

QoS and energy aware routing for real-time traffic in wireless sensor networks

Abinash Mahapatra, Kumar Anand, Dharma P. Agrawal*

OBR Center for Distributed and Mobile Computing, University of Cincinnati, Cincinnati, OH 45221-0030, USA

Available online 12 February 2005

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 improved QoS (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 paper, we propose an energy aware dual-path routing scheme for real-time traffic, which balances node energy utilization to increase the network lifetime, takes network congestion into account to reduce the routing delay across the network 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 claims are well supported by simulation results.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Real-Time communication; QoS routing; Sensor networks

1. Introduction

The paradigm of ad hoc network dates back to the 1970s, 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, small sized 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 capabilities. 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 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 forest fire detection, habitat monitoring, detecting radiation leakage, impurity level in sea 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

* Corresponding author. Tel.: +1 513 556 4756; fax: +1 513 556 7326.

E-mail addresses: mahapaa@ececs.uc.edu (A. Mahapatra), anandk@ececs.uc.edu (K. Anand), dpa@ececs.uc.edu (D.P. Agrawal).

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, reducing the delay by limiting the number of hops. Other factors 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.

To overcome the restrictions imposed by aforementioned factors, we have to reduce the number of hops a packet has to take to reach the destination by utilizing the GPS information and the radio range of the node. However, simple geographic forwarding can cause congestion at specific nodes, leading to significant delays. Routing should thus, also factor node congestion 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 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 the 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 paper also introduces an adaptive prioritized MAC, which assigns higher priority to real-time packets and reduces the MAC delay for time critical data.

2. Related work

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, 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. The standard 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 does not 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 [20], 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 into 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. However, flooding has extremely poor forwarding efficiency and results in lot of redundant transmissions, increased energy consumption, and hence decreased network lifetime. A better approach is suggested in [11], where a set of disjoint paths is maintained from source to destination over which the data is transmitted. This scheme also results substantial energy overhead, suffers from cache pollution and does not consider the time constraint nature of the packets. Certain schemes like [12] require both GPS and GIS capability to find out the best route. 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 to balance the node energy utilization. Furthermore, the region it chooses for forwarding and the priority selection 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.

3. Proposed protocol

3.1. 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 help 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.

3.2. 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 backoff 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. Fig. 1 illustrates the working of the routing protocol.

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 (Fig. 1). All the

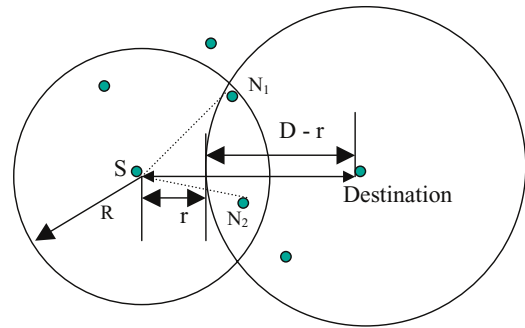


Fig. 1. First hop nodes selection.

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.

3.3. 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 , timestamp>

This HELLO_PKT includes the geographic location (xpos, ypos) of the node, the energy e 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

$$\text{delay} = \text{PKT_ORIG_TIME} - \text{PKT_RECV_TIME}$$

Now $\text{delay} \propto (T_p + T_d)$, 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.

3.4. 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
- S size of the packet
- 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

$$D = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}, \quad m = D/R, \quad (1)$$

$$T_p = S/L, \quad \text{If } (T_L < T_p m), \text{ packet is dropped.} \quad (1)$$

This is because $T_p m$ 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 (Fig. 1). All such neighboring nodes will satisfy the following criteria:

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

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

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

$$r = RK(D/T_L),$$

where D/T_L is the urgency factor and K is the normalization factor such that $0 < r \leq R$.

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 $10/5=2$, whereas 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:

$$\text{Priority} = \alpha(1/\text{delay}) + \beta(\text{energy}) \quad (4)$$

where

$$\alpha = K(D/T_L),$$

$$K = \text{normalization factor}$$

$$\beta = (1 - \alpha).$$

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 [16]. 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 (Eqs. (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 Fig. 2.

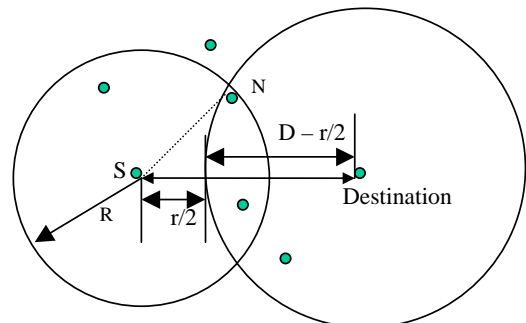


Fig. 2. Decreasing window size to half.

The modified equations now become:

$$(r/2) < \sqrt{(X - X_1)^2 + (Y - Y_1)^2} \Leftarrow r$$

$$D - r < \sqrt{(X - X_2)^2 + (Y - Y_2)^2} \Leftarrow D - (r/2)$$

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.

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] (Fig. 3). 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. However, if the channel is found to be busy, a backoff process is initiated. The value of the backoff timer is calculated as [18]

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

where

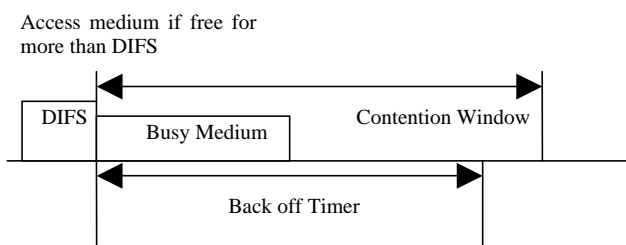


Fig. 3. 802.11 MAC layer.

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.

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-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-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 thus, 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, a node waits only for a smaller SIFS (Short Inter Frame Space) period (rather than DIFS) before transmitting a RTS packet. 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 backoff timer. This simple 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 therefore, be delivered with minimum delay.

The prioritized MAC layer also eliminates post backoff time for real-time packets. Post backoff implies that a node, after a successful transmission of a packet will wait for a random duration, before accessing the medium again. It is implemented to ensure fairness and provides other nodes a fair chance to access the medium. However, in our scheme if the node has more than one packet in the real-time queue, the post backoff 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 backoff timer is only activated for non-real-time packets.

5. Performance evaluation

We have implemented our QoS routing protocol in NS (version 2.26). The simulation environment models a sensor network of 100 nodes distributed randomly over an area of 500×500 . At the beginning of simulation, all nodes started with a starting energy of 500 units. With every reception and transmission, the energy of the nodes decreases based on [19] (ratio of energy spent for packet reception to packet transmission was kept at 1.05:1.4). The network stack of

each mobile node consists of link layer, an ARP module, modified interface priority queue, modified IEEE 802.11 MAC layer with 100 m transmission range, and a network interface. The link propagation time without congestion is assumed to be 1 time unit.

For the simulations that follow, we have considered CBR traffic (having different deadlines) with payload size set to 512 bytes. Data packets are generated at the source at a rate of 0.5, 1, 1.5, 2 packets/s. Each simulation runs for 500 s and there is no network partition during the course of simulation. We have compared the performance of our routing scheme with GEAR and Geographic Routing (GR).

5.1. Packet delay with different r

Fig. 4 shows the impact on average packet delay for different values of r . We have compared the packet delays for fixed value of $r=0.5R$, $0.6R$ to dynamic r as used in our QoS scheme. As expected, the fixed scheme gives better results for low network traffic. This happens because, when the network is not congested, least delay is achieved by forwarding the packet to the node closest to the destination. However, as the traffic increases, forwarding packets to a fixed region results in increased congestion and more traffic delay in that area. Our dynamic scheme selects different regions depending on the urgency factor of the data packet, thereby balancing the traffic in the network. This balancing helps to avoid hot regions in the network and reduces the delay for packets passing through the region. Thus for high network traffic, our scheme gives much improved performance.

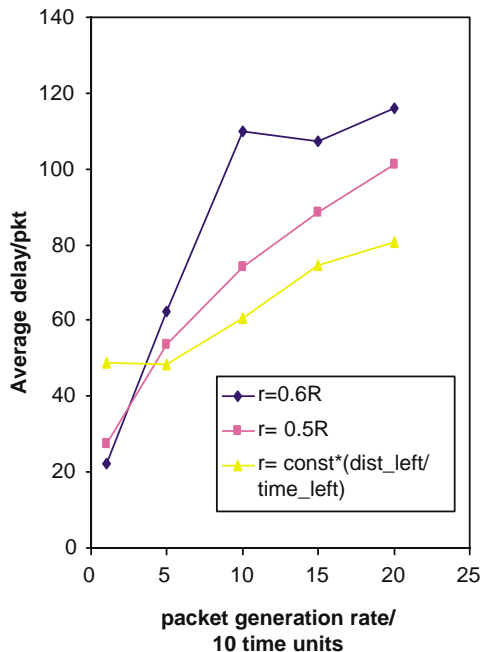


Fig. 4. Average packet delay for different r .

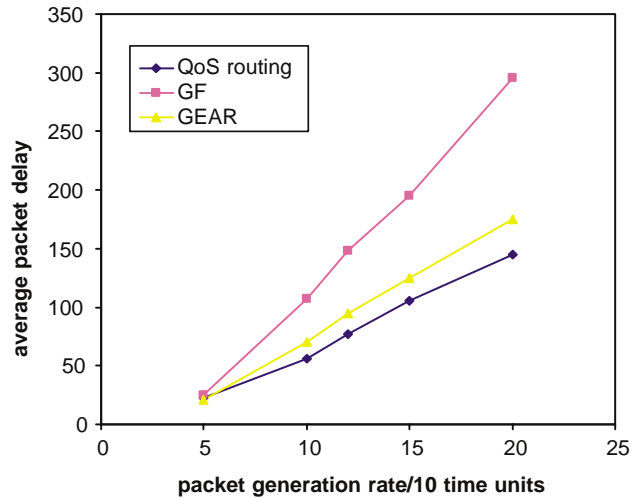


Fig. 5. Average packet delay for different routing schemes.

5.2. Average packet delay

In Fig. 5, 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 comparable. Since in GF the same set of nodes (closest to the destination) gets selected, as the traffic increases the congestion around the forwarding nodes also increases. GEAR, however, performs better than GF, because it does not select the same set of forwarding nodes just 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 both geographic and energy factor. Our QoS scheme shows better results than both GF and GEAR as the node selection is not restricted simply on geographic proximity and energy but also takes into account the delay at the neighboring nodes.

5.3. Network lifetime

Fig. 6 compares the network lifetime for different schemes, which is extremely critical for a sensor network.

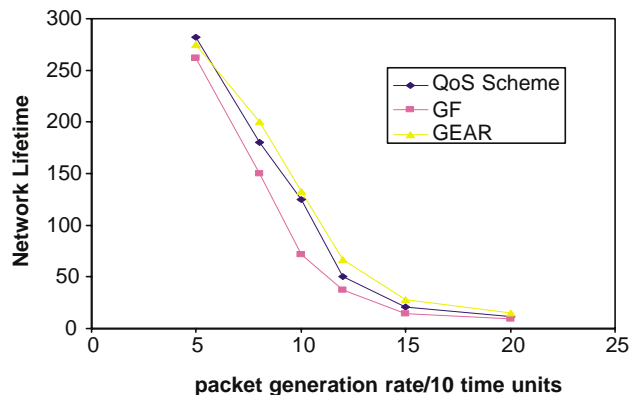


Fig. 6. Network lifetime for different routing schemes.

GF performs the worst under the circumstances because the same set of nodes is used for forwarding data packets every time. GEAR gives high priority to energy; therefore it gives the best performance. As evident by the graph, our QoS scheme is as efficient as GEAR and much better than GF. The proposed QoS scheme is able to balance node energy utilization like GEAR and also accounts for the delay critical to real-time applications.

5.4. 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. Thus sending the packet by two different routes increases the reliability. This is evident in Fig. 7, 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. Proposed QoS routing scheme has higher delivery ratio than other schemes for packets with aggressive deadlines.

5.5. Packet delay with varying α, β

As mentioned in the scheme, the priority of the node in the region r is determined from Eq. (4). In Fig. 8, we have compared the average packet delays for fixed values of the constants ($\alpha=0.5, 0.7$) to dynamic values of α , for a single source destination pair. From the graph, it is evident that the best performance is achieved by dynamic values of α , because of the uniform traffic distribution and reduced

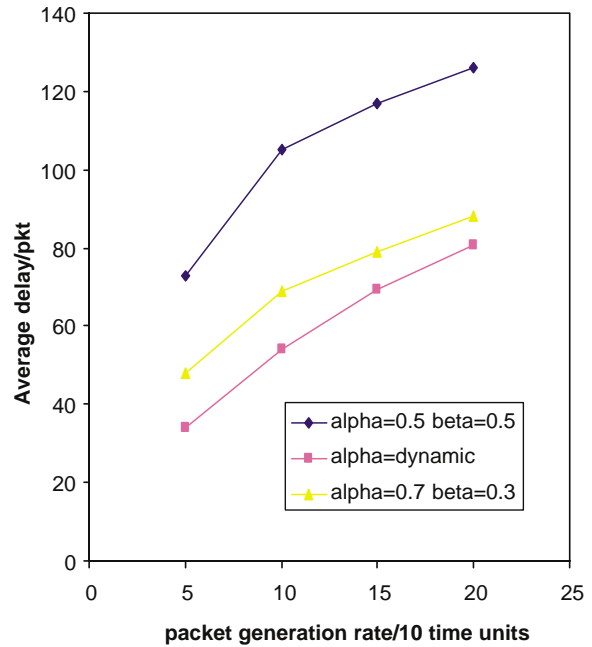


Fig. 8. Packet delay with varying α, β .

congestion. For $\alpha=0.7$ we get better results than $\alpha=0.5$, 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 α , gives better performance for real-time traffic.

5.6. 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 Fig. 9, 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

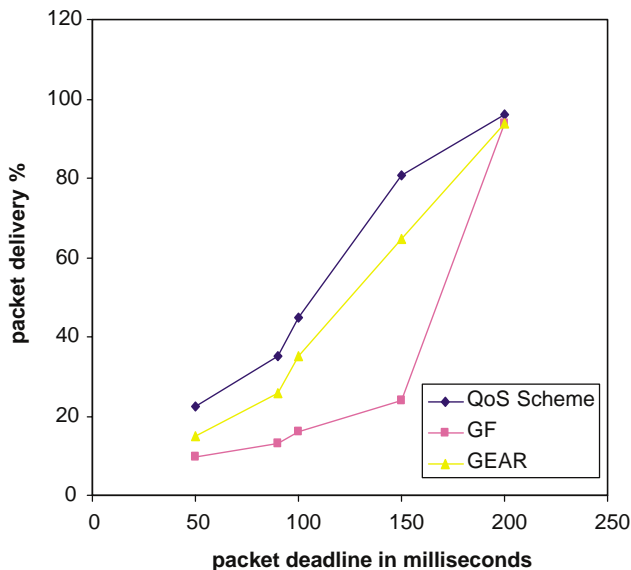


Fig. 7. Packet delivery ratio for different deadlines.

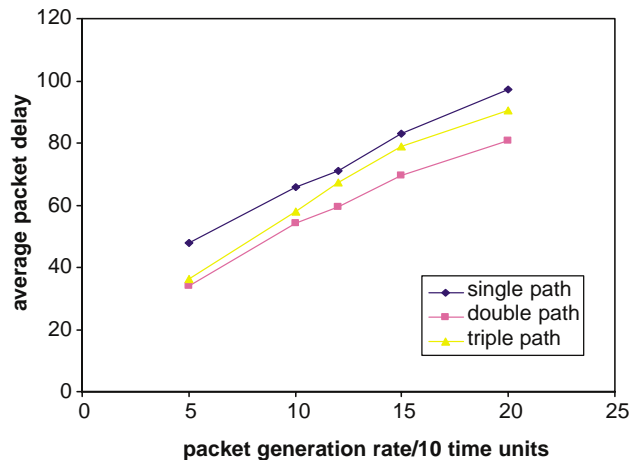


Fig. 9. Packet delay for varying number of paths.

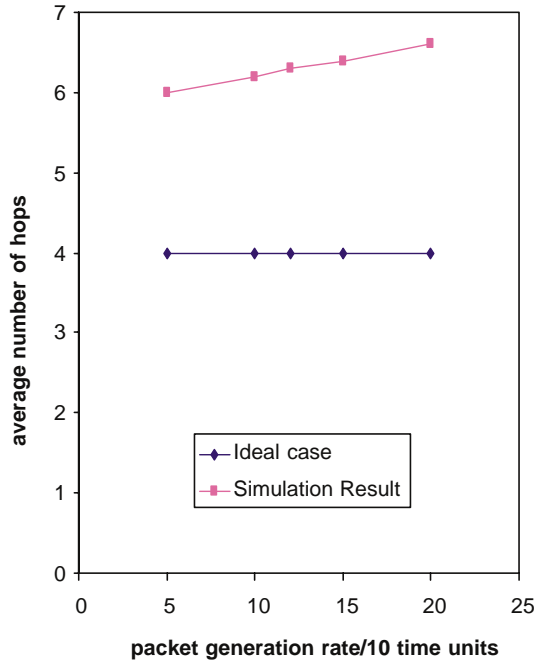


Fig. 10. Average number of hops for varying traffic.

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.

5.7. Number of intermediate hops

In Fig. 10, 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 four 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.

6. 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 ρ .

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

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

Probability of $X \geq a$

$$P(X \geq a) = \int_{a\lambda}^{\infty} e^{-\lambda x} dx = e^{-\lambda a}$$

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

$$P(X \geq D - r) = e^{-\rho[\pi R^2 - A(r,R)]} \tag{5}$$

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 $X \geq D - r$

$$P(X < D - r) = 1 - e^{-\rho[\pi R^2 - A(r,R)]} \tag{6}$$

where $r = \text{constant} \times (\text{Distance left} / \text{Time Left})$.

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:

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

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

$$\begin{aligned} A(r,R) &= \int_{x_1}^R 2\sqrt{R^2 + x^2} dx + \int_r^{x_1} 2\sqrt{(D-r)^2 - (x-D)^2} dx, \\ x_1 &= (R^2 + 2rD - r^2)/(2D) \end{aligned} \tag{8}$$

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

Using the above equations and using the average of the remaining distance as derived by Zorzi and Rao in [20], for

the distances normalized by the radio range R

$$E[\delta] = D - 1 + \int_{D-1}^D e^{-\rho A(r,R)} dr \quad (9)$$

where $E[\delta]$ = average of the remaining distance when the packet is pushed by a distance δ .

For the first hop

$$E[\delta_1] = D_f - 1 + \int_{D_f-1}^{D_f} e^{-\rho A(r,R)} dr = (D_f - 1) + I_1$$

where

r constant $\times (D_f/T_f)$
 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

$$E[\delta_2] = (E[\delta_1] - 1) + \int_{E[\delta_1]-1}^{E[\delta_1]} e^{-\rho A(r,R)} dr \\ = (E[\delta_1] - 1) + I_2$$

where $r = \text{constant}(E[\delta_1]) / (T_f - T_p - d_2)$

Assuming the packet reaches the destination at K th hop

$$E[\delta_k] = (E[\delta_{k-1}] - 1) + \int_{E[\delta_{k-1}]-1}^{E[\delta_{k-1}]} e^{-\rho A(r,R)} dr \quad (10)$$

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

$$E[\delta_k] = 0$$

From Eq. (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 .

7. Conclusions

The proposed routing protocol is stateless, 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 are able to achieve smaller packet delays, are able to maintain traffic balance and reduce node congestion at the forwarding nodes in the network. The energy metric ensures uniform energy depletion and thus increases the network lifetime. By employing a differentiated service model at the MAC layer, we are also successful in further reducing packet delay at lower layers. The MAC layer assigns higher priority to real-time packets and handles them differently from non-real-time packets. The use of a prioritized adaptive MAC scheme enables us to reap the full benefits of the upper routing layer

and prevents a lesser time critical packet from blocking a packet with more aggressive deadline. Therefore, both at the routing and the MAC layer, we successfully reduce the latency and are able to achieve higher packet delivery (in time) ratio.

References

- [1] P. Dharma, Q. Agrawal, Q. Zeng, Introduction to Wireless and Mobile Systems, Brooks/Cole, Belmont, CA, 2003. 436 pp.
- [2] J. Jubin, J. Tornow, The DARPA Packet Radio Project, Proceedings of the IEEE 1987;.
- [3] J.M. Kahn, R.H. Tornow Katz, 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, E.M. Royer, Ad-hoc on demand distance vector routing, Second IEEE Workshop on Mobile Computing Systems and Applications(WMCSA'99), New Orleans, LA, February 1999.
- [6] D.B. Johnson, D.A. Maltz, Dynamic source routing in ad-hoc wireless networks, Mobile Computing, Kluwer Academic Publishers, Dordrecht, 1996. pp. 153–181 (Chapter 5).
- [7] B. Karp, H.T. Kung, GPSR: greedy perimeter stateless routing for wireless networks, Proceedings AC Mobicom, Boston, MA, 2000.
- [8] Y. Yu, R. Govindan, D. 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, K. Ramamritham, A window protocol for transmission for time-constrained messages, IEEE Transactions on Computers 39 (9) (1990) 1186–1203.
- [10] D. Kandlur, K.G. Shin, D. Ferrari, Real-time communication in multi-hop networks, IEEE Transactions on Parallel and Distributed Systems 1994; 1044–1056.
- [11] S.-J. Lee, M. Gerla, Split multipath routing with maximally disjoint paths in ad hoc networks, Technical Report in University of California, 2000.
- [12] U. Bilstrup, P.-A. Wiberg, Routing Protocol for Wireless Real- Time Multihop Networks, www.hh.se/ide/publications/publ.files99/bilstrupWIP99.pdf
- [13] T. He, J.A. Stankovic, C. Lu, T. 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, 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, P. Steenkiste, On path selection for traffic with bandwidth guarantees, Proceedings of the IEEE International Conference on Network Protocols, October 1997.
- [16] O. Kasten, Energy Consumption, ETH-Zurich, Swiss Federal Institute of Technology. Available at http://www.inf.ethz.ch/~kasten/research/bathtub/energy_consumption.html, 2001.
- [17] J.-P. Sheu, C.-H. Liu, S.-L. Wu, Y.-C. 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 Consumption, ETH-Zurich, Swiss Federal Institute of Technology. Available at http://www.inf.ethz.ch/~kasten/research/bathtub/energy_consumption.html, 2001.
- [20] M. Zorzi, R.R. Rao, Geographic random forwarding (GeRaF) for ad hoc and sensor networks: multihop performance, IEEE Transactions on Mobile Computing 2 (2003).