UNIVERSITY OF CINCINNATI

DATE: 11/25/2003

I, Abinash Mahapatra ,

hereby submit this as part of the requirements for the Degree of:

Master of Science

In:

Computer Engineering

It is entitled:

QOS AND ENERGY AWARE ROUTING FOR REAL TIME TRAFFIC IN WIRELESS SENSOR NETWORKS

Approved by:

Dr Dharma P. Agrawal

Dr James Caffery

Dr Yiming Hu

QOS AND ENERGY AWARE ROUTING

FOR REAL TIME TRAFFIC IN WIRELESS

SENSOR NETWORKS

A thesis submitted to the
Division of Graduate Studies and Advanced Research
Of the University of Cincinnati

In partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

In the Department of Electrical & Computer Engineering and
Computer Science,

2003

By

Abinash Mahapatra

B.E. Electrical Engineering, University of Roorkee,
Roorkee, India, 2000.

Thesis Advisor and Committee Chair:
Dr. Dharma P Agrawal

Abstract

Wireless sensor networks are being built to facilitate automated information gathering in military, industrial, environmental and surveillance applications. Many such applications of Sensor Networks require desired/pre-fixed QoS (Quality of Service: packet delivery within a defined deadline) guarantees as well as high reliability. These applications demand high packet delivery ratio and are extremely delay sensitive. However, certain factors limit the ability of the multihop sensor network to achieve the desired goals. These factors include the delay caused by network congestion, hot regions in the network, limited energy of the sensor nodes, packet loss due to collisions and link failure. In this work, an energy aware dual-path routing scheme for real time traffic is proposed, which balances node energy utilization, increases the network lifetime, reduces the routing delay across the network by taking network congestion into account and increases the reliability of the packets reaching the destination by introducing minimal data redundancy. This paper also introduces an adaptive prioritized Medium Access Layer (MAC) to provide a differentiated service model for real time packets. Our intuitive claims are well supported by simulation results.

Acknowledgments

My M.S at University of Cincinnati has been a great learning experience for me. For this, I would like to thank Dr. Dharma P. Agrawal for his help, support and guidance.

I would also like to thank Dr. Caffery and Dr. Hu for taking the time to review my thesis and participating in my defense committee.

I would also like to thank my family and friends for their wishes.

I would also like to thank all my lab-mates, for their company and their help.

Thank you everyone for making my stay at the University of Cincinnati so enjoyable.

Contents

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Wireless Sensor Networks

The paradigm of Ad Hoc network dates back to the '70s, when these networks were originally called *packet radio networks* [2]. The primary objective of developing such networks was to develop military and surveillance applications. Subsequently, the need for developing smart sensing devices, coupled with recent advances in MEMS technology, resulted in introduction of cheap, miniturized sensor nodes [3] with formidable sensing capability. In the Smart Dust project at UC Berkeley [3] and Wireless Integrated Network Sensors [4] project at UCLA, researchers have tried to realize a functional network comprising of large number of sensors with wireless communication capability. These small, battery operated nodes, equipped with sensing, computing and wireless communication capabilities are finding increased usage in many civil, industrial and military applications.

A sensor network consists of a large number of smart disposable micro sensors deployed in a group, either on the ground, in the air, on the buildings, on the vehicles or under water, all interconnected by wireless radio to track and detect changes in the physical phenomenon around their environment. Each such micro sensor is an integral part of a sensor node that has an embedded processor, a low power radio, and a tiny battery and becomes a part of the multi-hop wireless ad hoc network. This is especially desirable in those applications where sensors may be thrown in an inhospitable terrain with the aid of an unmanned vehicle or a low flying aircraft. Distributed sensor networks aggregate complex data to provide rich, multi-dimensional pictures of the environment, which a single complex sensor, working alone cannot provide. Multiple sensors can help overcome line of sight issues and environmental effects by placing sensors close to the event of interest. Wireless sensor networks basically collect real time data, and offer unprecedented opportunities for a broad spectrum of applications such as industrial automation, situation awareness, tactical surveillance for military applications, and environmental monitoring. What makes sensor nodes so attractive is its miniaturization, low cost, low power radio and autonomous ad hoc connectivity; eliminating the need for any human intervention. As sensor data is mostly statistical in nature, the bandwidth requirement is low in sensor networks and data rate is expected to be of the order of 1-100kb/s. A sensor network can be said to be the nervous system of our engineering network world, extracting and transmitting useful and timely information reliably for efficient decision support and quick corrective actions.

A wireless sensor network is capable of functioning in hostile, inaccessible terrain without any infrastructure. However, one of the most important applications of the wireless sensor network is to provide unmanned surveillance of terrains where it is extremely difficult to bring up a traditional wireless infrastructure. These applications include detecting a forest fire, habitat monitoring, detecting radiation leakage, impurity level in sea or river-bed discharge, intrusion detection for military purposes, etc. A lot of these applications are delay sensitive and need the information to be transmitted to a central controller reliably within a certain deadline.

However, a wireless sensor network is resource constrained [1] and poses many challenges while designing an efficient routing protocol for deadline driven traffic. Due to the limited battery power of the sensor nodes, it is extremely important that the routing be energy efficient which aims at increasing the network lifetime. Besides limited energy, there are other factors which hinder the goal of transferring time critical information reliably across the network. The most common factor is the delay in routing. In typical routing schemes designed for ad hoc networks, like AODV [5], DSR [6] a lot of delay is caused because these schemes do not take advantage of the shortest path to the destination. If the sensor nodes are GPS enabled, then we can take the maximum advantage of the radio range by sending the packet to the node closest to the destination, thus, saving the delay by limiting the number of hops. Other issues include the delay caused by congestion at a node and hot regions in a network, which can introduce significant delays in the delivery of real time packets. Node mobility, link failure and node failure also add to the packet loss and affect the reliability of data delivery. All these factors together reduce the probability of successful packet delivery at the destination. Consequently, with an increase in the number of intermediate hops, the probability of packet loss also increases.

1.2 Characteristics of a Wireless Sensor Network

-Limited Energy: Sensor nodes are battery operated, which makes their operational lifetime short. Each node consumes a significant amount of energy in performing idle listening, receiving and transmitting packets. It costs 3 Joules of energy to transmit 1 Kb of data to a distance of 100 m [20]. On the other hand, a general-purpose processor with a modest specification of 100 million instructions per second (MIPS) processing capability can execute 300 million instructions for the same amount of energy.

-**Physical Limitation**: Sensor nodes are prone to failure because of the harsh environmental conditions where they are deployed. As nodes are randomly dispersed in the field, they might not be in each others transmission ranges all the time and they are unlikely to be replaced on failure.

-**Hardware Limitation**: The biggest hurdle is to make the sensor nodes cost effective, and achieving time synchronization, which is a critical component of the underlying infrastructure for any distributed system. Researchers are currently pursuing energy efficient physical layer design that provides a trade off between energy and quality/latency.

-**Data Aggregation**: Since the sensor nodes in close proximity may have similar data, the data need to be aggregated before transmission to the central controller. There are various clustering schemes, which achieve the desired goal of aggregating the data before transmitting to avoid redundancy and increase network lifetime. In all of the clustering schemes, a central controller cluster head assumes the responsibility of gathering data from its members. The data is further routed to the nest hop either using inter cluster gateways or the cluster head directly communicates with the central base station.

-**Distributed and Localized Algorithms**: Localized algorithms are attractive as they have very low overhead and can be implemented in a distributed manner. They make decisions based on local information collected by the local aggregation point to maintain a low control packet overhead to achieve scalability. They are also advantageous in mobile environments where the delay associated with collection/dissemination of global information can deteriorate the overall performance.

-**Collaborative Processing**: Local collaboration among sensor nodes is encouraged for enhanced tracking, detection and classification. Collaborative multiresolution interpretation and fusion of sensor data are essential in local information processing. Depending on available energy resources and delay tolerance of the application, adaptive signal processing provides a trade off between timeliness and fidelity of the signal processing.

1.3 Wireless Sensor Network Architecture

Energy efficiency and application specific requirements typically drive the design of the network infrastructure. Possible areas where energy consumption can be minimized while designing the communication architecture are topology discovery overhead, routing protocol overhead, actual transmission of data and period for the idle radio listening. Multihop communication in a sensor network can effectively overcome shadowing and path loss effects, if the node density is high. In fact propagation losses and channel capacity limitations are helpful in employing spatial frequency reuse.

Flat Network Architecture

In flat network architecture shown in Figure 1.1(b), all nodes are equal and connections are setup between nodes that are in close proximity to establish radio communications, constrained only by connectivity conditions and security limitations. Route discovery can be carried out in sensor networks using flooding that do not require topology maintenance as it is a reactive way of disseminating information. In flooding, each node receiving data packets broadcasts till all nodes or the node at which the packet was originated, gets back the packet. But in sensor networks, flooding is minimized or avoided as nodes could receive multiple or duplicate copies of the same data packet due to nodes having common neighbors or sensing similar data. Gossiping is a relatively less energy consuming variant of flooding where the packet is forwarded to a few randomly selected nodes. Intanagonwiwat et. al. [22] have introduced a data dissemination paradigm called directed diffusion for sensor networks, based on a flat topology. The query is disseminated (flooded) throughout the network with the querying node acting as a source and gradients are setup towards the requesting node to find the data satisfying the query. Events (data) start flowing to wards the requesting node along multiple paths. To prevent further flooding, a small number of paths can be reinforced among a large number of paths initially explored to form the multihop routing infrastructure so as to prevent further flooding. One advantage of the flat networks is the ease of creating multiple paths between communicating nodes, thereby alleviating congestion and providing robustness in the presence of failures. Route selection can also be made according to the traffic requirements, e.g., smaller delay and low capacity paths can be used for voice traffic, while voluminous data such as maps or video information can be sent over high capacity but possibly longer delay routes. In this protocol, data may be aggregated at intermediate nodes, forming virtual clusters in a locality as is done in the hierarchical topology. Rumor routing [23] is another routing protocol for flat networks that looks at routing queries to the nodes, which have observed a particular event. It creates paths leading to each event so that a query, which is generated, can be routed randomly till it finds the event path instead of flooding it across the network. The rumor routing algorithm uses a set of long-lived agents, which create paths that are directed towards the events they encounter. It is tunable and allows for trade offs between the setup overhead and the delivery reliability.

Hierarchical Network Architecture

One way of minimizing the data transmissions over long distances is to cluster the network so that signaling/control overheads can be reduced, while critical functions such as media access, routing, and connection setup could be improved. While all sensor nodes typically function as switches/routers, one node in each cluster is designated as the cluster head (CH), and traffic between nodes of different clusters must always be routed through their respective CHs or gateway nodes that are responsible for maintaining connectivity among neighboring CHs.

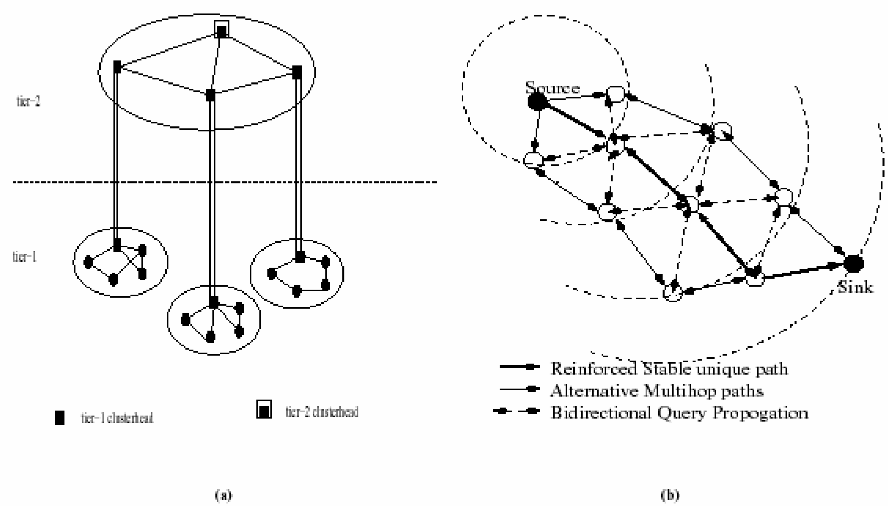

Figure 1.1: *Comparison of Hierarchical and flat networks*

Figure 1.1 (a) shows two-tier sensor network architecture; Figure 1.1 (b) shows a flat sensor network that uses directed diffusion for Routing.

The number of tiers within the network can vary according to the number of nodes, resulting in hierarchical network architecture as shown in Figure 1.1(a). A proactive clustering algorithm for sensor networks, called LEACH is one of the initial data gathering protocols introduced by MIT's researchers Heinzelman et. al. [21]. Each cluster has a CH that periodically collects data from its cluster members, aggregates it and sends it to an upper level CH. Only the CH needs to perform additional data computations such as aggregation, etc. and the rest of the nodes sleep unless they have to communicate with the CH. In order to evenly distribute this energy consumption, all the nodes in a neighborhood take turns to become the CH for a time interval called the cluster period.

1.4 Energy Aware Routing

Since Wireless Sensor Nodes have limited energy, it is extremely important to conserve the energy by using energy aware routing. By employing energy aware routing the system lifetime is enhanced, network partition due to holes (nodes with no energy) is avoided and energy balance is maintained through out the Sensor Network. In [24], Akyilidiz classifies routes based on the energy efficiency criterion and is illustrated in Figure 1.2. RE is the residual or remaining energy at the intermediate nodes. The link cost for transmission between any two nodes is represented by different values of t.

Route 1: Sink-A-B-Source, total RE= 4, total t=3

Route 2: Sink-A-B-C-Source, total RE= 6, total t=6

Route 3: Sink-D-Source, total RE= 3, total t=4

Route 4: Sink-E-F-Source, total RE= 5, total t=6

There values of energy consumption and remaining energy are used to classify routes as follows.

1. *Minimum Energy*: The route that consumes minimum energy to transmit the data packets between the sink and the sensor node is the minimum energy or minimum transmission energy (MTE) route.

2. *Maximum Residual Energy*: The route that has maximum total available energy is preferred.

3. *Minimum hop route*: The total cost of the route is the sum of each hop cost in the path in a minimum hop route. If all edges have the same cost, minimum hop route is same as minimum energy route.

4. *Maximum minimum residual energy node route*: The node that has minimum residual energy in a route is the bottleneck node. The route whose bottleneck node has the maximum available energy is preferred in this routing also called the max min routing

Probabilistic routing for obtaining uniform energy consumption using destination initiated reactive protocol like directed diffusion protocol, has been explored where instead of maintaining one optimal path, a set of good paths is maintained and chosen based on the probability that depends on the energy consumption of each path.

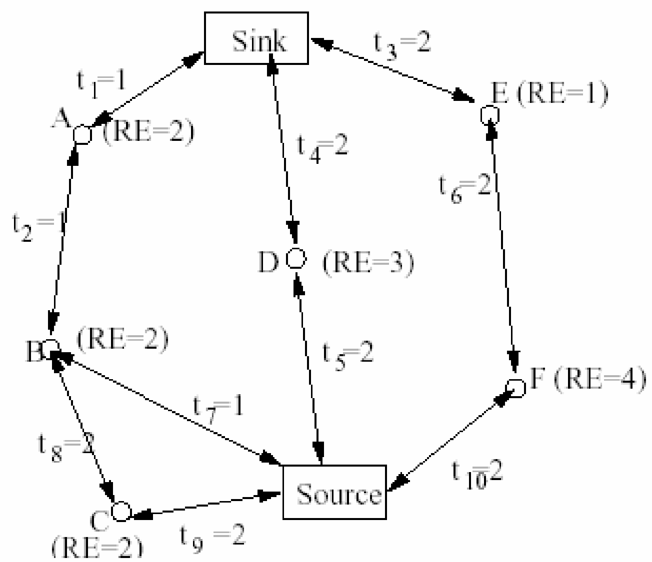

Figure 1.2: Energy Efficient Routes

1.5 Location Based Routing

By exploiting the geographic location information using GPS (Global Positioning System), we can employ location based routing. Geographic routing schemes by Karp *et. al.* [7], selects the next hop based on the geographic proximity to the destination. With every hop the packet reaches closer to the destination. By using geographic forwarding, the radio range of the nodes can be used to the maximum extent, thus reducing the number of hops. Another version of geographic forwarding is GEAR (Geographic and Energy Aware Routing Algorithm) [8]. Instead of using a classical greedy geographic forwarding, GEAR uses an energy aware cost estimate to balance the energy consumption among neighbors lying in the direction of the sink. GEAR gives better system lifetime than geographic routing due to its energy balancing feature. Along with utilizing the radio range of the node, another advantage of the location based routing is its stateless nature. The routing is not affected by cache pollution and the nodes need not store the entire route information as done in AODV [5], DSR[6].

1.6 Motivation of our Research

A lot of Sensor Network Applications need the packet to reach the destination within a certain deadline. However, due to the congestion in the network, node mobility, link failure, collisions and energy, the time critical packets may suffer unacceptable delays. To overcome the restrictions imposed by aforementioned factors, we have to reduce the number of hops a packet takes to reach the destination by utilizing the GPS and the radio range of the node. However, simple geographic forwarding can cause congestion at specific nodes, leading to significant delays. Routing should thus, factor node congestion also at the forwarding nodes to deliver packets within a given deadline. At the same time, it is equally important that the routing protocol be energy aware. Energy aware routing tries to increase the network lifetime by performing uniform resource utilization and tries to route packets in a way that, energy consumption is distributed uniformly across the forwarding nodes. Besides, since the packet information is extremely critical, we also need to ensure a reliable delivery of the data to the destination. Reliability can be significantly improved by injecting minimal redundant information in the network. Data redundancy, in spite of its routing and energy overhead, can increase the probability of successful packet delivery at the destination and provide high reliability. However, the usefulness of aforementioned techniques in reducing packet delay is often limited by the delay at the MAC layer. This work also introduces an adaptive prioritized MAC, which assigns higher priority to real time packets and reduces the MAC delay for time critical data.

1.7 Organization of the Dissertation

The dissertation is organized as follows.

Chapter 2, gives an overview on the related work done in the field of QoS aware routing in sensor networks. This section discusses the merits and limitations of the schemes. Chapter 3, provides a detailed discussion on the QoS and Energy Aware Routing scheme. In Chapter 4 we provide the performance evaluation of the scheme, comparing it with existing location based schemes like GPSR [7], GEAR [8]. The results show the contribution of our scheme in improving packet delivery percentage, packet delays and network lifetime, etc for various traffic scenarios. In Chapter 5 we present the conclusions and inferences of our work and describe a few extensions that could be made in future.

CHAPTER 2

QoS and Energy Aware Routing in Sensor Networks

2.1 Related Research in Real Time Routing

There has been a significant research in the area of real time routing in wired networks ([9], [10]). The wired networks, unlike wireless sensor networks, are not limited by energy, mobility, node failure due to physical reasons, and lack of a centralized controller. It is therefore, easier to design and model a real time wired network system. However, due to inherent problems of multihop wireless sensor networks, the design of a routing protocol which is both QoS and energy aware, poses many new challenges and not much work has been done in this direction. Existing on demand routing algorithms for ad hoc networks like AODV [5], DSR [6] do not consider time deadlines, energy or congestion at the forwarding nodes while routing a packet to its destination. GPSR [7] maintains stateless information; however, it doesn't take into consideration, the congestion or the energy of the intermediate nodes. GEAR [8], takes into consideration the energy and the geographic location while forwarding the packet, but does not factor node congestion or does not ensure reliability of data packets. GEAR also does not prioritize the real time packets over non real time packets to ensure better packet delivery (in time) for deadline driven traffic. In [25], Zorzi and Rao suggest a geographic forwarding scheme where contention is done at the receiver's side. This scheme is not reliable because of possible packet loss in case of a collision. Also the receiver contention scheme, only considers geographic proximity and does not take in to account the energy and congestion at other nodes.

One of the most common ways of ensuring real time packet delivery is to flood the network with the information as shown in Figure 2.1. By flooding, we increase the probability of the packet reaching the destination, thus ensuring reliability. However, flooding has extremely poor forwarding efficiency and results in a lot of redundant transmissions, increased energy consumption, and hence decreased network lifetime. The situation becomes more serious, when packets are getting transferred between multiple source destination pairs at the same time. Due to increased traffic in flooding, there are more collisions, increased MAC layer delays and hence the average packet delays increase.

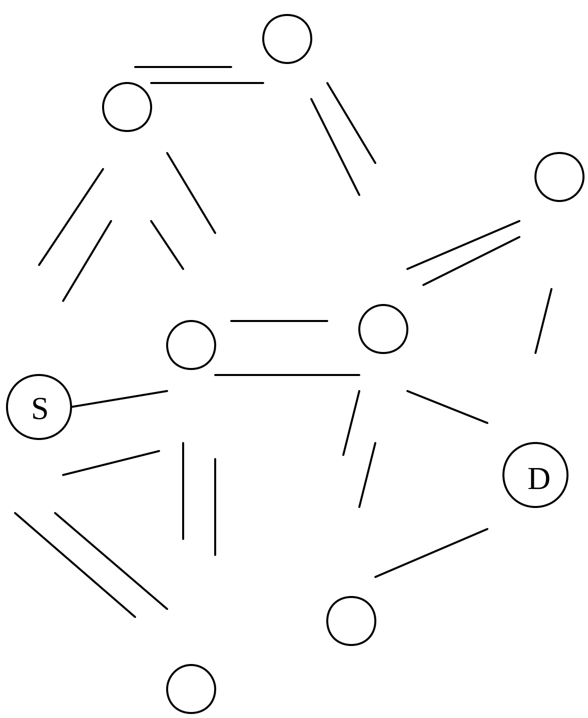

Figure 2.1 Flooding

A better approach is suggested in [11], where sets of disjoint paths is maintained from source to destination over which the data is transmitted. This Split Multipath Routing (SMR) protocol builds maximally disjoint paths. Multiple routes, of which one is the shortest delay path, are discovered on demand. Established routes are not necessarily of equal length. Data traffic is split into multiple routes to avoid congestion and to use network resources efficiently. Providing multiple routes is beneficial in network communications, particularly in mobile wireless networks where routes are disconnected frequently because of mobility and poor wireless link quality. When the source has data packets to send but does not have the route information to the destination, it transmits a RREQ packet. The packet contains the source ID and a sequence number that uniquely identify the packet. When a node other than the destination receives a RREQ that is not a duplicate, it appends its ID and re-broadcasts the packet. By doing so, we have a greater chance of finding maximum disjoint paths, as explained in Figure 2.2.

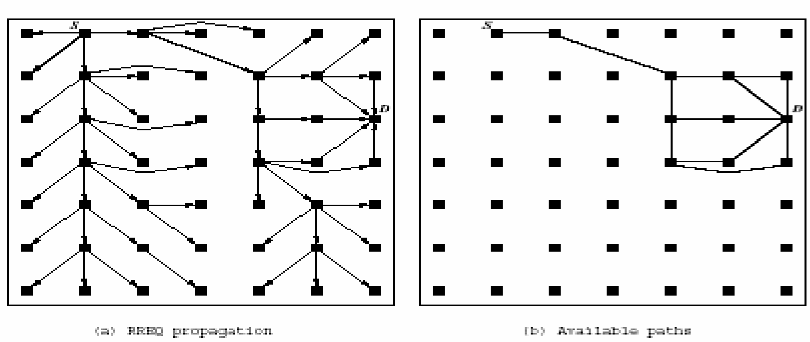

Figure 2.2 a. Overlapped Multiple Routes

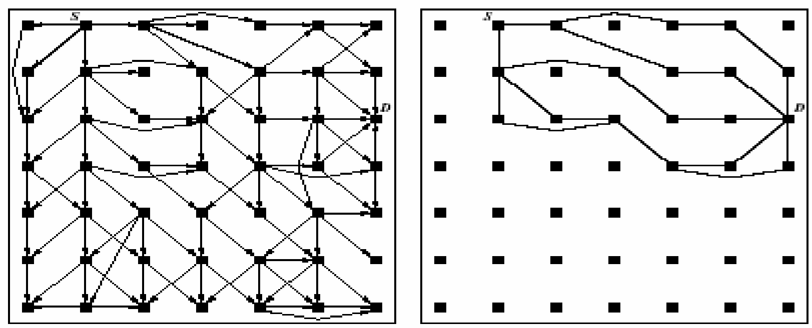

Figure 2.2 b. Multiple Routes with Maximally disjoint paths

Certain schemes like [12] require both GPS and GIS (Geographic Information System) capability to find out the best route. This results in higher infrastructure cost and needs a central controller to find out the shortest path. The SPEED protocol [13] achieves the goal of forwarding the packets closer to the destination and takes into account, the presence of hot regions and congestion at forwarding nodes into its routing strategy. However, it does not take into account the energy of the forwarding nodes so as balance the node energy utilization. Furthermore, the region it chooses for forwarding does not dynamically depend on the deadlines of the packets. SPEED also offers low reliability since it does not transmit any redundant data packets and uses a single route for data delivery.

There are other strategies to choose an optimal path for real time communication like minimal load routing [14], minimal hop routing, shortest distance path [15] etc. But these strategies do not specifically support the stateless architecture and the energy constraint of the sensor networks.

2.2 Issues in Existing Real Time Routing Schemes

All the above-mentioned schemes have the following certain shortcomings which makes them unsuitable for real time packet delivery.

- The routing strategy does not take into consideration the energy of the neighboring nodes.
- Mobility pattern of the nodes is ignored.
- Data reliability with minimal redundant duplication is not taken into account.
- The schemes do not differentiate between real time and non real time packets at the MAC layer.
- Packet forwarding does not dynamically depend on the remaining time of the deadline and the distance to the destination.

CHAPTER 3

QoS and Energy Aware Routing

3.1 Proposed Protocol

Our proposed routing protocol takes into consideration all the factors mentioned in section 2.2, which results in lower packet delivery and increased packet delays. We also try to maintain the energy balance so as to avoid network partition.

3.1.1 Protocol assumptions

The proposed routing scheme considers packet deadline, energy of the forwarding nodes and congestion at intermediate nodes to deliver real time traffic. It also introduces data redundancy by duplicating data packets at the source node to increase reliability. The basic assumptions of this scheme are:

- Nodes are GPS enabled and each node is aware of its geographic location. Our protocol uses geographic information to make routing decisions.
- Node distribution is uniform and the node density is high enough to avoid network partition. Sensor nodes are deployed in large numbers; hence it is a valid assumption. In the event of network partitions, a packet will be dropped.
- Each node is assigned a unique ID to helps us identify one node from other neighboring nodes.

- Presence of IEEE 802.11b MAC to facilitate reliable wireless communication.
- Radio Range of all the nodes is assumed to be equal to “R”. Range “R” is not affected by change in the energy of the nodes as time progresses.
- Network lifetime is defined as the time when the first node is depleted of its battery power and is rendered dead.
- All the sensor nodes start with the same energy before any traffic is routed through them.

3.1.2 Overview of the proposed approach

The basic working of our scheme is as follows. Each node exchanges periodic beacon messages (HELLO_PKT) with its neighboring nodes and maintains a neighbor table. Each entry in the neighbor table stores the geographic location of a neighboring node, the energy left, the estimated time delay (which includes the propagation delay and the MAC layer back off time) incurred by a HELLO_PKT in reaching from the neighboring node to this node and the mobility factor (indicating the frequency at which the node is changing locations). When a node has a packet to deliver, it computes its “urgency factor” which depends on the remaining distance and the time left to deliver the packet. Based on the calculated urgency factor, the routing protocol determines a distance “r” the packet needs to be pushed closer to the destination. The value of “r” is dynamic and is influenced by the “urgency factor” of the data packet. For extremely time critical packets, it is close to the radio range “R” of the sensor node and is smaller for lesser critical packets. Once “r” has been computed, routing protocol computes a priority factor, as explained below, for each of the neighboring nodes which are “r” units closer to the final destination. It then pushes the data packet “r” units closer to the destination by transmitting the packet to the neighbor node with the highest priority. The only exception to this rule is at the source, where the source sends a copy of the data packet to another neighbor node with second highest priority as well. This kind of data duplication is done only at the source node to achieve reliability by introducing minimal data redundancy. Figure 3.1, illustrates the working of the routing protocol.

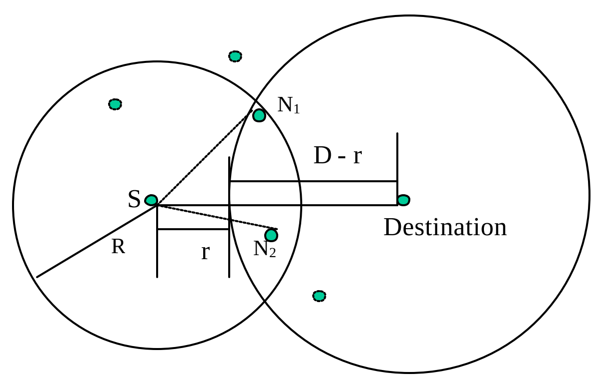

Figure 3.1 Packet forwarding in region “r”

At the first hop, the source S selects the best two nodes (N_1, N_2; ranked according to their calculated priority), which are “r” units closer to the destination, and transmits a copy of the data packet to both of them (Figure 1). All the intermediate nodes from now on forward the packet only along a single route to the destination. The destination node on receiving the duplicate second packet ignores it, if it has received the first packet already.

3.1.2 Neighbor Table management

Initially all nodes start with the same energy level and have a radio range R. At periodic time intervals, each node exchanges beacon messages (HELLO_PKT) with its neighboring nodes and constructs a neighbor table. The format of HELLO_PKT is as follows:

< *NodeId, xpos, ypos, e, m, timestamp*>

This HELLO_PKT includes the geographic location (*xpos, ypos*) of the node, the energy '*e*' of the node, the mobility factor '*m*' (based on the previous mobility pattern of the node), and the originating *timestamp* of the packet. By knowing the packet origination time, a receiving node can calculate the average delay experienced by a packet in reaching it.

delay = PKT_ORIG_TIME - PKT_RECV_TIME

Now delay   (Tp + T_d), where

T_p = propagation time across a link with no interfering traffic.

T_d = backoff time at the MAC Layer due to busy channel.

The incurred packet delay is thus an indication of the congestion around the neighboring node. Using this delay information, a node can factor node congestion into its routing algorithm and choose the next hop with the least delay for extremely time critical packets.

3.1.3 Packet Forwarding

Any packet originating from the source will be characterized by a packet ID, source ID, destination ID and the Time Left to deliver the packet. The source will forward the packet only if certain conditions are met:

T_p = minimum propagation delay across the link.

L = bandwidth of the node in Kbps.

T_L = Time Left to meet the deadline.

m = minimum hops to the destination.

R = Radio range of the node.

D = Distance from the current node to the final destination.

X, Y= co-ordinates of the next hop node.

X_1, Y_1= co-ordinates of the current node.

X_2, Y_2= co-ordinates of the destination node.

$$D = \quad (X_2-X_1)^2 + (Y_2-Y_1)^2,$$

$$m = D/R, \qquad Tp = S/L,$$

$$\text{If } (T_L < Tp*m)\text{, packet is dropped.} \tag{1}$$

This is because $\mathbf{T_p*m}$ represents the lower bound on the packet delivery time. If the time left to deliver the packet (T_L) is less than this lower bound, it is no use forwarding the packet any further, as it will not be able to reach the destination before its deadline. A check for this condition before forwarding ensures that no data packets will be unnecessarily forwarded, only to be dropped eventually at some point. This approach effectively saves energy and reduces traffic at the intermediate hops.

If the packet deadline meets the above criteria, then we scan the neighbor table and choose all such neighboring nodes, which are at least "r" distance units closer to the destination than the current node (Figure 1).

All such neighboring nodes will satisfy the following criteria:

$$r < \quad (X-X_1)^2+(Y-Y_1)^2 <= R \tag{2}$$

$$D-R < \quad (X-X_2)^2+(Y-Y_2)^2 <= D-r \tag{3}$$

The parameter "r" is itself dynamic and its value changes for different packets depending on the "urgency factor". In our scheme:

$r = R * K * (D/T_L)$, where D/T_L is known as the "urgency factor" and

K = normalization factor such that $0< r <=R$

The rationale behind this approach is to ensure fairness during real time packet forwarding and also achieve load balancing. A packet with higher value of (D/T_L) can be assumed to be more time critical as compared to one with a lower "urgency factor". For example, if a packet has its destination node at a distance D = 10 units away and the time left T_L = 5 units, its urgency factor will be 10/5 = 2, where as the packet with the same distance and having a T_L = 2 units, will have an urgency factor of 5. This means that, the second packet has to be delivered earlier than the first packet; and hence needs to be pushed closer to the destination than the first one. Consequently, only the most urgent packets are pushed to the boundary of the transmission range "R" and lesser urgent packets are not pushed to the fringes.

Once we have selected a set of neighboring nodes in the desired region, our next task is to pick the optimum node from the selected set for forwarding the data packet. To achieve this, the routing protocol computes priority factor of each of the node in the selected set. From the neighbor table it selects the nodes and calculates their priority in the following manner:

$$\text{Priority} = \ * (1/\text{delay}) + \ * (\text{energy}) + \ * (\text{mobility}) \qquad (4)$$

Where,

$= K * (D/T_L)$,

K = normalization factor,

$= (1- \ )$

= constant.

The philosophy behind the above equations is as follows. We try to assign maximum priority to the delay factor for packets with high emergency factor (D/T_L) and lesser priority to the energy factor. It makes perfect sense, because for time critical packets with aggressive deadlines, our major concern should be delivering the packets in time without having to worry about uniform energy utilization of neighboring nodes. However, a sensor network is limited by battery power and energy of nodes should not be overlooked altogether. Therefore, for packets with less aggressive deadlines (lower urgency factor), we assign more priority to the energy factor and try to locate nodes with high energy for forwarding data packets. The protocol thus factors both node congestion and node energy while routing real time traffic.

Once the nodes are prioritized based on the above equations, the session source, selects the best two nodes from the list and forwards a copy of data packet to each of them. This information duplication increases the reliability of data delivery. There is no data duplication at the intermediate nodes and packets are forwarded only to the node with the highest priority. Since the packet duplication is done only at the source node, minimal redundant data is injected into the network.

However, if there are no neighboring nodes in the desired region (Equations 3 and 4 are not satisfied), the window size "r" is decreased by a factor of 2 and nodes in the region "r/2" are searched for a possible forwarder as shown in Figure 3.2. The modified equations now become:

$$(r/2) < (X-X_1)^2+(Y-Y_1)^2 <= r$$

$$D-r < (X-X_2)^2+(Y-Y_2)^2 <= D-(r/2)$$

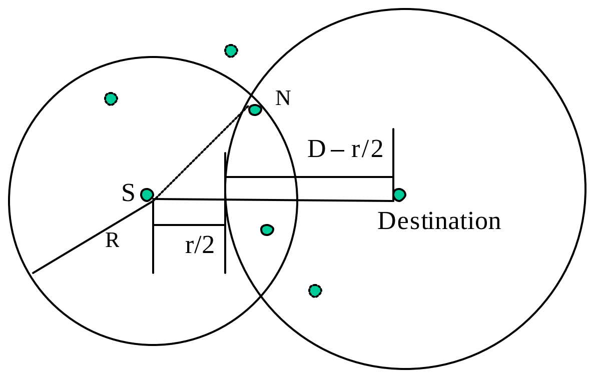

Figure 3.2 Packet Forwarding in region “r/2”

If intermediate nodes have both real time and non real time packets, then we maintain a buffer and real time packets are processed before the non real time packets. Amongst the real time packets, the packets are prioritized based on the emergency factor (D/T_L). This means that the most critical packets are sent first.

3.1.4 Prioritized MAC

This paper also introduces a prioritized MAC to reduce the delay in transmitting the packet at the MAC layer. Through simulation, we discovered that, the efficiency of a real time routing protocol is often limited by the delay at the MAC layer, which treats both real time packets and non real time packets alike. If a node has both kinds of packets (real time and non real time) to deliver, both these packets will be queued at the Interface Queue (IFQ). The MAC layer will then subsequently transmit each queued packet one at a time. A lesser critical data packet can therefore, block another packet with more aggressive deadline. It is therefore, extremely important to provide a differentiated service at the MAC layer as well, to reap the full benefits of an efficient real time routing protocol.

The IEEE 802.11 MAC DCF (Distributed Control Function) protocol is a carrier senses multiple access (CSMA) with collision avoidance (CA) protocol [17]. When operating in DCF mode, a node should sense the channel before transmitting any packet. If the channel is found to be idle for an interval greater than DIFS (Distributed Inter Frame Space), the node will reserve the channel by using RTS/CTS packet and then begin transmission eventually. However, if the channel is found to be busy, a backoff process is initiated. The value of the backoff timer is calculated as [18]

$$T = \text{Random}(0, CW) * T_{slot} \text{ where}$$

T_{slot} = slot time

CW = Contention Window

Once the backoff timer expires, the node senses the channel again. If the medium is found to be busy again, CW is doubled to decrease the probability of collision and backoff timer is recomputed. The node then, finally reinitiates the backoff process with the revised backoff value.

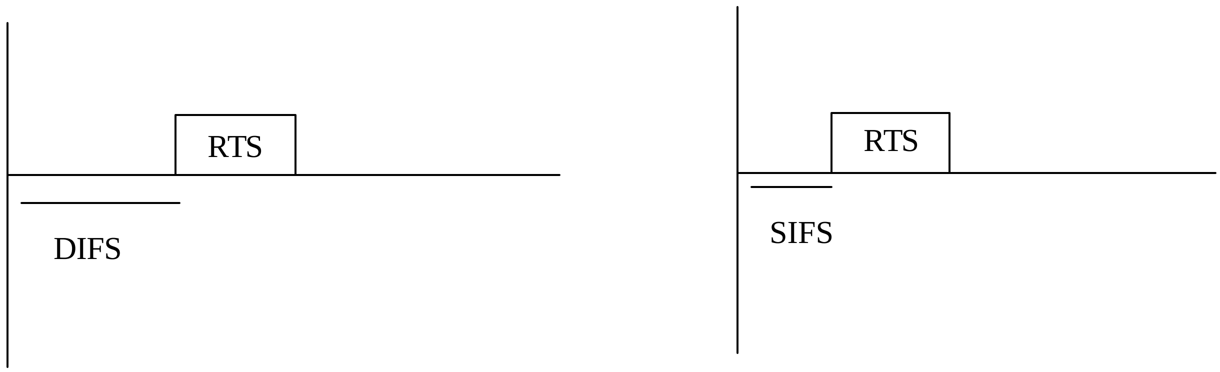

Figure 3.3 Reduced Waiting Time

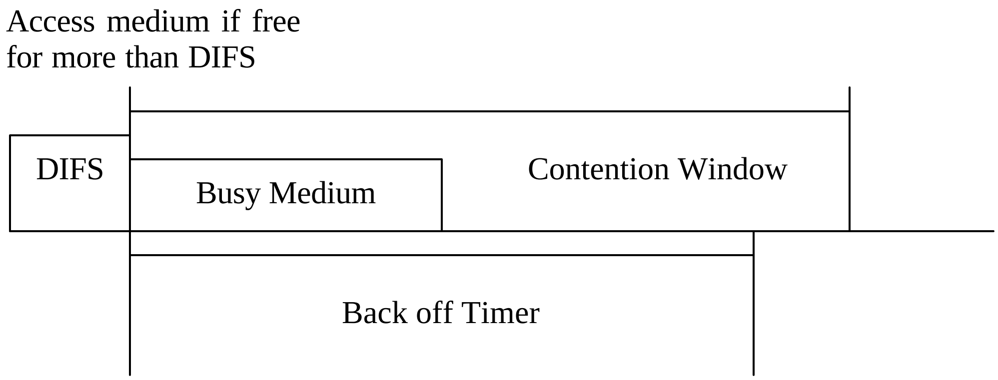

Figure 3.4 Contention Window

To avoid such latency for real time packets both the link layer and the 802.11 MAC layer have been modified to assign higher priority to real time packets. The link layer maintains two independent IFQs; IFQ_{REAL} for real time packets and $IFQ_{NON\text{-}REAL}$ for non-real time packets. Real time packets are queued in the IFQ_{REAL} according to their urgency factor (D/T_L) while the non real time packets are queued in the $IFQ_{NON\text{-}REAL}$ simply in the order of their arrival. The MAC layer assigns higher priority to the IFQ_{REAL} queue and processes it earlier. The MAC layer then follows a differentiated service model and handles the real time packets differently than non real time packets. If the packet to be delivered is a real time packet, then at the beginning of transmission, node waits for a smaller SIFS (Short Inter Frame Space) period (rather than DIFS) before transmitting a RTS packet (Figure 3.3). Also, contention window size is kept fixed for real time traffic and is not increased if the medium is found to be busy after the expiry of back off timer (Figure 3.4). This differentiated service model for real time traffic reduces fairness during channel contention and assigns higher priority to packets with aggressive deadlines. Real time packets with higher "urgency factor" have greater chances of acquiring the medium and can be delivered with minimum delay.

We also try to do away with post back off time for real time packets. Post back off is a time, where the node after a successful transfer of a packet waits for a random duration, before accessing the medium again. It is implemented to ensure fairness and provides other nodes a chance to get a fair share of the medium. However, in our scheme if the node has more than one packet in the real time queue, the post back off is turned off till all the real time packets are transmitted. This reduces the delay of the real time packets waiting in the queue to be processed. The post back off timer is only activated for non real time packets.

CHAPTER 4

Performance Evaluation

We simulated our QoS scheme using the ns-2 simulator. 200 nodes were distributed randomly over an area of 500x500. Initially all nodes started with a starting energy of 500. With every reception and transmission, the energy of the nodes decrease based on [19] (ratio of energy spent for packet reception to packet transmission was kept at 1.05:1.4). The range of the "R" was taken to be 100 and the link propagation time without congestion as 1 time unit.

4.1 Packet delay with different "r"

Figure 4.1, shows the average packet delay for varying traffic with different values of "r". We have compared the delays for fixed value of "r = 0.5R, 0.6R" to dynamic "r" as used in our QoS scheme. As evident, for low network traffic, the fixed scheme gives better results. This happens because when the network in not congested, the least delay is achieved by forwarding the packet to the node closest to the destination. As evident in the graph, for low traffic "r=0.6R" gives the best result. From the graph, we also observe that the packet delays in the region "r=0.5R" gives a better performance than "r=0.6R" for increased traffic. The reason behind is that, in "r=0.6R" region we are restricting the number of possibly less congested next hops as compared to the region with "r=0.5R". Also, with increasing "r" we move more towards geographic forwarding, hence more congestion. As the traffic increases, forwarding packets to a fixed region results in an increased congestion and more traffic delay in that area.

The best performance is achieved by our dynamic scheme. Our dynamic scheme selects different regions depending on the urgency factor of the data, thereby balancing the traffic in the network. This balancing helps to avoid the hot regions in the network and reduces the delay for the packets passing through the region. Thus for high traffic, our dynamic scheme gives much improved performance.

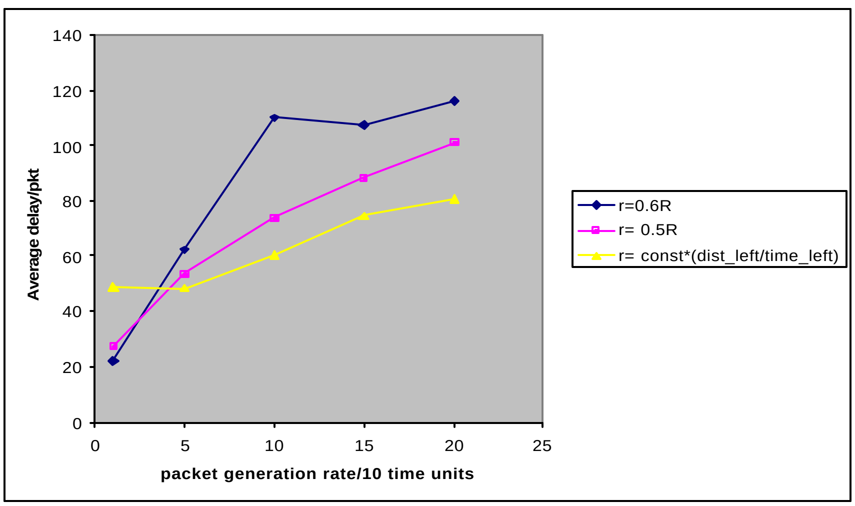

Figure 4.1. Average packet delay for different “r”

4.2 Average packet delay for different routing schemes

In Figure 4.2, we compare the average packet delay of our scheme to Geographic Forwarding (GF) and Geographic and Energy Aware Routing (GEAR). For low traffic, the delay experienced by all the schemes is almost comparable, because of decreased congestion, less collisions and low MAC layer traffic.

GF gives the worst performance with increased traffic. This is because in GF the same set of nodes (closest to the destination) get selected, as the traffic increases the congestion or delay around the nodes increases.

GEAR performs better than GF, because it doesn't select the same set of nodes on the basis of the geographic proximity. As the traffic increases and the network's energy decreases, GEAR chooses different set of nodes depending on geographic and energy factor. Since GEAR also gives high priority to the geographic location, it also results in building hot regions around the nodes. Our QoS scheme shows better results that both GF and GEAR, as the node selection is not restricted just to geographic proximity and energy, it also takes in to account the delay of the neighboring nodes. In nutshell, more we move away from geographic, lesser is the average packet delay experienced in the network.

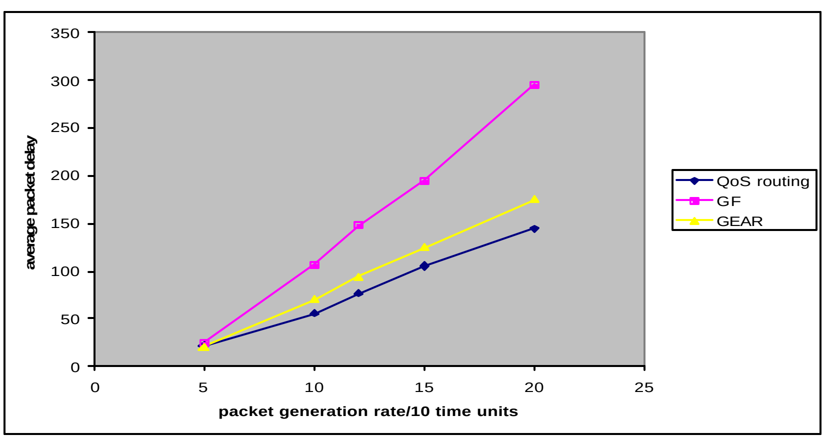

Figure 4.2. Average packet delay for different routing schemes

4.3 Network Lifetime

Figure 4.3, compares the network lifetime, which is extremely critical for a sensor network. Network lifetime is defined as the instant when the first node dies in the system. This parameter gives us a very clear picture of the future connectivity and life span of the Sensor Network. GF performs the worst under the circumstances because the same set of nodes is strained every time. With every packet reception and transmission, the fixed set of nodes (closest to the destination) gets depleted of the energy. Since, GEAR gives high priority to energy, therefore it gives the best performance. GEAR successfully routes the packets to the nodes with high energy, there by performing energy balance in the network. As evident by the graph, our QoS scheme is as efficient as GEAR and much better than GF, because of the priority given to energy while selecting the next hop. Overall our scheme gives much better results and accounts for the delay critical to real time applications. At the same time, it does not expend a lot of energy and increases the network lifetime.

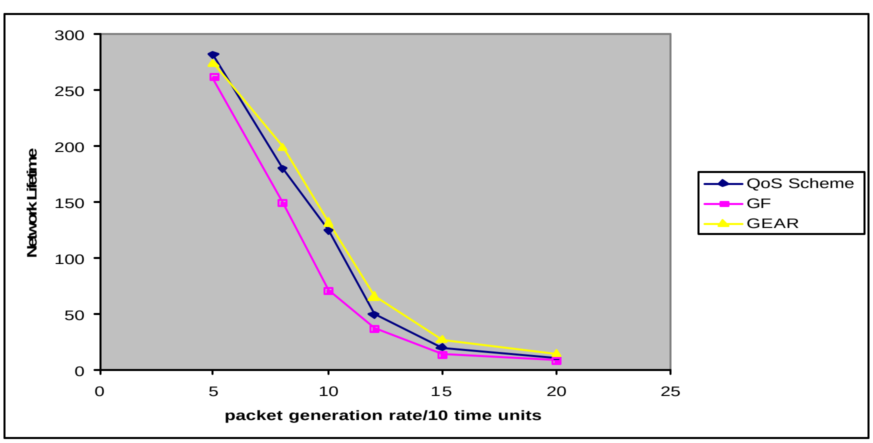

Figure 4.3. Network Lifetime for different routing schemes

4.4 Reliability

For real time packets it is very critical for the data to reach the destination within the deadline. Our strategy of packet duplication increases the probability of at least one of the packets reaching the destination before the deadline. Even if one of the routes encounter hot regions, it is possible that the same packet through another route experiences lesser-congested region. Thus sending the packet by two different routes increases the reliability. This is evident in Figure 4.4, where we have compared the packet delivery percentage with the deadlines. When the deadline is long enough, all three schemes achieve very high packet delivery percentage. As we make the deadlines more aggressive, we observe that the delivery percentage reduces drastically for GF and GEAR. However, our QoS scheme still maintains high delivery percentage. The only demerit of our scheme is that packet duplication consumes more energy as compared to a single path route. But in time critical applications, our foremost concern is achieving the deadline, rather than conserving the energy. Moreover, as shown in Figure 4.3, energy wise we still perform as close as GEAR.

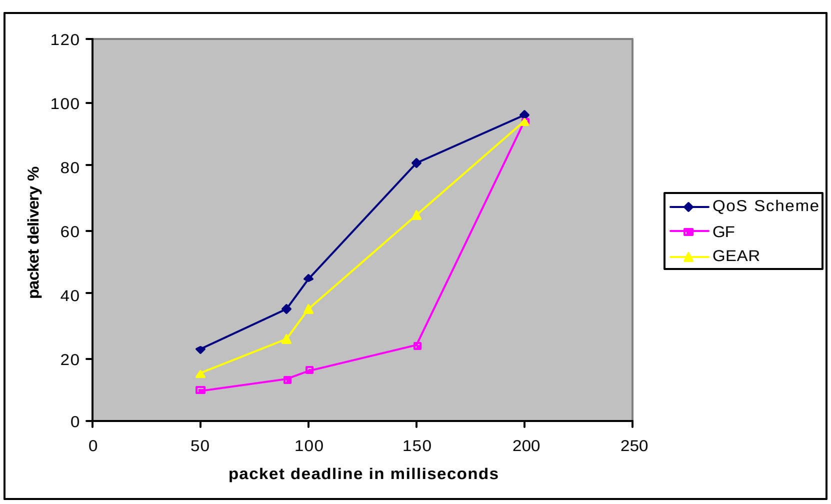

Figure 4.4. Packet delivery percentage for different deadlines

4.5 Packet delay with different , values

As mentioned in the scheme, the priority of the node in the region “r” as follows:

Priority = * (1/delay) + * (energy) + * (mobility) (4)

Where = K * (D/T_L),

K = normalization factor, = 1-

In Figure 4.5, we have compared the average packet delays for fixed values of the constants (=0.5, 0.7) to the dynamic values of , for a single source destination pair. From the graph it is evident that the best performance is achieved by dynamic values of , because of the uniform traffic distribution and reduced congestion. For =0.7 we get better results than =0.5, because of the increased priority given to the delay factor. Hence our scheme of selecting a dynamic value of the region “r” coupled with dynamic value of “ ”, gives the best performance for real time traffic.

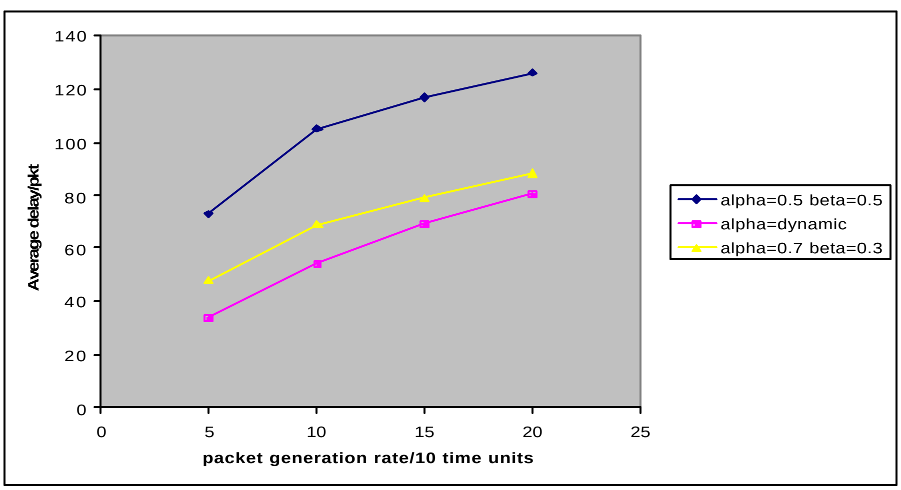

Figure 4.5. Packet delay for different values of alpha, beta

4.6 Packet delay for varying number of paths/routes

In our scheme we select two alternative routes to transmit the duplicated packets. As we observe in Figure 4.6, the selection of two paths gives the least delay as compared to sending the packet in three paths or in a single path. In single path routing, it is possible that in the intermediate hops the packet incurs high congestion due to the cross traffic. By choosing double path we increase the possibility of at least one of the routes incurring much lesser delay than the other one. If we further increase the number of paths to three, for low traffic the performance is similar to double path routing. However, as the traffic increases, due to more number of redundant traffic introduced by triple path routing, the congestion increases. This congestion due to high cross traffic significantly increases the delay and also depletes the nodes of the energy thereby reducing the network lifetime.

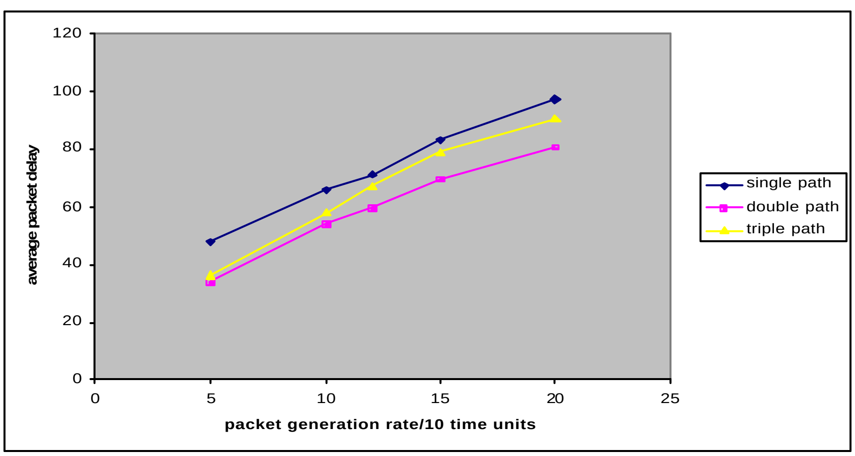

Figure 4.6 – Packet delay for varying number of paths

4.7 Number of intermediate hops

In Figure 4.7, we compare the average number of hops for the ideal case to our scheme. We have considered a scenario where the minimum number of hops (distance/radio range) between the source and destination pair is 4. Therefore in the ideal case, all packets should take 4 hops to reach the destination. However, due to the random topology, congestion and energy factors, number of actual hops taken is different in our scheme. It is evident from the figure that, as the traffic increases number of hops for the packets also increase due to the increased congestion around the fringe of the radio range.

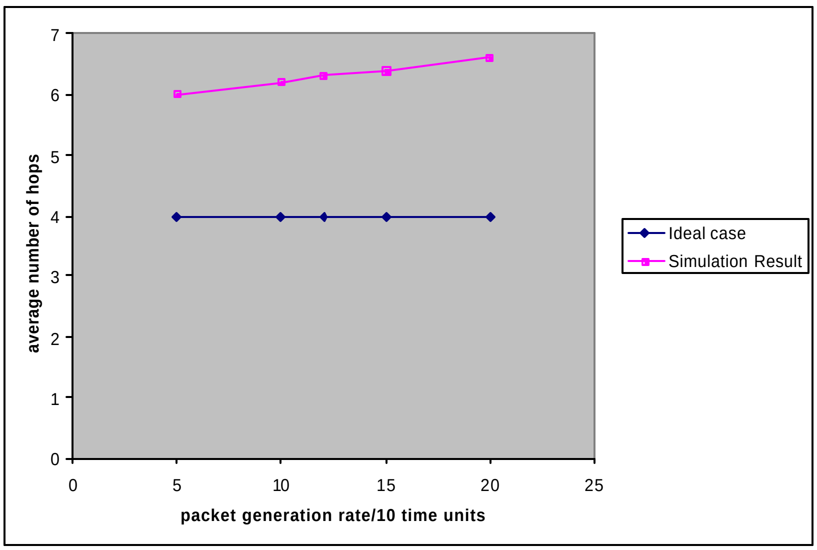

Figure 4.7 – Average number of hops for varying traffic

4.8 Routing Analysis

In this section, we perform the geographic analysis of our routing scheme for multihop packets with dynamic “r”. For simplicity we assume that there is no cross traffic and the nodes are randomly placed in the network according to Poisson distribution. The node density is assumed to be ?.

We know that the probability distribution function for Poisson arrival is as follows:

$$f_x(X) = \begin{cases} ?\ e^{-?x} & \text{if } x>=0, \\ 0 & \text{otherwise} \end{cases}$$

Probability of X>=a,

$$P(X >= a) = \int_a^{8} ?\ e^{-?x}\ dx = e^{-?a}$$

Therefore, the probability of pushing the packet with the remaining distance at least "*D-r*" (pushing the packet at max r distance units closer), is

$$P(X >= D\text{-}r) = e^{-?[pR2\ -\ A(r,R)]} \qquad (5)$$

The probability that the packet is in the region "X" such that *X < (D-r)* (finding the next hop in region "r") is equal to the probability of not finding the node in the region specified by *X >= D-r*.

$$P(X < D\text{-}r) = 1 - e^{-?[pR2\ -\ A(r,R)]} \qquad (6)$$

where,

r = constant * (Distance left/ Time Left)

The probability of finding the node in the region “r/2” is equal to the probability of not finding the node in the region “r” multiplied by the probability of finding the next hop in the region “r/2”, which is:

$$
\begin{aligned}
P(X< D-r/2) &= [1- P(X < D-r)]\, P[(X < D-r/2)] \\
&= e^{-\lambda[\pi R2\ -\ A(r)]}\, [1 - e^{-\lambda[\pi R2\ -\ A(r/2,r)]}] \qquad (7)
\end{aligned}
$$

In case of pure geographic forwarding, the probability of forwarding the packet to the next hop, assuming a dense network will always be 1. This results in a hot region around the route. However in our scheme since the probability of selecting a node in a region is not 1, we get an even distribution of load around the route.

where,

$$
A(r,R) = \int_{x1}^{R} 2\sqrt{R^2 - x^2}\, dx + \int_{r}^{x1} 2\sqrt{(D-r)^2 - (x-D)^2}\, dx \qquad (8)
$$

$$
x1 = (R^2+2rD-r^2)/(2D)
$$

It is the area of the two intersecting circles as shown in Figure 1. For simplicity we have assumed the forwarding node has coordinates at origin and destination at (D, 0).

Using the above equations and the average of the remaining distance as derived by Zorzi and Rao in [25], for the distances normalized by the radio range R,

$$
E[d] = D-1 + \int_{D}^{D} e^{-\lambda A(r,R)}\, dr \qquad (9)
$$

Where E[d]= average of the remaining distance when the packet is pushed by a distance d.

For the first hop,

$$E[d_1] = D_f - 1 + \int_{D_f-1}^{D_f} e^{-?A(r,R)}\, dr = (D_f - 1) + I_1$$

Where,

$r = constant*(D_f / T_f)$

D_f = Initial distance between the source and the destination

T_f = Initial Deadline

d_i = delay due to congestion at i^{th} hop

For the second hop,

$$E[d_2] = (E[d_1]-1) + \int_{E[d1]-1}^{E[d1} e^{-?A(r,R)}\, dr$$

$$= (E[d_1]-1) + I_2$$

Where,

$r = constant* (E[d_1])/(T_f - T_p - d_2)$

Assuming the packet reaches the destination at K^{th} hop

$$E[d_k] = (E[d_{k-1}]-1) + \int_{E[dk-1]-1}^{E[dk-1]} e^{-?A(r,R)}\, dr \qquad (10)$$

Since the packet reaches the destination at K^{th} hop, the average distance left after K hops will be equal to 0.

$$E[d_k] = 0$$

From Equation (10), we see that there is an inductive relation between the average number of hops for a packet. Hence by solving for the above equation for k, we can get the average number of hops for packets with dynamic “r”.

CHAPTER 5

Conclusions and Future Work

5.1 Conclusions

Our routing protocol is stateless in nature, energy aware and deadline driven. From the results, it is evident that our scheme gives much improved performance for high traffic real time packets as compared to other geographic routing schemes. By using dynamic value of "r", we obtain less packet delays and maintain traffic balance in the network. The energy metric ensures uniform energy depletion, thus increasing the network lifetime. We also reduce the MAC delays in the packets by using a prioritized MAC layer. This layer waits for a reduced time SIFS before sending the RTS, doesn't increase the contention window and eliminates post back off time for real time packets. Therefore, both at the routing and the MAC layer, we successfully reduce the latency and achieve higher packet delivery percentage.

5.2 Future Work

As future work, this work could be extended to achieve performance optimization for multicast routing. The other possible extension could be developing a dynamic equation for the mobility pattern and including probabilistic routing as one of the parameters for selecting the next hop node. We can also study the performance of our routing schemes for regular topologies, like grid etc.

References

[1] Dharma P. Agrawal and Q. Zeng, *Introduction to Wireless and Mobile Systems*, Brooks/ColePublishing, 436 pages, ISBN 0534-40851-6, 2003.

[2] J. Jubin and J.D Tornow, "The DARPA Packet Radio Project ," *Proceedings of the IEEE*, 1987.

[3] J. M. Kahn, R. H. Tornow Katz, and K. S. J Pister, "Next Century Challenges: Mobile Networking for Smart Dust," *MOBICOM'99*, Seattle, WA, 1999.

[4] The WINS Project, http://www.janet.ucla.edu/WINS.

[5] C. E Perkins and E. M. Royer, "Ad-Hoc On Demand Distance Vector Routing," *2nd IEEE Workshop on Mobile Computing Systems and Applications(WMCSA'99)*, New Orleans, Louisiana, February 1999.

[6]David B. Johnson and David A. Maltz, "Dynamic Source Routing in Ad-Hoc Wireless Networks," *Mobile Computing*, *Chapter 5,* pp. 153-181, Kluwer Academic Publishers, 1996.

[7]Brad Karp and H. T. Kung, "GPSR:Greedy Perimeter Stateless Routing for Wireless Networks," Proceedings AC Mobicom, Boston, MA, 2000.

[8] Yan Yu, Ramesh Govindan, and Deborah Estrin, "Geographical and Energy Aware Routing: A Recursive Data Dissemination Protocol for Wireless Sensor Networks," *UCLA Computer Science Department Technical Report UCLA/CSD-TR-01-0023*, May 2001.

[9]W. Zhao, J. A. Stankovic, and K. Ramamritham, "A Window Protocol for Transmission for Time-Constrained Messages," *IEEE Transactions on Computers*, Vol. 39, No. 9, pp.1186-1203, Sept. 1990.

[10] D. Kandlur, K. G. Shin, and D. Ferrari,"Real-time Communication in Multi-hop Networks," *IEEE Transactions on Parallel and Distributed Systems*, October 1994, pp.1044-1056.

[11] Sund-Ju Lee and Mario Gerla, "Split Multipath Routing with Maximally Disjoint Paths in Ad Hoc Networks," *Technical Report in University of California*, 2000.

[12] Urban Bilstrup and Per-Arne Wiberg, "Routing Protocol for Wireless Real- Time Multihop Networks," www.hh.se/ide/publications/publ.files99/ bilstrupWIP99.pdf

[13] Tiam He, John A. Stankovic, Chenyang Lu, and Tarek Abdelzaher, "SPEED: A Real Time Routing Protocol For Sensor Networks," *University of Virginia Tech. Report CS-2002-09*, March 2002.

[14] A. Shaikh, J. Rexford and K. Shin,"Dynamics of Quality-of-service routing with inaccurate link-state information," *Technical Report CSE-TR-350-97*, Dept. of Electrical Engineering and Computer Science, the University of Michigan, Ann Arbor, MI, Nov. 1997.

[15] Q. Ma and P. Steenkiste, "On path selection for traffic with bandwidth guarantees," *Proceedings of the IEEE International Conference on Network Protocols*, Oct. 1997.

[16] O. Kasten. Energy Comsumption. ETH-Zurich, Swiss Federal Institute of Technology. Available at http:/www.inf.ethz.ch/~kasten/research/bathtub/energy_consumption.html%, 2001.

[17] Jang-Ping Sheu, Chi-Hsun Liu, Shih-Lin Wu, and Yu-Chee Tseng, "A Priority MAC Protocol to Support Real-time Traffic in Ad Hoc Networks," *European Wireless*, 2002.

[18] *Wireless LAN Medium Access Control(MAC) and Physical Layer (PHY) Specifications*, IEEE Standard 802.11, June 1999.

[19] O. Kasten. Energy Comsumption. ETH-Zurich, Swiss Federal Institute of Technology. Available at http:/www.inf.ethz.ch/~kasten/research/bathtub/energy_consumption.html%, 2001.

[20] K. Sohrabi, J. Gao, W. V. Ailawadhi, G. J. Pottie, "Protocols for self-organization of a wirelesssensor network," *IEEE Personal Communications*, vol. 7, no. 5, pp. 16-27, Oct. 2000.

[21] Heinzelman W., Chandrakasan A., and Balakrishnan H.:, "Energy-Efficient Communication Protocols forWireless Microsensor Networks," In *Proceedings Hawaaian International Conference on Systems Science*, January 2000.

[22] C. Intanagonwiwat, R. Govindan, and D. Estrin, Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks. In *Proceedings of the 6th Annual ACM/IEEE International Conference on Mobile Computing and networking (MOBICOM)*, pp. 56–67, August 2000.

[23] Braginsky D. and Estrin D., "Rumor Routing Algorithm For Sensor Networks," *In Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications*, pp. 12-21, September 2002.

[24] Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., and Cayirci, E., "A Survey on Sensor Networks,"*IEEE Communications Magazine*, Vol. 40, No. 8, pp. 102-114, August 2002.

[25] M. Zorzi and R. R. Rao, ``Geographic Random Forwarding (GeRaF) for ad hoc and sensor networks: multihop performance," *IEEE Trans. on Mobile Computing*, Vol. 2, Oct.-Dec. 2003.