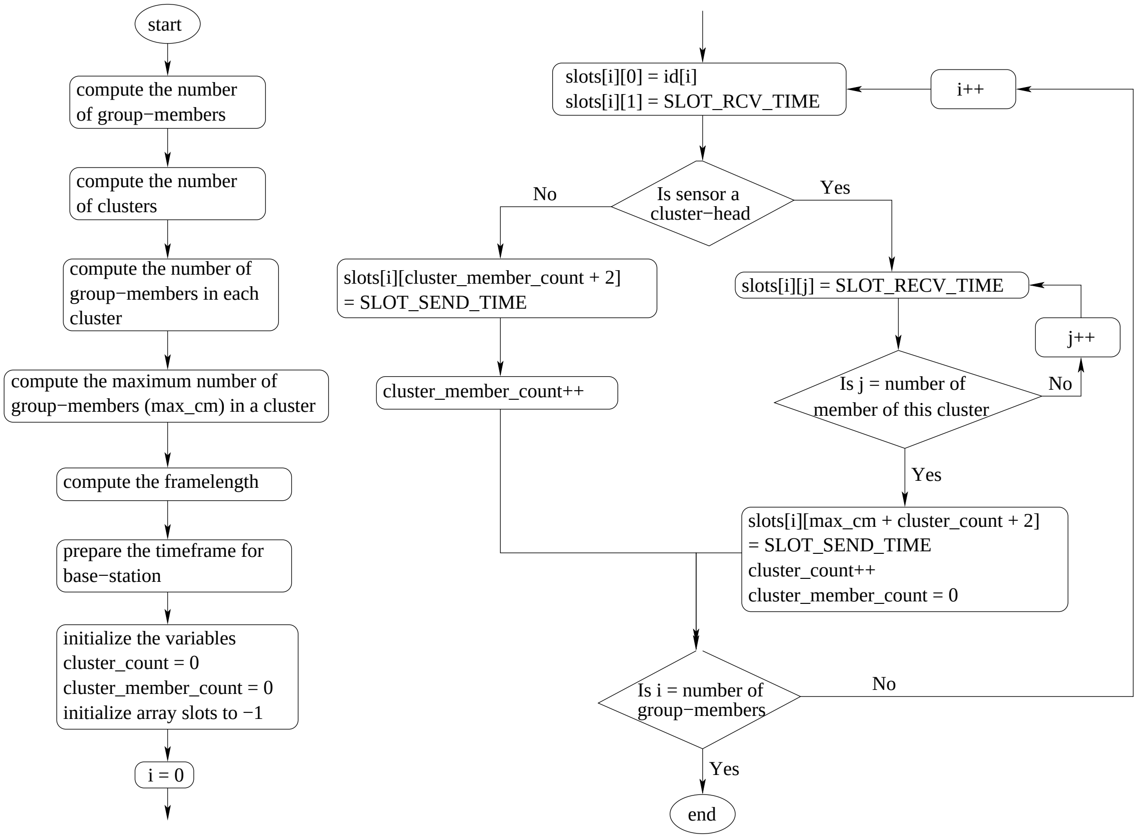
Figure 11. **State chart diagrams for Protocol 4 and 5**

station and the sensors that are group-members. Cluster-members that are the group-members respond to the message in their respective time-slot to their cluster-heads. The cluster-head then forwards the cluster-members message to the base-station in its time-slot. If no cluster-member of a cluster responds to the base-station's message and the cluster-head of that cluster is also not a group-member then this cluster-head need not send a message to the base-station. Based on the messages received from the cluster-heads, base-station designs the new architecture (i.e. new slot allocation), which consists of time-slots only for the sensors that had responded to the base-station's message. This architecture will be similar to the protocol 4 architecture with fewer number of sensors in each time-frame. The algorithm for computing this new slot allocation resulting in slot minimization is described in figure 12.

This whole process gives rise to a new base-station-cluster-heads and clusterhead-groupmembers TDMA frames setup over the old basestation-clusterheads and clusterhead-clustermember TDMA frames setup. After this, the base-station starts the communication. When data aggregation session terminates, the base-station, then, broadcasts a wakeup message that wakes-up all the nodes and thus, they revert back to the initial architecture for other communications. Figure 11 shows the state chart diagrams for protocol 5. The events and actions used in the state chart diagram are explained in table 4

2. Performance Metrics for Comparing Protocols

As already discussed, the sensors are energy constraint and have limited memory and computation power. So the performance metrics should incorporate the metrics to measure these constraints. In view of these considerations, the following performance metrics are defined for evaluating the performance of proposed protocols:-



NOTES: SLOT_RECV_TIME represents the Rx slots
SLOT_SEND_TIME represents the Tx slots
slots[][] datastructure stores the slot allocation information
and it is send to the sensors in next timeframe

Figure 12. **Algorithm for slot minimization in protocol 5.**

Total Power Consumed by an algorithm is defined as the total power consumed in a session for network m, generated by an algorithm i.

Session Efficiency$(M_i(m))$ is defined as the ratio of the total number of packets (N_{rx}) received by all receivers in a session for network m, generated by an algorithm i to the total number of packets (N_{sx}) being transmitted during that session for the same network m, generated by the same algorithm i.

$$M_i(m) = \frac{N_{rx}}{N_{sx}} \times 100, \tag{3.1}$$

Session efficiency gives the idea of number of packets lost in the network due to collisions. So for an algorithm having no collisions, session efficiency is 100%.

Communication Overhead(C_i) is defined as the ratio of the total number of control packets $(N_{control})$ transmitted in a session for network m, generated by an algorithm i, to the total number of data packets (N_{data}) transmitted during that session for the same network m, generated by the same algorithm i.

$$C_i(m) = \frac{N_{control}}{N_{data}} \times 100, \tag{3.2}$$

Communication overhead tells about the number of control packet transmitted per hundred data packets for an algorithm. Lower the value of communication overhead, better the algorithm.

Latency(L_i) is defined as the time difference between the time (T_{start}) when the base-station transmits first message in a session for a network m, generated by an algorithm i, and the time (T_{stop}) when the last receiver receives last message in that session for the same network m, generated by the same algorithm i.

$$L_i(m) = T_{start} - T_{stop}, \tag{3.3}$$

Latency gives the idea about the delays in the network. If the latency is more, there is more delay for a packet to reach all the receivers. Generally, Lower the latency, better the algorithm.

Any algorithm that consumes less power, has low communication overhead, high session efficiency and low latency is considered as the best algorithm for data aggregation.

CHAPTER 4

Performance Evaluation and Comparison

This chapter describes the experiments conducted to study the performance of the protocols proposed for “Selective Data Aggregation in Wireless Sensor Networks”. This research work uses Network simulator [26] to perform the simulations. The results obtained from the simulator are logged on a trace file. These results are further analyzed with the help of PERL scripts to evaluate the performance of the algorithms. The protocols are also analyzed to validate the simulation results. The analytical results are presented later in the chapter. Both the simulation and analysis results suggest that the protocol 5 overall performs better than other protocols. Finally the best performing protocol i.e. protocol 5 is compared with PEGASIS using the mathematical analysis.

The objective of the experiments is to evaluate the performance of the algorithms in terms of energy efficiency, communication overhead, session efficiency, and latency and to suggest the best algorithm among the algorithms described in this research work. Experiments are conducted with varying number of sensors, group members, and clusters.

1. Simulation Environment: Network Simulator

Network Simulator is a discrete event simulator used for simulating networking protocols. NS provides support for simulation of TCP, routing, MAC, and multicast protocols over wired and wireless networks.

This research work use Network Simulator (NS) for simulating and evaluating the protocols. The protocols are implemented at MAC layer inspite of the fact that some of them involves two hop communication. This is because sensors are small devices, which probably have only single layer architecture as opposed to multi-layer architecture. This implies that its better to simulate the protocols at one layer in order to make them resemble more to the real life scenarios.

Constant Bit Rate (CBR) traffic model is used for data generation. All the data packets are of length 20 bytes. The nodes have a transmission range of $250mm$ for protocol 1 and 2. For protocol 3, 4 and 5, the base-station and cluster-heads have a transmission range of $250mm$ and cluster-members have transmission range of $50mm$. This is due to the fact that only the base-station and cluster-heads need to transmit to large distances. Cluster-members transmit only to their respective cluster-heads that are in short range distance. The parameters of the simulator are adjusted in such a way that it works like a $914MHz$ Lucent WaveLAN DSSS radio interface with $2Mbps$ of bandwidth.

For the simulation purpose, a session is defined to consist of 5000 phases. A phase is one complete data aggregation request. In other words, a phase starts when the base-station sends a request for data aggregation to the sensors and ends when the base-station receives the last message from a sensor for that data aggregation request.

2. NS Extensions

In order to increase the simulation accuracy, this research work has made several extensions to NS. These extensions are described below:-

2.1. Propagation Model: A New Propagation Model. NS supports free space and two ray propagation model. These models are used for wireless networks in general scenarios. The applications emphasized in this research work require a model for propagation of RF waves in human body. For this purpose, A propagation model for communication within the human body has been developed by the research team [1]. According to that propagation model, the power received by a receiver, P_r is given by the formula

$$P_r = \frac{(P_t - P_{NF} - P_{FF})\lambda^2}{(4\pi.d)^2} G_t G_r, \tag{4.1}$$

where

$$\begin{aligned} P_{NF} &= \sigma \frac{\mu\omega}{\sqrt{\sigma^2+\epsilon^2\omega^2}} \left(\frac{I^2.dl^2}{6\pi}\right) [e^{-2\alpha r} \left(\frac{\gamma^2}{2\alpha} + \frac{d_0 - r}{4r^2} + \frac{\gamma(d_0 - r)}{2r}\right) \\ &\quad + e^{-2\alpha d_0} \left(\frac{-\gamma^2}{2\alpha} + \frac{d_0 - r}{4d_0^2} + \frac{\gamma(d_0 - r)}{2d_0}\right) \\ &\quad + e^{-\alpha(d_0+r)} \left(\frac{2(d_0 - r)}{(d_0 + r)^2} + \frac{2\gamma(d_0 - r)}{(d_0 + r)}\right)] \end{aligned} \tag{4.2}$$

and,

$$P_{FF} = \sigma\eta^2\gamma^2 \frac{I^2 dl^2}{12\pi\alpha} (e^{-2\alpha d_0} - e^{-2\alpha d}). \tag{4.3}$$

The variables used in equations 4.1- 4.3 are described in table 5. This propagation model has been implemented in NS[2]. This research work use this propagation model as underlying propagation model for wireless communication. This has also been noticed

---

[1] The work is submitted in ICC 2003

[2] The propagation model is implemented by Valliappan Annamalai

Table 5. **Description of variables**

P_t	Power at which signal is transmitted by the transmitter
P_r	Power at which signal is received by the receiver
d	Distance between two antennas
d_0	Distance separating near field and far field regions
G_t	Gain of the transmitter antenna
G_r	Gain of the receiver antenna
λ	Wavelength of the RF wave
σ	Conductivity of the medium
ϵ	permittivity of the medium
μ	Permeability of the medium
ω	Angular Frequency
η	Complex intrinsic impedance $\left(\frac{\gamma}{\sigma+j\omega\epsilon}\right)$
γ	Complex propagation constant $(\alpha + j\beta)$
α	Attenuation constant
β	Phase constant
I	Current in the transmitting antenna
r	radius of the antenna
dl	length of the antenna

that using a different propagation model does not affect the relative performance of the protocols. This is due to the reason that changing the propagation model only changes the energy consumed in one transmission. The number of transmissions, which are same for any propagation model, depend on the protocol used and thus, the protocols perform relatively same with different propagation models.

2.2. Energy Model: Extensions For Amplifier Circuitry Power Dissipation. The version of NS used for simulation supports energy model that considers only the power consumed in the transmitter and receiver circuitry. The amplifier circuitry power dissipation at the transmitting antenna is ignored. Thus to incorporate the amplifier circuitry power dissipation, energy model is extended as described in next section.

The energy consumed in the antenna at the time of transmission for free space is given by [7] as:

$$\begin{aligned} E_{T_x}(k,d) &= E_{T_{x-elec}}(k) + E_{T_{x-amp}}(k,d) \\ E_{T_x}(k,d) &= E_{elec}k + \epsilon_{amp}kd^2 \qquad (4.4) \end{aligned}$$

In NS, The $E_{T_{x-elec}}(k)$ part is implemented but $\epsilon_{amp}kd^2$ part has not been implemented. Second part is implemented by this research work in order to increase the simulation accuracy by creating a subclass of existing energy model class. Further, the above equation is valid only for freespace. So a new equation for energy consumption corresponding to the propagation model for human body is developed, which is described here.

The transmitter power dissipation is same for all media. The only thing that changes is the power consumed in the amplifier circuitry. This is because for different propagation model, the signal needs to be amplified to different power levels and thus incur different power dissipation. For freespace, paper [7] considers d^2 energy loss due to channel transmission. This is not true for this model. After carefully investigating the propagation model formula, it is observed that P_t is proportional to $(d^2 + A - Be^{-2\alpha d})$, where A and B are constants given by:

$$A = \frac{\left[P_{NF}(12\pi\alpha) + \sigma\eta^2\gamma^2 I^2 dl^2 e^{-2\alpha d_0}\right] G_t G_r \lambda^2}{48\pi^3 \alpha P_r} \qquad (4.5)$$

$$B = \frac{\left[\sigma\eta^2\gamma^2 I^2 dl^2\right] G_t G_r \lambda^2}{48\pi^3 \alpha P_r} \qquad (4.6)$$

Thus, considering a $(d^2 + A - Be^{-2\alpha d})$ energy loss due to channel transmission, to transmit a k bit message a distance d, the energy consumed in transmitter circuit is:

$$\begin{aligned} E_{T_x}(k,d) &= E_{T_{x-elec}}(k) + E_{T_{x-amp}}(k,d) \\ E_{T_x}(k,d) &= E_{elec}k + \epsilon_{amp}k\left(d^2 + A - Be^{-2\alpha d}\right) \qquad (4.7) \end{aligned}$$

The values of A, B, and *alpha* are calculated by writing a C program and are then used in the simulator. The above formula is implemented in to the simulator along with the freespace formula. This extended energy model is used for simulation. The variables used in equations 4.5 and 4.6 are described in table 5.

3. Performance Evaluation Using Simulation Results

In this section, this research work investigates the performance of the protocols using the simulation results. This research work has also conducted the statistical analysis of the simulated data. According to this analysis, The confidence level of each datum collected is 95% and the confidence limit is within acceptable limit of 4%. In other words, there is a 95% probability that the results obtained for each protocol does not differ from the actual average value by more than 4%. Details about this statistical analysis and formulae used is given in the book "The art of computer systems performance analysis: techniques for experimental design, measurement, simulation, and modeling" [27]. The experiments are described in the following sub-sections: -

3.1. Experiment 1. The simulations are conducted to see the performance of protocols for varying number of group members by keeping the total number of nodes and number of nodes in a cluster as constant. Total number of nodes is kept equal to 31 and percentage of nodes in a cluster is fixed to 25%. Figure 13, 14, 15, and 16 shows the plots for experiment 1. It is observed that the power consumed and latency increases with the increase in the percentage of group members for all protocols. Session efficiency decreases and average communication overhead increases with the increase in percentage of group members for protocol 1 and 3. For protocol 2, 4 and 5, average session efficiency and communication overhead is constant. Protocol 5 is performing best in terms of power consumed and latency and performing equally well to protocol 2 and 4 in terms of session efficiency and communication overhead.

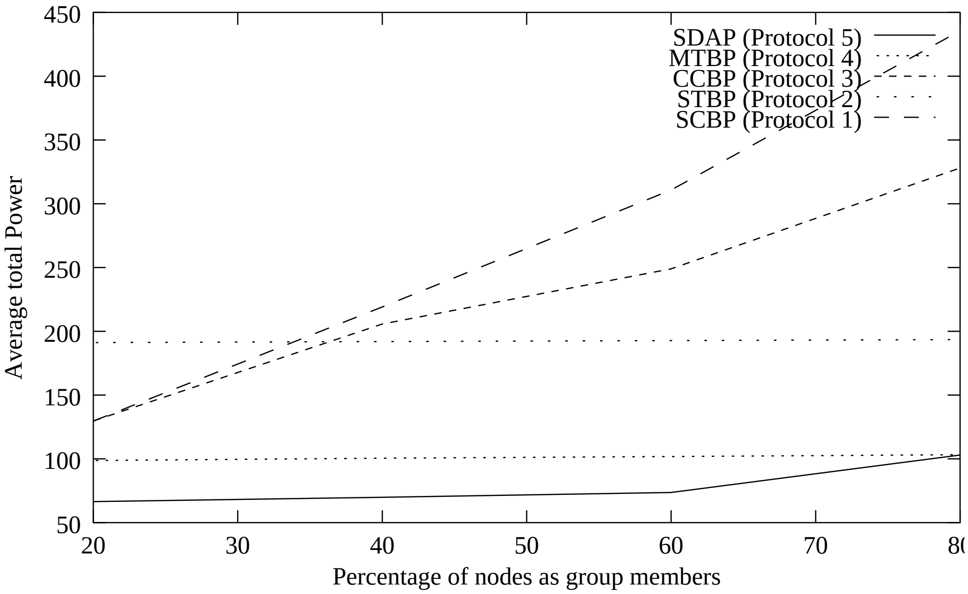


Figure 13. **Comparison of power consumed for varying percentage of group members by keeping number of nodes = 31 and percentage of nodes in a cluster = 25**

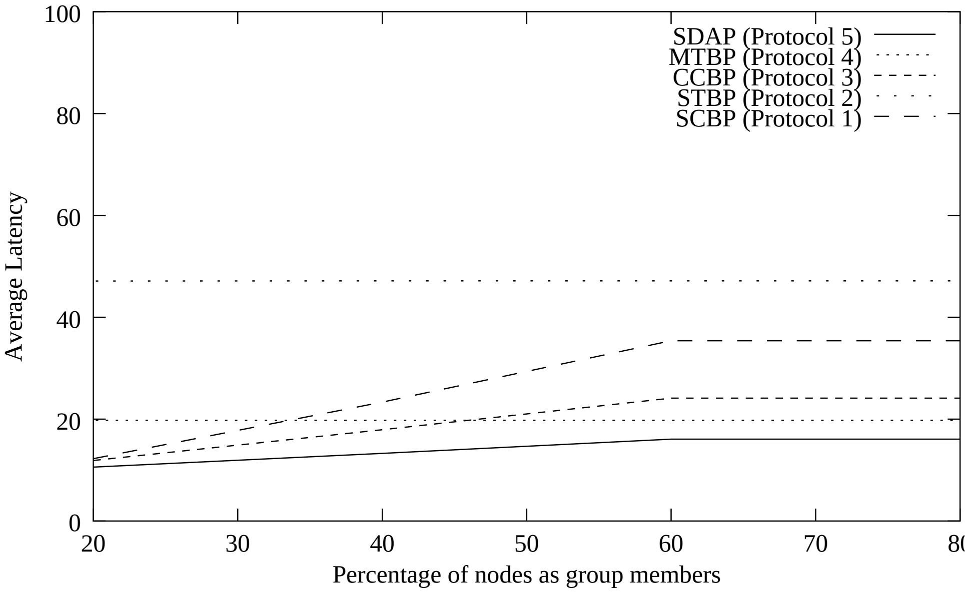


Figure 14. **Comparison of latency for varying percentage of group members by keeping number of nodes = 31 and percentage of nodes in a cluster = 25**

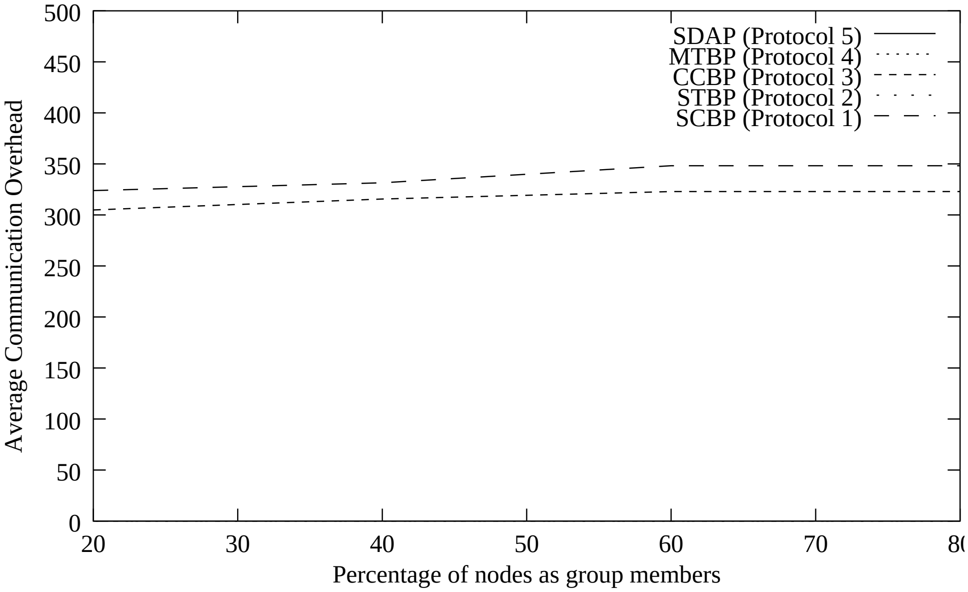


Figure 15. **Comparison of communication overhead for varying percentage of group members by keeping number of nodes = 31 and percentage of nodes in a cluster = 25**

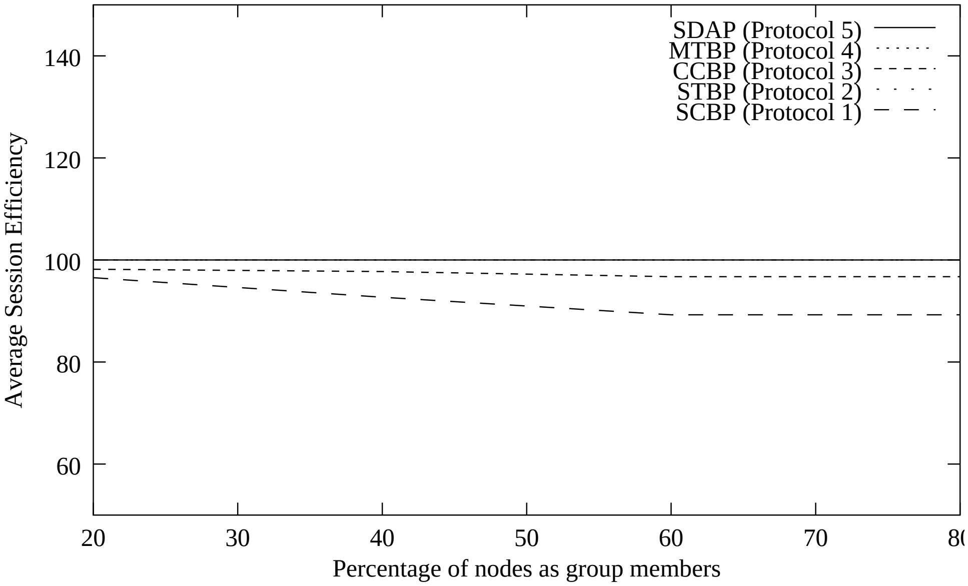


Figure 16. **Comparison of session efficiency for varying percentage of group members by keeping number of nodes = 31 and percentage of nodes in a cluster = 25**

3.2. Experiment 2. The objective of this experiment is to evaluate the performance of protocols in terms of power consumption, latency, communication overhead, and session efficiency by varying the percentage of nodes in a cluster. The total number of nodes is kept equal to 21 and percentage of group members is kept equal to 60. Figure 17, 18, 19, and 20 shows the plots for experiment 2. Protocol 1 fails for 50 percentage of nodes in a cluster. Power consumed and latency increases with the increase in percentage of number of nodes in a cluster for all protocols. Communication overhead and session efficiency is almost constant for all protocols. Protocol 5 performs best in terms of power consumption and latency. In terms of average communication overhead and session efficiency, protocol 2, 4, and 5 performs equally well.

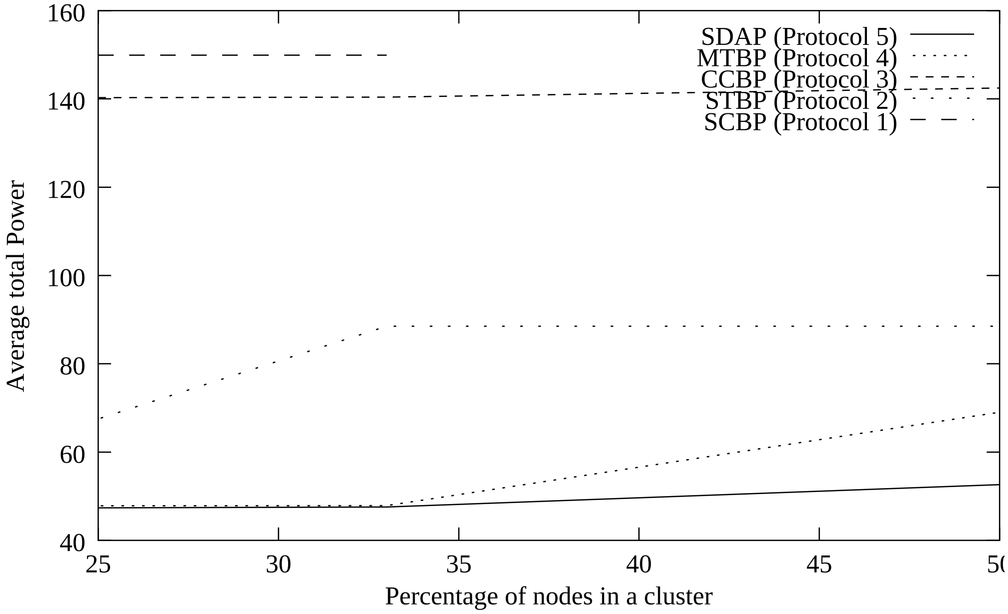


Figure 17. **Comparison of power consumed for varying percentage of nodes in a cluster by keeping number of nodes = 21 and percentage of group members = 60**

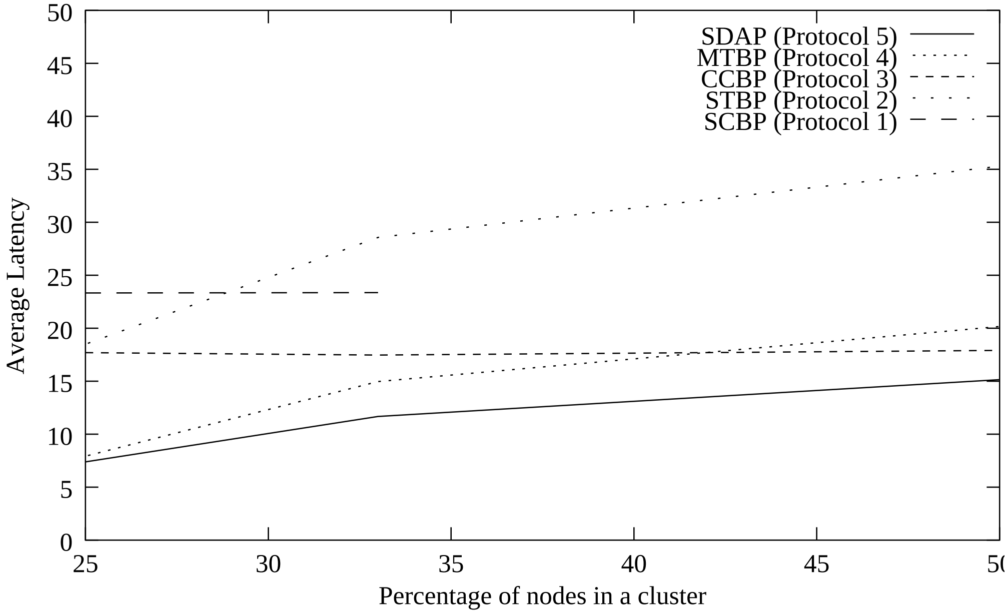


Figure 18. **Comparison of latency for varying percentage of nodes in a cluster by keeping number of nodes = 21 and percentage of group members = 60**

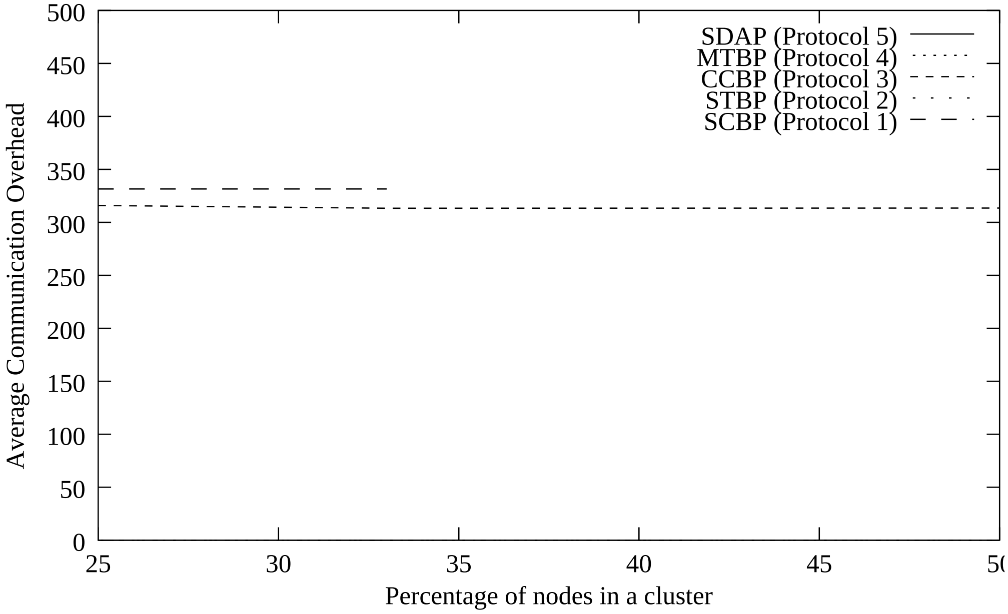


Figure 19. **Comparison of communication overhead for varying percentage of nodes in a cluster by keeping number of nodes = 21 and percentage of group members = 60**

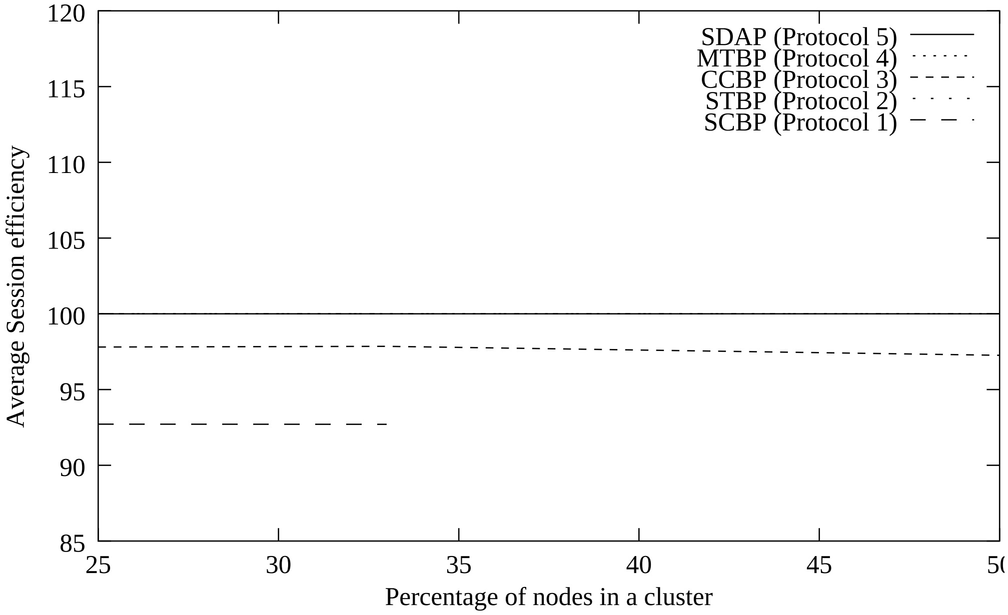


Figure 20. **Comparison of session efficiency for varying percentage of nodes in a cluster by keeping number of nodes = 21 and percentage of group members = 60**

3.3. Experiment 3. In this experiment, the performance of protocols is evaluated by varying the total number of nodes. The percentage of nodes in a cluster is kept equal to 33 and percentage of group members is kept equal to 60. Figure 21, 22, 23, and 24 shows the plots for experiment 2. Total power consumed and latency increases with the increase in total number of nodes for all protocols. Communication overhead increases with the increase in the total number of nodes for protocol 1 and 3 and session efficiency decreases with the increase in the total number of nodes for protocol 1 and 3. Protocol 5 performs best in terms of power consumption and latency. In terms of communication overhead and session efficiency, protocol 2, 4, and 5 performs equally well.

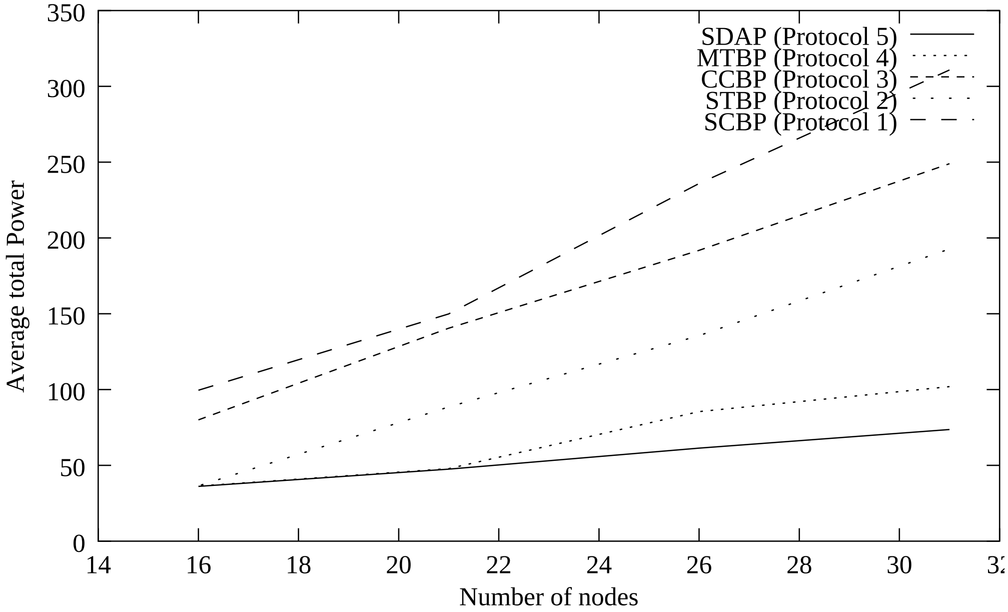


Figure 21. **Comparison of total power consumption for varying number of total nodes by keeping percentage of group members = 60 and percentage of nodes in a cluster = 33**

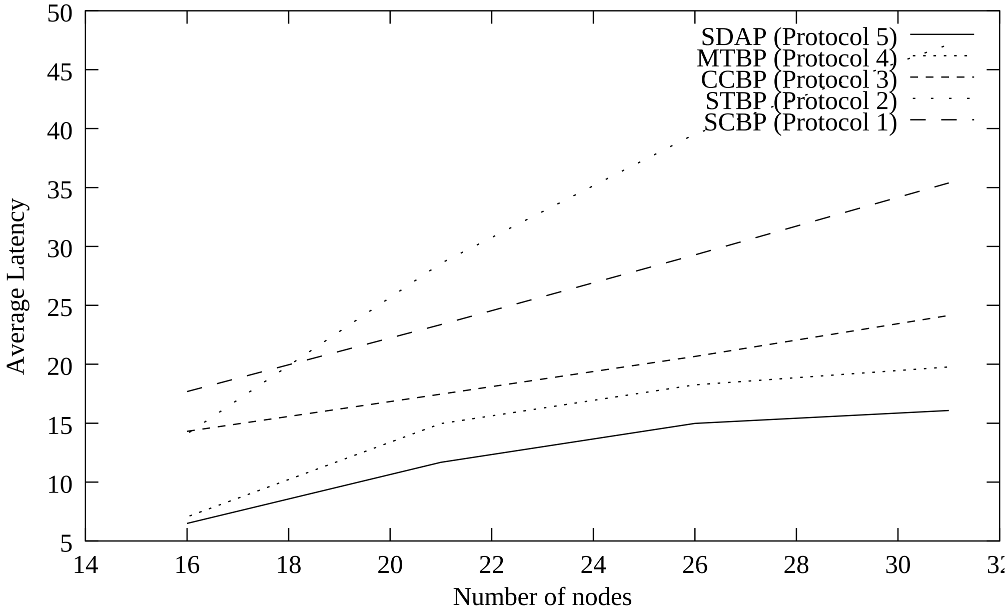


Figure 22. **Comparison of latency for varying number of total nodes by keeping percentage of group members = 60 and percentage of nodes in a cluster = 33**

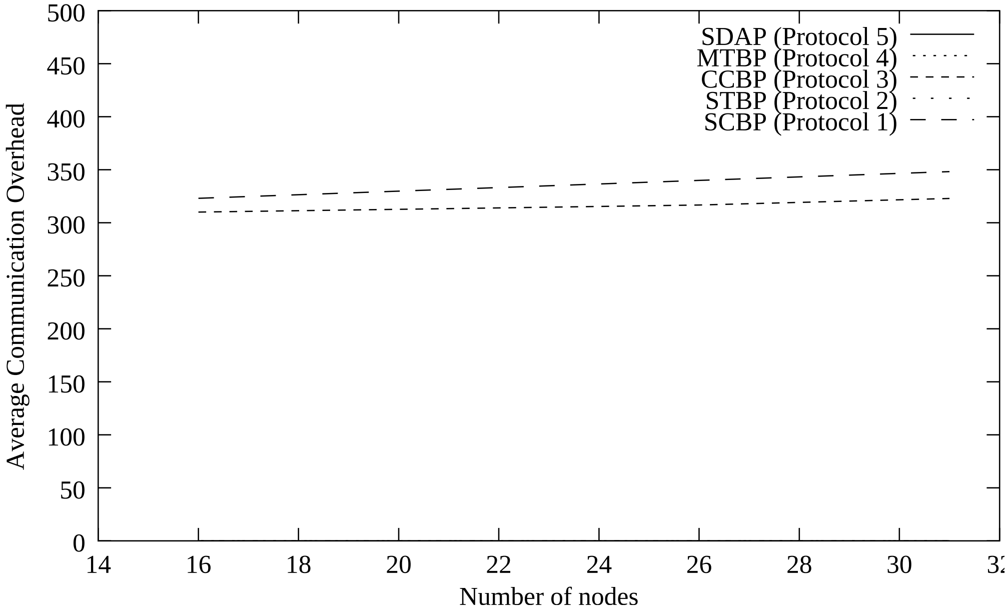


Figure 23. **Comparison of communication overhead for varying number of total nodes by keeping percentage of group members = 60 and percentage of nodes in a cluster = 33**

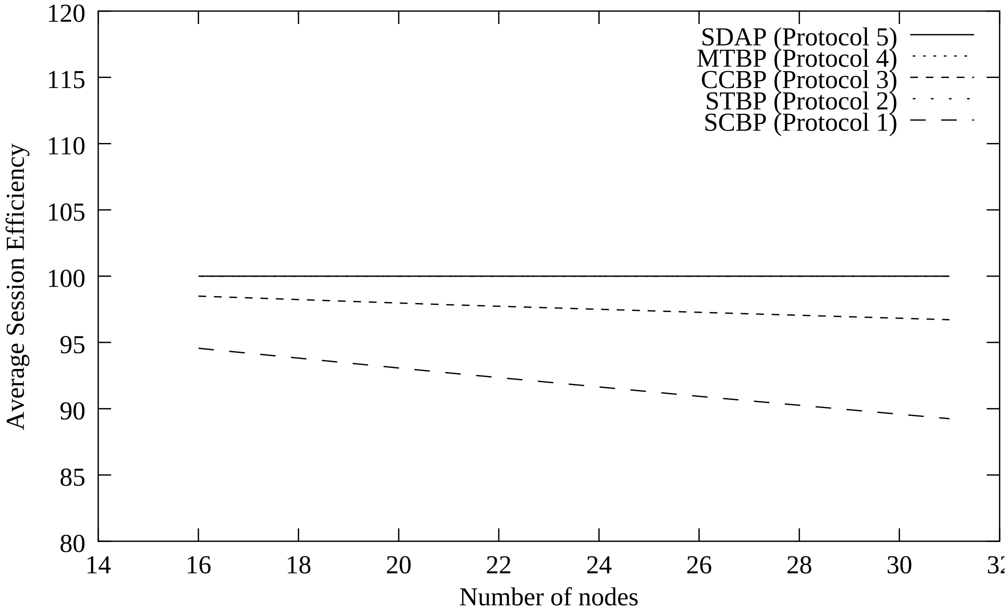


Figure 24. **Comparison of session efficiency for varying number of total nodes by keeping percentage of group members = 60 and percentage of nodes in a cluster = 33**

4. Performance Evaluation Using Analysis

The protocols are also analyzed in order to validate the simulation results. the following assumptions are made during the analysis.

- The energy consumed by a receiving antenna is not considered i.e. only energy consumed by a transmitting antenna is considered.
- There is no data compression at the time of data fusion, which is carried out after each consecutive transmission. In other words, when a sensor sends a message to its neighbor, the neighbor attaches its data to that message and sends it to the next neighbor.
- The formula is derived only for free space since this can be easily extended for human model.

Table 6 shows the list of variables used in the derivation along with the values used for plotting graphs using C. There are two kind of transmissions during the process of data aggregation; long range transmission and short range transmission. Long range transmission occurs when a sensor sends data to the base-station, which is located at a long range distance where as short range transmission occurs when a sensor sends data to another sensor located at a short range distance. Generally, the long range distance is much larger than the short range distance and thus long range transmission consumes much more energy than the short range transmission. In this analysis, the ratio of long range distance to short range distance is taken as α. In the following sections, the energy consumed and delay for the protocols is analyzed.

Table 6. **List of variables used in analysis**

E	Total energy consumed in all transmitters	
L	Latency	
E_{T_x}	Energy consumed in the transmitter	
P	Probability that a sensor transmits data successfully	
P_t	Probability that a sensor will transmit data	0.2 and 0.3
E_{elec}	Energy dissipated in the transmitter circuitry	$50nJ/bit$
ϵ_{amp}	Energy dissipated in the transmitter amplifier	$100pJ/bit/m^2$
k	Number of bits transmitted by the transmitter	$20bits$
d	Short range distance	$20m$
D	Long range distance	$200m$
N	Total number of sensors	varies from 10 to 40
C	Number of clusters or cluster-heads	varies from 10% to 40%
α	Ratio of long range distance to short range distance	10

The energy consumed in a transmitting antenna for free space is given by the formula [7]

$$\begin{aligned} E_{T_x}(k,d) &= E_{T_{x-elec}}(k) + E_{T_{x-amp}}(k,d) \\ E_{T_x}(k,d) &= E_{elec}k + \epsilon_{amp}kd^2 \qquad (4.8) \end{aligned}$$

4.1. Protocol 1. Let N be the number of sensors, P_t be the probability that a node will transmit, and P be the probability of success i.e. the node transmits successfully. In other words, P is the probability that only one node will transmit at a given time. P is given by the following formula:-

$$P = NP_t(1 - P_t)^{N-1} \qquad (4.9)$$

If E_t is the energy required for one transmission then the total energy required by a sensor to successfully transmit a message is given by

$$\begin{aligned} E_s &= PE_t + (1-P)\left[E_t + PE_t + ...\right] \\ &= PE_t + (1-P)\left[E_t + E\right] \end{aligned}$$

$$\begin{aligned} &= \frac{E_t}{P} \\ &= \frac{E_t}{NP_t(1-P_t)^{N-1}} \\ &= \frac{E_{elec}k + \epsilon_{amp}kD^2}{NP_t(1-P_t)^{N-1}} \end{aligned} \tag{4.10}$$

where E_s is the energy consumed for successfully transmitting a message. Thus, the total energy consumed is given by

$$\begin{aligned} E &= NE_s \\ &= \frac{E_{elec}k + \epsilon_{amp}kD^2}{P_t(1-P_t)^{N-1}} \end{aligned} \tag{4.11}$$

Latency is defined as the time taken by the last message to reach the base-station. Let T be the time taken for a message to reach the base-station when there is no collision. The time taken by the first message to successfully reach the base-station sent by any of a sensor is given as:-

$$\begin{aligned} L_{s1} &= PT + (1-P)[T + PT + (1-P)[...]] \\ &= \frac{T}{P} \\ &= \frac{T}{NP_t(1-P_t)^{N-1}} \end{aligned} \tag{4.12}$$

where L_{s1} is the time taken by the first message to successfully reach the base-station. The time taken by the second message to successfully reach the base-station is equal to the time taken by the first message plus additional time elapsed due to transmission and contention. It is given as:-

$$L_{s2} = L_{s1} + \frac{T}{(N-1)P_t(1-P_t)^{N-2}} \tag{4.13}$$

The factor $N-1$ is because now there are $N-1$ sensors in the competition for transmitting message to the base-station.

Similarly, the time taken by the N^{th} message to reach the base-station is

$$L = \sum_{i=1}^{N} \frac{T}{(N-i+1)P_t(1-P_t)^{N-i}} \tag{4.14}$$

where, L is the time taken by the last message to reach the base-station, which is nothing but latency.

4.2. Protocol 2. In this case, there is no collision so the energy consumed in one transmission

$$E_s = E_{elec}k + \epsilon_{amp}kD^2 \tag{4.15}$$

and thus, total energy consumed is

$$E = N\left(E_{elec}k + \epsilon_{amp}kD^2\right) \tag{4.16}$$

Latency, in this case, is nothing but the frame-length minus one. Frame-length is equal to the number of nodes.

$$L = N - 1 \tag{4.17}$$

4.3. Protocol 3. Let the number of clusters = number of cluster-heads = C. This implies that the number of cluster-members = $N - C$, and
average number of sensors in a cluster = $\frac{N}{C}$.

The cluster-members communicate using short range transmission where as cluster-heads communicate to the base-station using long range transmission.
Therefore, number of short range transmissions = $N - C$, and
number of long range transmissions = C.
All short range transmissions are of size = k, and

all long range transmissions are of size $= \left(\frac{N}{C}\right) k$ as there is no data compression.

Energy consumed by a cluster-member in one transmission is given as

$$E_s = \frac{E_{elec}k + \epsilon_{amp}kd^2}{(N-C)P_t(1-P_t)^{N-C-1}} \tag{4.18}$$

Therefore, total energy consumed by the cluster-members is

$$\begin{aligned} E_{cm} &= (N-C)E_s \\ &= (N-C)\left[\frac{E_{elec}k + \epsilon_{amp}kd^2}{(N-C)P_t(1-P_t)^{N-C-1}}\right] \\ &= \frac{E_{elec}k + \epsilon_{amp}kd^2}{P_t(1-P_t)^{N-C-1}} \end{aligned} \tag{4.19}$$

where E_{cm} is the energy consumed by all the cluster-members.

Energy consumed by a cluster-head in one transmission is given as

$$E_s = \frac{E_{elec}\frac{N}{C}k + \epsilon_{amp}\frac{N}{C}kD^2}{CP_t(1-P_t)^{C-1}} \tag{4.20}$$

Therefore, total energy consumed by the cluster-heads is

$$\begin{aligned} E_{ch} &= CE_s \\ &= C\frac{E_{elec}\frac{N}{C}k + \epsilon_{amp}\frac{N}{C}kD^2}{CP_t(1-P_t)^{C-1}} \\ &= \frac{E_{elec}\frac{N}{C}k + \epsilon_{amp}\frac{N}{C}kD^2}{P_t(1-P_t)^{C-1}} \end{aligned} \tag{4.21}$$

where E_{ch} is the energy consumed by all the cluster-members.

Thus, the total energy consumed is

$$\begin{aligned} E &= E_{ch} + cm \\ &= \frac{E_{elec}\left(\frac{N}{C}\right)k + \epsilon_{amp}\left(\frac{N}{C}\right)kd^2}{P_t(1-P_t)^{C-1}} + \frac{E_{elec}k + \epsilon_{amp}kD^2}{P_t(1-P_t)^{N-C-1}} \end{aligned} \tag{4.22}$$

Latency, in this case, is equal to the time taken for last message from a cluster-member to reach the cluster-head plus the time taken by the last message from a cluster-head to reach the base-station.

The time taken by the last message transmitted by a cluster-member to reach the cluster-head is

$$L_{cm} = \sum_{i=1}^{N-C} \frac{T}{(N-C-i+1)P_t(1-P_t)^{N-C-i}}, \tag{4.23}$$

and the time taken by the last message transmitted by a cluster-head to reach the base-station is

$$L_{ch} = \sum_{i=1}^{C} \frac{T}{(C-i+1)P_t(1-P_t)^{C-i}}, \tag{4.24}$$

Thus, the latency is

$$\begin{aligned} L &= L_{cm} + L_{ch} \\ &= \sum_{i=1}^{N-C} \frac{T}{(N-C-i+1)P_t(1-P_t)^{N-C-i}} + \sum_{i=1}^{C} \frac{T}{(C-i+1)P_t(1-P_t)^{C-i}} \end{aligned} \tag{4.25}$$

4.4. Protocol 4 and 5. Since it is assumed in the analysis that all the sensors are group members and the energy is dissipated only during the transmission time, protocols 4 and 5 will have the same formula. This is due to the fact that protocols 4 and 5 differ only when there are sensors that are not part of data gathering group. Hence, the analysis of both the protocols is similar.

The energy consumed in the transmitting antenna for free space is given by the formula [7]

$$\begin{aligned} E_{T_x}(k,d) &= E_{T_{x-elec}}(k) + E_{T_{x-amp}}(k,d) \\ E_{T_x}(k,d) &= E_{elec}k + \epsilon_{amp}kd^2 \end{aligned} \tag{4.26}$$

Let the number of sensors = N, and

the number of clusters = number of cluster-heads = C.

This implies that the number of cluster-members = $N - C$, and

average number of sensors in a cluster = $\frac{N}{C}$.

The cluster-members communicate using short range transmission where as cluster-heads communicate to the base-station using long range transmission.

Therefore, number of short range transmissions = $N - C$, and

number of long range transmissions = C.

All short range transmissions are of size = k, and

all long range transmissions are of size = $\left(\frac{N}{C}\right) k$ as there is no data compression.

Thus, the total energy consumed during one phase of data transmission (or aggregation) is given as:-

$$\begin{aligned} E &= (n-c)\left[E_{elec}k + \epsilon_{amp}kd^2\right] + \\ &\quad c\left[E_{elec}\left(\frac{n}{c}\right)k + \epsilon_{amp}\left(\frac{n}{c}\right)kD^2\right] \\ &= E_{elec}k\,(2n-c) + \epsilon_{amp}k\left((n-c)d^2 + nD^2\right) \qquad (4.27) \end{aligned}$$

Latency is the time taken by the last message to reach the base-station, which is equal to the length of the TDMA frame, thus latency is equal to frame-length minus one.

$$\begin{aligned} L &= frame - length - 1 \\ &= number\ of\ clustermembers + number\ of clusters \\ &= \left(\frac{N-C}{C}\right) + C \\ &= \frac{N}{C} + C - 1 \qquad (4.28) \end{aligned}$$

4.5. Analytical Results. Figures 25 and 26 show the $energy * delay$ performance metric versus the number of nodes for value of P_t equal to 0.2 and 0.3. The number of nodes varies from 10 to 40. The number of cluster as percentage of total number of sensors is fixed at 10%. The C program is also run for different values of percentage of clusters. It is noticed that the results are more or less the same, thus the figures for different values of percentage of clusters are not presented. Figures 25 and 26 suggest that the protocol 1 and 3 have high values of $energy * delay$. Protocol 4 and 5 are performing better than other protocols. It is also evident in the figures that difference in the performance of protocol 4 and 5 and protocol 2 is not very significant when compared to the simulation results. This is due to the assumptions that have been made during the analysis. Protocol 2 occurs a greater loss of energy for total number of receives than the protocol 5 or Protocol 4.

5. Comparison of Protocol 5 and PEGASIS

In this chapter, the protocol developed is compared with PEGASIS [9] in terms of energy and delay. PEGASIS is a protocol for data gathering in an energy-efficient manner from sensor networks. It establishes a chain structure in which each node communicates only with a close neighbor and takes turns transmitting to the base-station. In order to simplify the analysis, following assumptions are made:-

- The energy consumed by the receiving antenna is not considered i.e. only energy consumed by transmitting antenna is considered.
- There is no data compression at the time of data fusion after each consecutive transmission. In other words, when a sensor sends a message to its neighbor, the neighbor attaches its data to that message and sends it to the next neighbor.

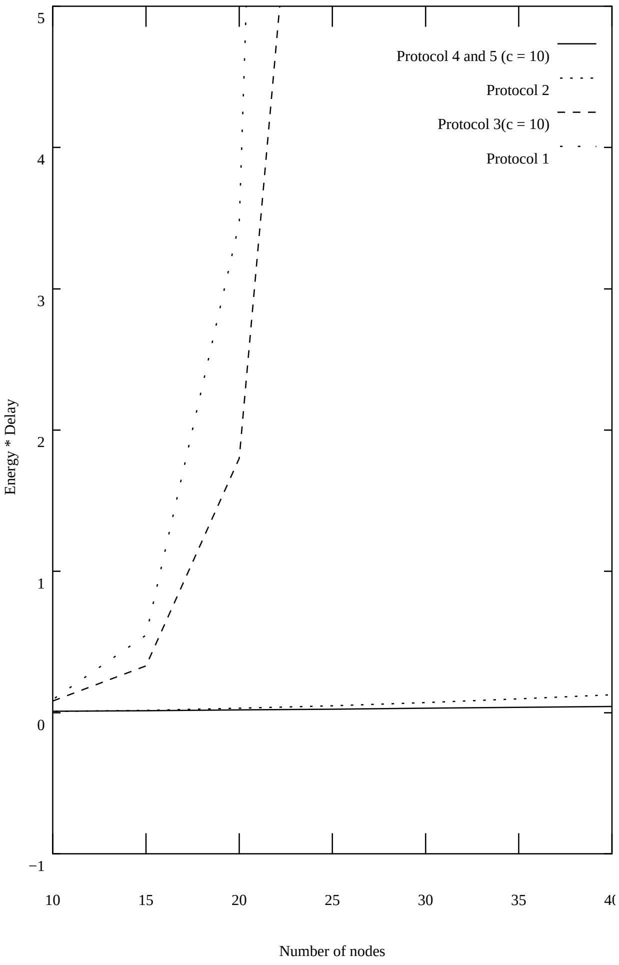


Figure 25. **Energy * delay versus number of nodes for $P_t = 0.2$**

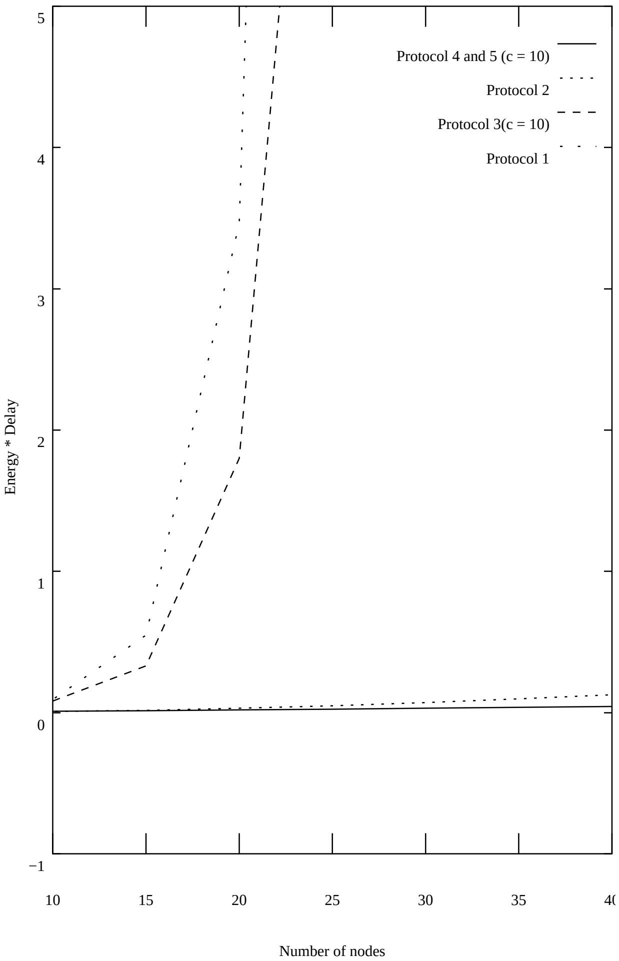


Figure 26. **Energy * delay versus number of nodes for** $P_t = 0.3$

- The formula is derived only for free space since this can be easily extended for human model.
- The number of sensors is even.

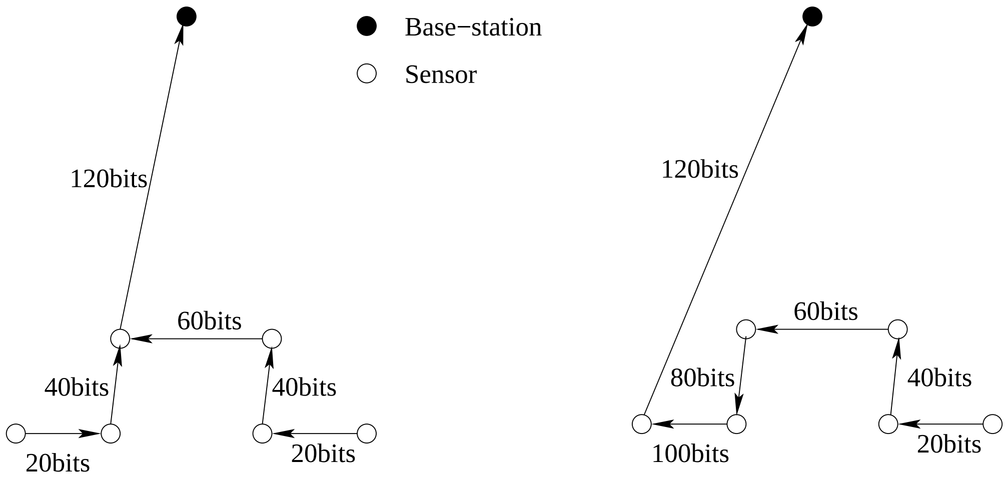
This research work is considering only the best case and the worst case situations. An example showing both the cases is shown in figure 27.

5.1. Worst Case PEGASIS. In the worst case scenario, the last sensor of the chain transmits data to the base-station. all other sensors send data to their neighbor. Say there are n sensors. First sensor sends k bit data to its neighbor, second sensor sends $2k$ bits data to its next neighbor and similarly i^{th} sensor sends ik bits data to $(i+1)^{th}$ sensor and nth sensor sends nk bits data to base-station. Thus, the total energy consumed is given as

$$\begin{aligned} E &= E_{elec}\left[1k + 2k + ... + nk\right] d^2 + \\ &\quad \epsilon_{amp}\left[1k + 2k + ... + (n-1)k\right] d^2 + \epsilon_{amp} nkD^2 \\ &= E_{elec}k\left[\frac{n(n+1)}{2}\right] d^2 + \\ &\quad \epsilon_{amp}k\left[\frac{n(n-1)}{2}\right] d^2 + \epsilon_{amp}kD^2 n \\ &= E_{elec}k\left[\frac{n(n+1)}{2}\right] d^2 + \epsilon_{amp}k\left[\frac{n(n-1)}{2}d^2 + nD^2\right] \end{aligned} \tag{4.29}$$

Latency, in worst case, is given by the number of nodes since the last message propagates through all nodes before reaching the base-station.

$$L = n; \tag{4.30}$$



(a) Best Case Scenario

(b) Worst case Scenario

Figure 27. **Data aggregation scenarios for PEGASIS**

5.2. Best Case PEGASIS. In this case, the middle sensor transmits data to the base-station. First and last sensors transmit k bits data to second and second from last sensors respectively and similarly i^{th} and i^{th} from last sensors transmits ik bits data to $(i+1)^{th}$ and $(i+1)^{th}$ from last sensors respectively. It is assumed that n is even so the total energy consumed is given as

$$\begin{aligned} E &= E_{elec}\left[2\left(1k+2k+...+\left(\frac{n}{2}-1\right)k\right)+\frac{n}{2}k+nk\right]d^2+ \\ &\quad \epsilon_{amp}\left[2\left(1k+2k+...+\left(\frac{n}{2}-1\right)k\right)+\frac{n}{2}\right]d^2+ \\ &\quad \epsilon_{amp}nkD^2 \\ &= E_{elec}k\left[\frac{n(n+4)}{4}\right]d^2+ \\ &\quad \epsilon_{amp}k\left[\frac{n^2}{4}\right]d^2+\epsilon_{amp}kD^2n \\ &= E_{elec}k\left[\frac{n(n+4)}{4}\right]d^2+\epsilon_{amp}k\left[\frac{n^2}{4}d^2+nD^2\right] \end{aligned} \tag{4.31}$$

Latency, in best case, is equal to half the number of nodes plus one, since the last message propagates through half the number of nodes plus one.

$$L = \frac{n}{2} + 1 \tag{4.32}$$

5.3. Comparison Results. The protocols are compared with the help of C programs developed using the above analysis. Figure 28 shows the results comparing the energy consumed for PEGASIS and Our Protocol. The x-axis shows the number of nodes that varies from 10 to 40 and y-axis shows the energy consumed. Figure 29 shows the results comparing the delay for PEGASIS and Our Protocol. The x-axis shows the number of nodes that varies from 10 to 40 and y-axis shows the latency. Figure 30 shows the results comparing the energy*delay metric for PEGASIS and Our Protocol. The x-axis shows the number of nodes that varies from 10 to 40 and y-axis shows the energy*latency metric. All the three plots have six curves; worst case PEGASIS, best case PEGASIS, proposed protocol with number of clusters equal to 10%, 20%, 30% and 40%. It can be seen from the plots that proposed protocol for number of clusters varying from 10% to 40% and for nodes varying from 10 to 40 performs better than the worst case PEGASIS and best case PEGASIS. Thus it can said that proposed protocol performs better than PEGASIS for all cases.

6. Comparison of Protocol 5 with an Ideal Scheduling Scheme

6.1. Description of Ideal Scheme. In the ideal scheme, first slot will be allotted to BS in all TDMA frames. Next consecutive slots in a clusters TDMA frame will be allotted to the group-members. During these slots, the base-stations TDMA frame's slot will be empty as no cluster-head will transmit during this time. Once the group-members transmission is over, cluster-heads send the data to base-station in the following slots.

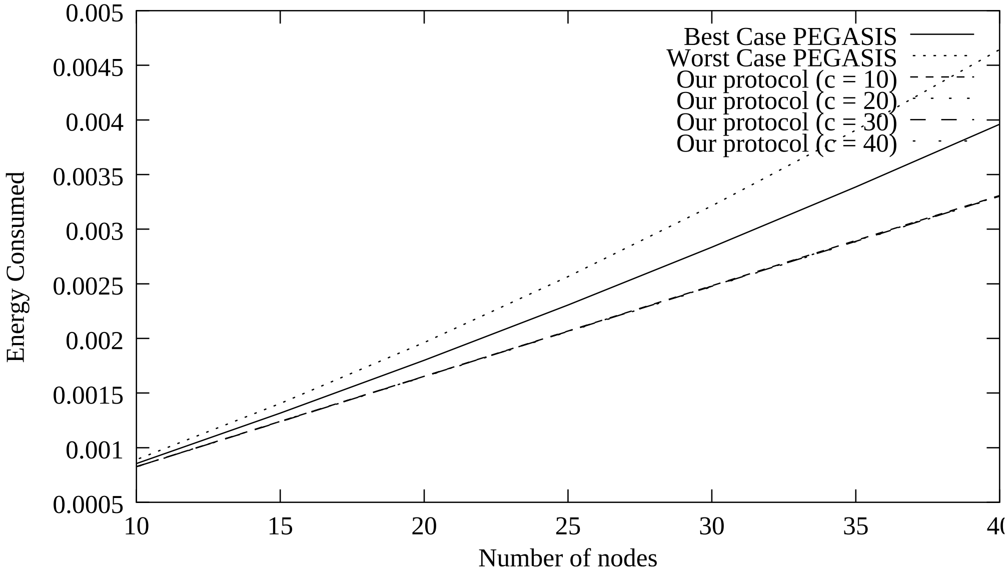


Figure 28. **Comparison of PEGASIS And Our Protocol in terms of energy consumption.**

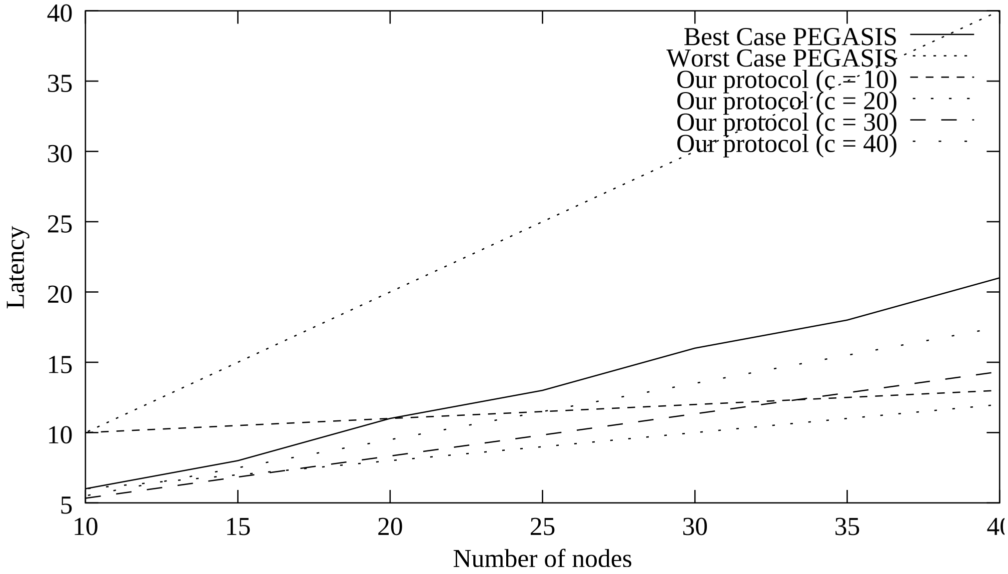


Figure 29. **Comparison of PEGASIS and Our Protocol in terms of latency.**

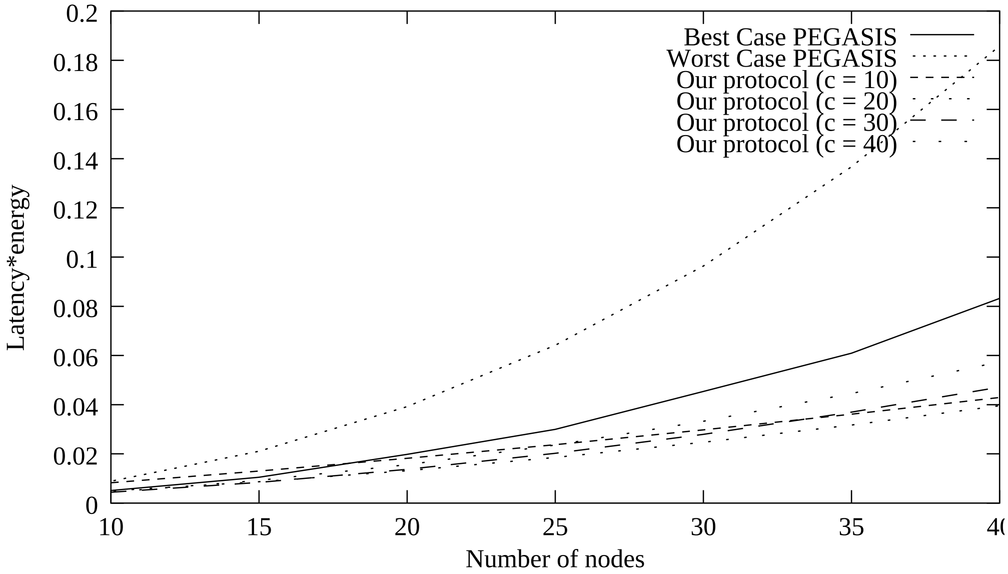


Figure 30. **Comparison of PEGASIS and Our Protocol in terms of energy*latency.**

This can be explained clearly with the help of an example. Consider 2 clusters (A and B) with 4 and 2 group-members respectively. In this case, the total length of frame will be 1+4+1 = 6 slots. In the first slot base-station will transmit. Second and third slots in cluster A will be allotted to group-members of cluster A. Fourth slot in cluster A and base-station's TDMA frame will be allotted to the cluster-head of clusterA. Second, third, fourth, and fifth slot of cluster B will be allotted to group-members of cluster B. Sixth slot in cluster B and base-station's TDMA frame will be allotted to the cluster-head of cluster B. All the remaining slots will be empty since there will not be any transmission.

6.2. Comparison with Ideal Scheme. There is slot wastage in our scheme, because we use a two level TDMA scheme as mentioned in the thesis document. The nodes are divided into clusters with a separate CDMA channel for each cluster. The cluster-head of a cluster communicate with the base-station in another channel. The slot wastage occurs due to the following reasons:-

1. Cluster-members have to transmit first and the cluster-heads then transmit data to the base-station. This results in slot wastage in clusterhead-basestation TDMA frame.
2. Number of cluster-members in a cluster is not equal so a cluster can need more time than the other clusters to collect data from members and transmit it to the base-station. Since the TDMA frames are synchronized and the size of frames is equal therefore, a cluster with less number of members have to wait, which result in slot wastage in that TDMA frame.

Any ideal scheduling algorithm will be the one, which has no slot wastage. Such an algorithm is not possible in a 2 level TDMA scheme. Consider an example of 3 clusters with 1, 4, and 7 members. The first clusters' communication will be over in 1(for member) + 1(for cluster-head) + 1(for base-station) = 3 slots. The second clusters' communication will be over in 4(for member) + 1(for cluster-head) + 1(for base-station) = 6 slots. The third clusters' communication will be over in 7(for member) + 1(for cluster-head) + 1(for base-station) = 9 slots. So it may be seen that there will be 6 slots of idle time for first cluster. Thus, an ideal scheme is not possible unless the number of cluster-members in each cluster is exactly same.

Even in the case of equal number of cluster-members in each cluster, there will be slot time wastage in clusterhead-basestation TDMA channel because the base-station has to wait for data from cluster-heads, which in-turn, wait for data from cluster-members. The above discussion suggest that an ideal scheduling algorithm which involves no slot wastage in two level TDMA schemes is not feasible but the proposed protocol can be compared with an ideal scheduling algorithm which has minimum slot wastage and no overhead. As described in protocol 5 descriptions, it is clear that protocol 5 has some overhead, which will not present in case of any ideal scheme. Thus, the protocol 5 is compared with ideal

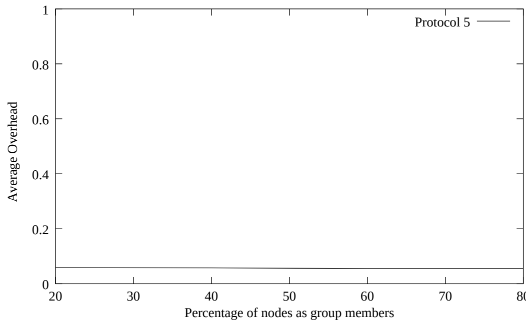


Figure 31. **Overhead in protocol 5 for experiment 1**

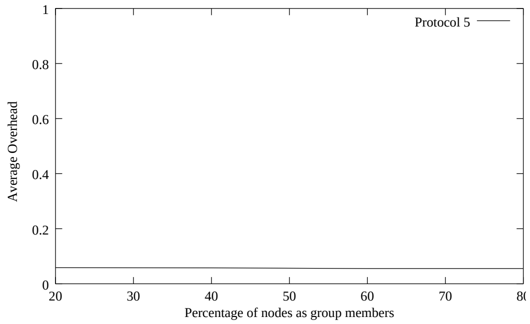


Figure 32. **Overhead in protocol 5 for experiment 2**

scheme in terms of overheads. There are two kinds of overheads: first is communication overhead, which comes from the fact that the base-station communicate with the sensors to get the group information and distribute the slot information. Second overhead is the processing overhead, which occurred when base-station computes the scheduling scheme for communication. As it is assumed that base-station has ample supply of energy and processing power, and communication cost is much cheaper than processing cost, so the processing overhead is neglected. The figures 31, 32, and 33 show the overhead in protocol 5. There is no overhead in an ideal scheme. It can be noted from the figures that the overhead in protocol 5 is very small.

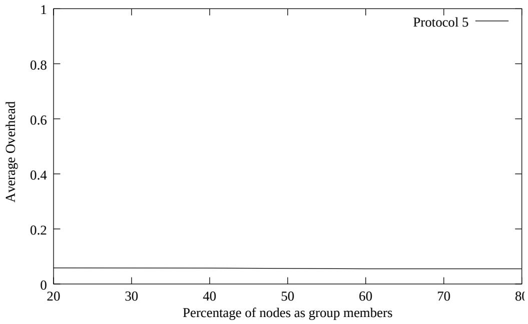


Figure 33. **Overhead in protocol 5 for experiment 3**

CHAPTER 5

Conclusions and Future Works

In this research work, a novel scheme for energy-efficient data aggregation is proposed. The proposed protocol use combination of CDMA and TDMA along with clustering to achieve high energy efficiency, high session efficiency, low communication overhead, and low latency. It is evident from the simulation results that the protocol 5 outperforms all other protocol in terms of normalized power and latency and performs equally well in terms of session efficiency as compared to Protocol 2 and 4. Protocol 5 has negligible communication overhead. This communication overhead further decreases with the increase in the session length. Communication overhead for protocol 2 and 4 is zero and for protocol 1 and 3 is more than 300%. The simulation results are validated with the help of mathematical analysis. The simulation results and analytical results confirm to the same results that the protocol 5 performs better than other protocols. The protocol 5 is also compared to an existing protocol PEGASIS with energy consumption, delay, and energy * delay as metrics. The comparison results shows that proposed protocol outperforms the PEGASIS.

This research work has assumed no data compression. The comparison of protocols with different levels of data compression at MAC layer during the time of data fusion is left for the future research work. The simulations and analysis presented in this research work need to be further validated by actual implementation in real life sensors to see their feasibility and other security issues.

REFERENCES

[1] J. Heidemann, F. Silva, C. Intanagonwiwat, R. Govindan, D. Estrin, and D. Ganesan. Building Efficient Wireless Sensor Networks with Low-Level Naming. 18th ACM Symposium on Operating Systems Principles, pp. 146-159, October 2001.

[2] M. Bhardwaj, T. Garnett, A.P. Chandrakasan. Upper Bounds on the Lifetime of Sensor Networks. In IEEE International Conference on Communications, Volume: 3, pp. 785-790, 2001.

[3] L. Schwiebert, S.K.S. Gupta, J. Weinmann, A. Salhieh, V. Shankar, V. Annamalai, M. Kochhal, and G. Auner. Research Challenges in Wireless Networks of Biomedical Sensors. In Proceedings of MobiCom, pp. 151-161, 2001.

[4] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless Sensor Networks: A Survey. Computer Networks (Elsevier) Journal, Vol. 38, No. 4, pp. 393-422, March 2002.

[5] P. Rentala, R. Musunuri, S. Gandham, and U. Saxena. Survey on Sensor Networks. http://citeseer.nj.nec.com/479874.html, 2001.

[6] P. Bhagwat, C. Bisdikian, I. Korpeoglu, A. krishna and M. Naghshineh. System Design Issues for Low-Power, Low-Cost Short Range Wireless Networking. In IEEE International Conference on Personal Wireless Communications, pp. 264-268, 1999.

[7] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan. Energy-Efficient Communication Protocol for Wireless Microsensor Networks. In Hawaii International Conference on System Sciences, 2000.

[8] V. Shankar, A. Natarajan, S.k.S. Gupta, and L. Schwiebert. Energy-Efficient Protocols for Wireless Communication in Biosensor Networks. In proceedings of IEEE Personal, Indoor, and Mobile Radio Conference, pp. D114-D118, 2001.

[9] S. Lindsey, C.S. Raghavendra. PEGASIS: Power Efficient GAthering in Sensor Information Systems. In IEEE Aerospace Conference, March 2002.

[10] S. Lindsey, C. Raghavendra, k. Sivalingam. Data gathering in sensor networks using the energy*delay metric. In Proceedings of 15th International Parallel and Distributed Processing Symposium, pp. 2001-2008, 2001.

[11] W. Ye, J. Heidemann, and D. Estrin. An Energy-Efficient MAC Protocol for Wireless Sensor Networks. In Proceedings of the 21st International Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2002), New York, June 2002.

[12] B. Krishanamachari, D. Estrin, and S. Wicker. The Impact of Data Aggregation in Wireless Sensor Networks. In International Workshop of Distributed Event Based Systems (DEBS), Vienna, Austria, July 2002.

[13] J. Kuri, S. Kasera. Reliable Multicast in Multi-access Wireless LANs. In *Proceedings of INFOCOM '99, New York, March 1999.*

[14] J.J. Garcia-Luna-Aceves, E.L. Madruga. The Core Assisted Mesh Protocol. In *IEEE Journal on Selected Areas in Communications, vol. 17, no. 8, Aug 1999.*

[15] S.J. Lee, M. Gerla, and C.C. Chiang. On-Demand Multicast Routing Protocol, In *Proceedings of IEEE WCNC'99, New Orleans, LA, pp. 1298-1304, September 1999.*

[16] S.K.S. Gupta and P.K. Srimani. Cored-Based Tree with Forwarding Regions(CBT-FR); A Protocol for Reliable Multicasting in Mobile Ad Hoc Networks. In JPDC Special Issue on Routing, pp. 1249, 2001.

[17] T. Ozaki, J.B. Kim and T. Suda. Bandwidth-Efficient Multicast Routing Protocol for Ad-Hoc Networks. In proceedings of IEEE ICCCN, pp. 10-17, 1999.

[18] E.M. Royer, C.E. Perkins. Multicast Operation of the Ad hoc On-Demand Distance Vector Routing Protocol. In *Proceedings of Mobicom'99, Seattle, WA, pp 207-218, August 1999.*

[19] E. Bommaiah, M. Liu, A. McAuley, and R. Talpade. AMRoute:Ad hoc Multicast Routing Protocol. *Internet Draft, draft-talpade-manet-amroute-00.txt, Aug 1998.*

[20] J.E. Wieselthier, G.D. Nguyen, and A. Ephremides. Multicasting in Energy-Limited Ad-Hoc Wireless Networks. In proceedings of IEEE MILCOM, pp. 723-729, 1998.

[21] J.E. Wieselthier, G.D. Nguyen, and A. Ephremides. Algorithms for Energy-Efficient Multicasting in Ad-Hoc Wireless Networks. In proceedings of Mobile Networks and Applications (MONET), pp. 1414-1418, 1999.

[22] J.E. Wieselthier, G.D. Nguyen, and A. Ephremides. On the Construction of Energy-Efficient Broadcast and Multicast Trees in Wireless Networks. In proceedings of IEEE Infocom, pp. 585-594, 2000.

[23] I. Chlamtac, and S. Kutten. Tree-Based Broadcasting in Multihop Radio Networks. IEEE Transactions On Computers, VOL. C-36, NO. 10, pp. 1209-1223, 1987.

[24] I. Stojmenovic, M. Seddigh, and J. Zunic. Internal Nodes Based Broadcasting Algorithms in Wireless Networks. In proceedings of the 34th Annual Hawaii International Conference on System Sciences (HICSS-34).

[25] J. Wu and H. Li. On Calculating Connected Dominating Set for Efficient Routing in Ad Hoc Wireless Networks. In proc. DIAL M, Seattle, pp.7-14, August 1999.

[26] The Network Simulator - ns-2, http://www.isi.edu/nsnam/ns/, 2001

[27] Raj Jain. The art of computer systems performance analysis: techniques for experimental design, measurement, simulation, and modeling. New York: Wiley Publication, 1991.