



QoS routing based on multi-class nodes for mobile ad hoc networks

Xiaojiang Du *

Department of Engineering, Indiana University–Purdue University, 2101 E. Coliseum Blvd., Fort Wayne, IN 46805, USA

Available online 17 April 2004

Abstract

In this paper, we present a new quality of service (QoS) routing protocol for mobile ad hoc networks (MANETs). Most of the existing routing protocols assume homogeneous nodes in MANETs, i.e., all nodes have the same communication capabilities and characteristics. However, in many ad hoc networks, nodes are not the same. Some nodes have longer transmission range, larger transmission bandwidth, and are more reliable and robust than other nodes. We take advantage of the non-homogeneous property to design more efficient QoS routing protocol. And node location information is used to aid routing. We also develop a new algorithm to calculate end-to-end bandwidth for a given path. Our QoS routing protocol contains end-to-end bandwidth calculation and bandwidth reservation. QoS route is discovered and setup only when it is needed. Extensive simulation studies demonstrate the good performance of the QoS routing protocol.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Mobile ad hoc networks; Quality of service; Routing; Location-aided routing; Backbone nodes

1. Introduction

Mobile ad hoc networks form a class of dynamic multi-hop network consisting of a set of mobile nodes that intercommunicate on shared wireless channels. While the infrastructured cellular system is a traditional model for a mobile wireless network, MANETs are self-organizing and self-configuring multi-hop wireless networks, where the network structure changes dynamically due to node mobility.

QoS routing is important for a mobile network to interconnect wired networks with QoS support (e.g., Internet). The QoS routing protocol is also needed in a stand-alone multi-hop mobile network for real-time applications (like voice, video, etc.).

QoS routing requires not only to find a route from a source to a destination, but a route that satisfies the end to end QoS requirement, often given in terms of bandwidth or delay. QoS routing in wired networks has been well studied. A novel and interesting direction in network QoS research is the cognitive packet networks (CPN) [4–6,11,19], proposed by Gelenbe. They proposed “smart packets”, which use a reinforcement learning algorithm (based on random neural network) to achieve intelligent peer-to-peer routing, and thus

* Tel.: +1-260-481-6389; fax: +1-260-481-6281.

E-mail address: du@engr.ipfw.edu (X. Du).

provide effective QoS routing. In [20,21], a concept called self-awareness is introduced, which is the ability of a network to observe its own behavior using internal probing and measurement mechanisms. Self-awareness can effectively improve QoS performance [21]. Gelenbe also proposed QoS routing with genetic algorithms [8], using loss and delay as QoS goal [12], etc. Deploying artificial intelligence techniques in network QoS research is an interesting and promising research topic. Other interesting research in wired network QoS include distributed routing [25,26], where the path is computed by a distributed computation during which control messages are exchanged among the nodes, and the state information kept at each node is collectively used in order to find a path. In [22], Xue proposed an efficient approximation algorithm for minimum-cost QoS multicast routing problems and an efficient heuristic algorithm for unicast routing problems in communication networks.

Quality of service is more difficult to guarantee in ad hoc networks than in most other type of networks, because the network topology changes as the nodes move and network state information is generally imprecise. This requires extensive collaboration between the nodes, both to establish the route and to secure the resources necessary to provide the QoS. In recent years, however, QoS in ad hoc networks as a research topic has started to receive attention from a growing number of researchers [1–3,23,24,27,29,30], and major advances are expected in the next few years.

QoS needs a set of service requirements to be met by the network while transporting a packet stream from source to destination. The ability to provide QoS is heavily dependent on how well the resources are managed at the MAC layer. Among the QoS routing protocols proposed so far, some use generic QoS measures and are not tuned to a particular MAC layer [8,13]. Some use CDMA to eliminate the interference between different transmissions [2,3,14,15]. Different MAC layers have different requirements for successful transmissions, and a QoS routing protocol developed for one type of MAC layer does not generalize to others easily. In this paper, we develop a QoS routing protocol for ad hoc networks using TDMA. The goal is to establish bandwidth guaranteed QoS routes in

MANETs. The protocol is an on-demand routing protocol, and builds QoS routes only as needed. A flow specifies its QoS requirement as the number of transmission time slots it needs from a source to a destination. For each flow, the QoS routing protocol will find both the route and the transmission time slots for each node on the route.

In MANETs, the QoS issues include delay, delay jitter, bandwidth, probability of packet loss, and so on. In this paper, we mainly concern about bandwidth. The main contributions of this paper are: (1) We develop a new algorithm that can compute the maximum available bandwidth for a given path. (2) We propose a new QoS routing protocol that calculates bandwidth and reserves slots for MANET. This is the first QoS routing protocol (to our best knowledge) that exploits multi-class nodes with different communication capabilities. (3) We prove that the probability of having one backbone node in each grid is large by VC dimension theorem. The rest of the paper is organized as follows. In Section 2, we review the related work on routing protocols considering non-homogeneous nodes in MANETs, and routing with location information. In Section 3, we present a new algorithm that calculates the available maximum bandwidth in a given path. We describe our QoS routing protocol in Section 4 and a routing example is also given. In Section 5, we discuss the simulation experiments performed with the QoS routing protocol and we compared the performance with best effort AODV routing protocol. In Section 6, we prove that the probability of having a backbone node in each grid is very large. And we conclude the paper in Section 7.

2. Related works

Usually nodes in MANETs are considered homogeneous, i.e., all nodes are the same. They have the same transmission range, the same bandwidth, the same reliability and security. But in many real mobile ad hoc networks, nodes have different communication capabilities. For example, in a battlefield network, there are soldiers carrying portable wireless devices, and there are vehicles and tanks carrying more powerful and reliable

communication devices. They have different communication characteristics in terms of transmission power, bandwidth, processing capability, etc. So it would be more realistic to model these network elements as different type of nodes. Also there are several advantages that can be exploited to design better QoS routing protocol when nodes in MANETs are modeled as different type of nodes.

There are some previous works considered non-homogeneous MANETs. In [7], Xu presents a routing algorithm with mobile backbones. They propose to build a physically hierarchical ad hoc network to achieve good scalability. They consider non-homogeneous ad hoc network, e.g., nodes have different radio capabilities. A hierarchical large-scale ad hoc network is built using different types of radio capabilities at different layers. In such a structure, nodes are first dynamically grouped into multi-hop clusters. Each group elects a cluster-head to be a backbone node (BN). Then higher-level links are established to connect the BNs into a backbone network. The backbone network helps routing and achieves good performance. However, it is a best effort routing algorithm, which does not provide QoS routing. Also the deployment of backbone nodes is different between our QoS routing protocol and the routing algorithm in [7].

Ye considered non-homogeneous MANETs in [9]. The authors suggest to using *reliable nodes* (in terms of both being robust to failure and being secure) in the network for efficient operations. The main contribution of [9] is to provide a deployment strategy that determines the positions and the trajectories of these reliable nodes such that a framework for reliably routing information is achieved. Also, [9] only considers best-effort routing.

Our QoS routing protocol takes advantage of the different communication capabilities of heterogeneous nodes in many ad hoc networks. Some physically more powerful nodes are chose as backbone nodes for routing. The idea of using backbone in routing has appeared in several previous works. The CEDAR [13] algorithm establishes and maintains a routing infrastructure called *core* in ad hoc networks. And routing is based on the *core*. There are several differences between CEDAR and our QoS routing protocol. We list some of the differences in the following: (1) CE-

DAR considers homogeneous nodes, while B-QoS considers heterogeneous nodes. The heterogeneous node model is more realistic and provides efficient routing. (2) In CEDAR, some complex algorithm is used to generate and maintain the core nodes, and the algorithm introduces large overhead, since every node needs to broadcast messages to its neighbors periodically. While in our QoS routing, the election of backbone nodes is very simple, the first backbone-capable (more powerful) node that sends out claim message becomes the backbone node. (3) In addition, CEDAR needs to broadcast route probe packet to discover the location of a destination node. While in our QoS routing, global positioning system (GPS) is used to provide node location information, and an efficient algorithm is used to disseminate node location information.

In the new QoS routing protocol, node location information is used to simplify the routing strategy. The advances in the development of global positioning system (GPS) nowadays make it possible to provide location information with a precision within a few meters. Location information can be used for directional routing in distributed ad hoc systems. Research has shown that geographical location information can improve routing performance in ad hoc networks. Additional care must be taken into account in a mobile environment, because locations may not be accurate by the time the information is used.

Routing with assistance from geographic location information requires each node to be equipped with the GPS. This requirement is quite realistic today since such devices are inexpensive and can provide reasonable precision. Several routing algorithms based on location information have been proposed. The well-known location based routing algorithms are location-aided routing (LAR) protocol [10], geographic addressing and routing [18], distance routing effect algorithm for mobility (DREAM) [16] and greedy perimeter stateless routing (GPSR) [14].

3. The path bandwidth calculation algorithm

In TDMA, to provide a bandwidth of B slots on a given path P , it is necessary that every node

along the path finds at least B slots to transmit to its downstream neighbor, and these slots do not interfere with other transmissions. Because of these constraints, the end-to-end bandwidth on the path is not simply the bandwidth on the bottleneck link. The path bandwidth calculation problem can be formulated as follows:

In a network $G = (N, L)$, where N is the finite set of nodes and L is the set of bi-directional links. Given the current transmission schedule $\{T_i\}$ at each node i , for a given path $P = \{n_S \rightarrow n_{S-1} \rightarrow \dots \rightarrow n_1 \rightarrow n_0\}$, where n_S is the source and n_0 is the destination, and $(n_i, n_{i-1}) \in L$. Need to find the set $\{T_i^P\}$, for all nodes from n_S to n_1 , where $T_i \cap T_i^P = 0$, and the set $\{T_i \cup T_i^P\}$ still satisfy the conflict-free property, and the end-to-end bandwidth on P is maximized. The set $\{T_i^P\}$ is the set of slots when node n_i along P transmits to the downstream node n_{i-1} . The objective is to find the set of new transmission slots $\{T_i^P\}$ for each node along P so that these transmissions are conflict-free, and the path bandwidth is maximized.

To resolve slot scheduling at the same time as available bandwidth is searched on the entire path is equivalent to solve the satisfiability problem (SAT) which is known to be NP-complete [28]. We assume the number of hops in path P is not very large. And this is true for many ad hoc networks, especially for the heterogeneous ad hoc networks considered in this paper. Based on the above assumption, we present a new algorithm that computes the optimal solution for the bandwidth calculation problem, i.e., it gives the maximum bandwidth in P , and the sets of transmission slots— $\{T_i^P\}$ for each node n_i , given the current transmission schedule $\{T_i\}$ at each node i .

Each node n_i computes the maximum bandwidth from source n_S to n_i . The current available transmission slots at each node n_i is denoted as AS_i . The required bandwidth is B_0 . The algorithm is given below.

3.1. The bandwidth calculation algorithm

At node n_k ,

If the number of available slots at node n_k — $|AS_k|$ is less than the required bandwidth B_0 , then stop. The QoS cannot be satisfied.

Update the upper bound of the maximum bandwidth $UB = \min\{UB, |AS_k|\}$.

Initially set the bandwidth B as UB , then reduce 1 each loop.

For all the B -slot set T_S^B at node n_S ,

For all the B -slot set T_{S-1}^B at node n_{S-1} ,

.....,

For all the B -slot set T_k^B at node n_k ,

Select a set T_k^B with B -slots from node k that is conflict free with the slot sets of the two upstream neighbors T_{k+1}^B and T_{k+2}^B .

If there is no such slot set available in node k , back to the outer loop, and change the B -slot set T_{k+1}^B in node $k+1$.

If all the B -slot sets in T_{k+1}^B have been tried and none work out, change the B -slot set T_{k+2}^B in node $k+2$. And so on.

If all the B -slot sets in node n_S have been tried, and none work out, then reduce B to $B-1$, and start over. Whenever B is less than the required bandwidth B_0 , the computation stops. And the required QoS cannot be met.

If a serial of B -slot sets $\{T_k^B\}$ are found, then B is the maximum bandwidth in path P . And the slot sets $\{T_k^B\}$ are the transmission slots for the nodes along P . The slot sets $\{T_k^B\}$ are conflict free, because during the calculation, we make sure every set T_k^B is conflict free with the two upstream sets T_{k+1}^B and T_{k+2}^B .

Basically, the above algorithm tries all the combinations of B -slot sets from different nodes, and check if there is one combination that satisfies the conflict free requirement. To make the computation more efficient, an upper bound of the slot number is maintained. The upper bound is the minimum of the available slots from the source node to current node. Also, if the slot number is less than the required bandwidth, the computation is stopped. When the number of hops in the path is not very large, the computation load of the above algorithm is no heavy.

Table 1 shows an example of using the bandwidth computation algorithm to calculate the maximum available bandwidth in a path. The second row is the available transmission slots at

Table 1
Bandwidth calculation for a given path

Node	4	3	2	1	0
Slots	1, 5	2, 3, 4, 6	1, 2, 3, 5, 6	2, 6	1, 3, 4
	1, 5	2, 3	6 ●		
	1, 5	2, 4	3, 6	●	
	1, 5	2, 6	3●		
	1, 5	3, 4	2, 6	●	
	1, 5	3, 6	2●		
	1, 5	4, 6	2, 3	●	
	1	2	3	6	1

each node. In order to make the transmission conflict free, the transmission slot at node k cannot be the same as the slots at the two nearby upstream node $k + 1$ and $k + 2$. The (red) spot means that the slot allocation is stopped because no enough conflict-free slots available.

After seven times unsuccessfully tries, a feasible slot allocation is found. And the available bandwidth in the path is obtained. In this example, it is 1. As the example shows, it does not require much computation to get a conflict free slot sets for the nodes along a given path.

4. Utilizing backbone nodes in QoS routing protocol

Most existing routing protocols for ad hoc networks consider homogeneous nodes, i.e., all nodes are the same. They have the same transmission range, the same bandwidth, etc. But in many real mobile ad hoc networks, quite often nodes have different communication capabilities. So it would be more realistic to model these network nodes as different type of nodes. Also there are advantages in terms of better routing performance when nodes in MANET are modeled as different type of nodes.

This proposition is realistic in the sense that in typical ad hoc deployments one can envision the presence of multiple types of nodes. For simplicity, we only consider there are two types of nodes in the network. One type of the nodes has longer transmission range, larger bandwidth and is more reliable. We refer to these powerful nodes as

Backbone-Capable nodes, in short as BC-nodes. BC-nodes can be elected as Backbone nodes, in short as B -nodes. Other nodes are referred to as general nodes.

Some of the advantages of deploying B -nodes are

1. Provide more reliability and fault tolerance, since B -nodes are more reliable.
2. B -nodes have larger bandwidth than general nodes. Routing traffic via B -nodes has larger chance to meet the QoS requirement.

Among all the nodes in the network, there are some backbone nodes (B -nodes) with large transmission range R and large bandwidth. These B -nodes serve as the backbone for the QoS routing in the mobile ad hoc networks. And other nodes are general nodes with transmission range r . Usually transmission range R of B -node is much larger than that of general node r . We consider the routing area to be divided into several small, equal squares—refer to as grids.

The side length of a grid is set as $a = R/2\sqrt{2}$. The relationship of a and R is shown in Fig. 1, where it is the worst case (longest distance) of two neighbors. There is a B -node in each grid whenever it is possible. So each B -node can communicate with B -nodes in neighbor grids, including the diagonal one. The QoS routing protocol is given in the follows.

4.1. The QoS routing protocol

- Each B -node can reach nearby B -nodes. If a B -node i find there is no B -node in the nearby grid (e.g., after i forwards a packet to the nearby grid, i does not hear any transmission from a B -node for a certain time) i will initiate a B -node election process for the nearby grid. The B -node election process is described in the sequel.
- There is no routing table maintained among B -nodes and general nodes. The QoS route is

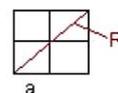


Fig. 1. The relationship between a and R .

discovered on demand. When a B -node moves out of a grid, it initiates a B -node election process in this grid. And a new B -node will be elected.

- When a source node S wants to setup a QoS route for a flow to a destination D , assume S knows the location of D .
- We first consider the case where source S is a B -node. And we will consider the case where source S is not a B -node later. If S is a B -node, it will find out the grid g_D that node D belongs to based on the location information of D . Assume the B -node in grid g_D is C . And the rectangle including node S and D is called the flooding zone. The flooding zone is illustrated in Fig. 3 as the black rectangle. The rectangle can also include the circle that centers at node D with radius being the expected moving distance, like the LAR routing protocol in [10]. A general assumption is that nodes in MANETs do not move very fast. And usually the circle is within the transmission range of the B -node C .
- Node S floods the route request (RR) packets to all the B -nodes in the flooding zone, i.e., node S uses limited flooding to find out the QoS supported route. Each flooding has a unique Flooding_ID. When forwarding the RR packet, each B -node adds its available transmission slots T_i to RR packet. And when a B -node receives the RR packet, it uses the available transmission slot set $\{T_i\}$ at all upstream nodes and the bandwidth calculation algorithm to compute the maximum bandwidth up to its location. If the maximum bandwidth is less than the required bandwidth, the RR packet is dropped. Otherwise, the B -node forwards the RR packet to its neighbor B -nodes. When a B -node receives the RR packet with the same Flooding_ID, it simply discards the RR packet. This is to reduce the flooding overhead. Although there is such possibility that a detour path may have large bandwidth than a direct link. Although there is such possibility that a detour path may have large bandwidth than a direct link, e.g., like the case in Fig. 2. The direct link on the top only has two slots, while the detour path below has three slots.
- When the RR packet arrives at B -node C , C also computes the maximum bandwidth from S to C . If the required bandwidth is satisfied, B -node C

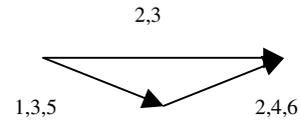


Fig. 2. The detour path with larger bandwidth.

broadcasts a probe packet to all nodes in the nearby grids. Since node C is a B -node with large transmission range, the broadcast can reach nodes in all nearby grids. If node D is still within the nearby grids, then D will receive the probe and send an Ack (acknowledge) packet back to node C . The Ack packet finds a route to C via limited-hop flooding. Since we assume nodes in MANETs do not move very fast, most of the time, node D is still within the transmission range of the B -node C . Then C will send a route reply (RP) packet along the route back to the source node S . And all the intermediate nodes reserve the slots T_i for the path P . When the RP packet arrives at the source node S , the QoS path is set-up and the bandwidth is reserved. If node C receives multiple RR packets from the same source node S (with the same Flooding_ID), node C will reply to 2 or 3 of them, and discard the rest of them. This is to setup 1 or 2 backup QoS routes in addition to the primary route. In case when primary route is broken because of node moving, the backup route is used.

- If B -node C does not receive an Ack from D within a certain waiting time, it means that D has move out of C 's nearby grids, and C will forward the RR packet to all its neighbor B -nodes. And the neighbor B -nodes will compute the maximum bandwidth from source node S to them. If the required bandwidth is satisfied, then these B -nodes will broadcast the probe packet to the nearby grids. When node D receives the probe, it sends an Ack packet to its nearby B -node by limited-hop flooding. Then the nearby B -node sends out a RP packet back to source node S . If D is still not found by neighbor B -nodes of C , those B -nodes will send the packet to their neighbor B -nodes, and so on. But this is not likely to happen. Because usually a node in mobile ad hoc networks does not move very fast, so node D should not be far away from C .

- If S is not a B -node, then S will first find out a route to the nearby B -node with enough bandwidth. Node S sends out the route request (RR) packet by flooding to all the nodes in its grid. Only nodes in the same grid will process and forward the RR packet. This is to limit the routing overhead from flooding. When nodes receive the RR packet, they will calculate their available bandwidth and compare with the required bandwidth. If QoS is satisfied, the RR packet is forwarded to their neighbors. Otherwise, the RR packet is dropped. When the B -node in this grid receives the first RR packet, it will forward the RR packet to other B -nodes if it has enough bandwidth. Since B -nodes have much larger bandwidth than general nodes, most of the time the B -node will have enough bandwidth. In case the B -node does not have enough bandwidth, the B -node will notice the source node to flood the RR packet to nodes in neighbor grids and find out a nearby B -node with enough bandwidth. And the rest is the same as above.

In the above QoS routing protocol, node location information is used in several places to aid routing. The location information is used to determine which B -node is the nearest one to the destination. One B -node is maintained in each grid based on node location information. Otherwise, some grids will have more than one B -node and some grids will have no B -nodes at all. Whether a B -node is in the flooding zone is based on the node location information.

An important idea in the above QoS routing protocol is to let most of the transmission based on B -nodes. Because B -nodes have large bandwidth, it increases the chance to find the route that satisfies the QoS requirement. Also the number of hops in the route is greatly reduced because of the long transmission range of B -nodes.

4.2. More protocol details

4.2.1. Route maintenance

Node mobility can cause established QoS route broken. Route maintenance is very important for

QoS routing in MANETs. Since the bandwidth QoS routing involves the non-trivial computation of path bandwidth, we propose a simple route maintenance method. Moving away (or failure) of any node in the route can cause the route broken, and the broken route is detected by the upstream node (closer to source), e.g., assume the upstream node $k + 1$ sends a packet to node k . Node $k + 1$ will assume the route is broken if it does not hear any transmission from the node k for a certain time. If the existing QoS route is broken because a node moves away or dies out, the upstream node in the route will send a `route_broken` packet to source, and source will start a new route discovery process and try to find a new QoS route.

4.2.2. Election of B -node

Initially, one B -node is set in each grid. Since B -nodes also move around, we need an election algorithm for new B -node. When a B -node leaves its current grid, or for any other reason (e.g., current B -node failed) there is no B -node in a grid, a B -node election process is initiated. Assume there are more BC-nodes than the number of grids in the network. The election process works as following. When a node discovers there is no B -node in the grid, it floods the election message to all the nodes in the grid. When a BC-node T receives the election message, it sends out a claim message that claims it will become the B -nodes to all the BC-nodes in the grid. Other BC-nodes will not send out claim message if it receives one claim message. And a new B -node is elected.

When the B -node is moving near the border of the grid, it may move back to the grid after shortly move out of the grid. To reduce the overhead of frequent elections, we can set a safe distance for the B -node to start the election process. For example, only when the current B -node is more than $a/5$ away from the grid border, it will send out an election message.

4.2.3. A routing example

Next, we will give a routing example by using the QoS routing algorithm in Fig. 3. Assume node S wants to send a packet to node D . And node S knows node D 's current location.

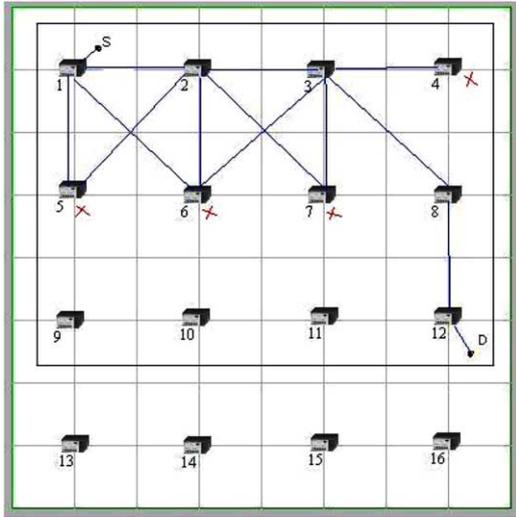


Fig. 3. The QoS routing example.

1. Node *S* sends out the route request (RR) packet by flooding to all the nodes in its grid. A *B*-node 1 is within the transmission range of node *S* and has enough bandwidth.
2. Then *B*-node 1 forwards the RR packet to neighbor *B*-nodes 2, 5 and 6. Each *B*-node calculates the available bandwidth and compares it with the required bandwidth. In the example, nodes 5 and 6 do not have enough bandwidth to meet the QoS requirement, so they drop the RR packet. Node 2 has enough bandwidth, so it adds its available slots to RR packet and forwards it to the neighbor *B*-nodes. Node 7 does not have enough bandwidth, but node 3 does. And this process continues, until the RR packet reaches node 12.
3. Node 12 has enough bandwidth and it finds out the destination node *D* is in its grid. So node 12 sends a route reply (RP) packet back to node 1 along the reverse path. Then node 1 sends the RP packet to node *S*. And all the intermediate nodes reserve the slots for the QoS flow. The QoS path $s \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 8 \rightarrow 12 \rightarrow D$ is successfully setup.

For simplicity, only *B*-nodes, source and destination are plot in Fig. 3. As we can see from the routing example, most of the transmissions are

based on *B*-nodes. And this provides efficient and effective QoS routing.

4.2.4. Routing overhead

Also we can see from the above routing example, the routing overhead is much smaller than traditional on demand routing. It only involves a small area of flooding in the grid of the source node *D*, plus the flooding among the *B*-nodes. *B*-nodes are only a small fraction of the total nodes, so the overhead from *B*-node flooding is not large.

5. Simulations and test results

The QoS routing protocol is implemented in QualNet, a scalable packet-level simulator with an accurate radio model. TDMA is used as the MAC protocol. The transmission rate of the general node and the *B*-node are 1 and 5 Mbps respectively. There are 40 slots in a TDMA frame. The simulation test-bed that we used consists of 30 general nodes and nine *B*-nodes uniformly distributed at random in an area of 1000 m × 1000 m. The radio transmission ranges of the general node and the *B*-node are 100 and 400 m respectively. Each simulation was run for 600 simulated seconds. The mobility in the environment was simulated using a random-waypoint mobility model, wherein each node randomly chooses a point in the field and moves towards it at a randomly chosen velocity. The node then pauses for a specified period at the destination before continuing the same pattern of motion. In our simulations, the pause time was set to 0 s, which corresponds to constant motion. We control the node mobility by varying the maximum node velocities. The maximum velocities range from 0 to 30 m/s.

The data traffic is generated with CBR sources, where the source and the destination are chosen before the simulation. We ran each simulation 20 times to get an average result for each simulation configuration.

We conduct two kinds of simulation tests. First, we compare the performance of three types of QoS traffics at different node mobility. The three types of QoS traffics are QoS-1, QoS-2 and QoS-4, which need one, two and four data slots in each

frame, respectively. Second, we compare our QoS routing protocol with best effort AODV routing protocol. The details are given in the following.

5.1. Performance comparison of three QoS traffics

In the simulation, we study how the mobility affects the system performance. In the first experiment, we consider the effect of variable mobility on the rerouting due to a broken path. If any link on the path is broken because nodes move away, the QoS path needs to be rerouted. Fig. 4 presents the simulation result.

In the simulation, we vary the maximum node speed from 5 to 30 m/s. And we observe that the percentage of rerouted traffic increases as the mobility increases for all three types of QoS traffics. This is because high node mobility causes path to be broken more frequently. However, the rerouting percentage is not very high even when the maximum node speed is 30 m/s, where the rerouting percentage is about 25%. This is because most links in the QoS path are links between *B*-nodes. And *B*-nodes have large transmission range. So even nodes move fast, for most of the time, *B*-nodes are still within the transmission range of the neighbor *B*-nodes, and hence the links are still good.

Also we observe that the rerouting percentage is independent of the QoS type. Even though a path

for a QoS-1 request is easier to setup than a path for a QoS-4 request, once they are setup, the effects of node mobility to the path broken are the same for the two cases, i.e., we measure the fraction of connections that have already setup a QoS path and need to be rerouted during their active periods.

The second experiment is to measure the average throughput at different node mobility. As we can see from Fig. 5, the average throughput decreases as the node speed increases. High mobility causes more link broken and frequent rerouting, and thus causes more packet loss and larger end-to-end transmission delay. The high QoS connection (e.g., QoS-4) has higher throughput on average because of the high data rate (more data slots). Again, since most of the links in the QoS path are links between *B*-nodes, these links are more reliable and less affected by node mobility. The average throughput does not decrease much as the mobility increases. When the maximum node speed is 30 m/s, the average throughput is about 75% of the average throughput when maximum node speed is 5 m/s.

Fig. 6 reports the percentage of the first *B*-node in the same grid as the source node for different QoS requirement and mobility. *B*-nodes have larger bandwidth (5 Mbps) than general nodes (1 Mbps). So in most cases, the *B*-node in the same grid can satisfy the QoS requirement of the source

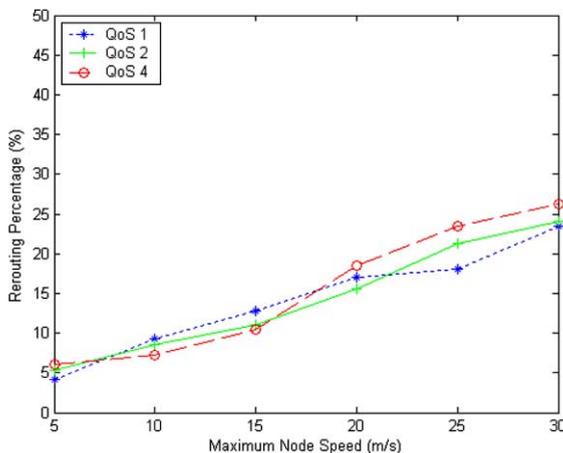


Fig. 4. The rerouting percentage vs. node mobility.

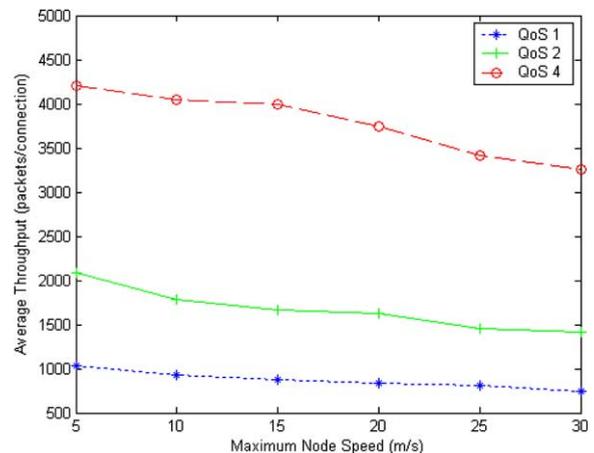


Fig. 5. Average throughput vs. node speed.

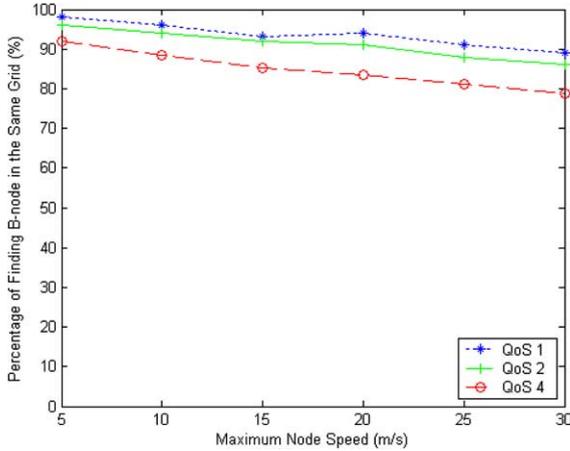


Fig. 6. The percentage of finding *B*-node in the same grid for different QoS.

node, and serve as the first *B*-node in the path. Fig. 6 shows that the percentage is very high. For QoS-1 and QoS-2 traffics, the percentage is always higher than 90%. For QoS-4 traffic, the percentage is still higher than 80%. This means that the source node only needs to flood the RR packet in a small area to find the starting *B*-node. So the routing overhead is greatly reduced.

Also we can see from Fig. 6, the percentage of finding *B*-nodes in the same grid decreases as the required bandwidth increases. The percentage of QoS-1 is higher than the percentage of QoS-2, which is higher than the percentage of QoS-4. For all the three types of QoS traffics, the percentage decreases slightly as the node mobility increases. This is because when node moves fast, *B*-nodes are more likely to move out of current grid, which causes no *B*-node in the current grid for a short time (before the new *B*-node is elected).

5.2. Performance comparison with AODV

The following set of experiments is to compare the performance of the QoS routing protocol with best effort AODV routing protocol. AODV [17] is an on-demand routing protocol that uses flooding to discover route. We choose AODV as the routing protocol for comparison because AODV has a similar route discovery mechanism as the QoS routing protocol.

First we measure the average packet delay for both AODV and the QoS routing protocol. In the simulation, the QoS traffic is a mixture of three types of QoS traffics: QoS-1, QoS-2 and QoS-4. The source–destination pair randomly determines its QoS type with equal probability. Fig. 7 presents the average packet delay for the two routing protocols under different network load. The network traffic load is varied by changing the number of CBR sessions generated during the simulation. The packet delay also depends on the node moving speed. In the test, the maximum node speed is set as 15 m/s for both routing protocols.

When the traffic is light, e.g., number of transmitted packets is less than 6×10^3 , the packet delay is very close for AODV and the QoS routing protocol. The QoS routing protocol performs much better than AODV when the traffic gets heavy. As the network traffic becomes heavy, the average packet delay of AODV increases very fast, reaching over 500 ms when the network load is 10^4 packets. While the packet delay of the QoS routing protocol does not increase much, only about 200 ms under the same network load. This is because for the best effort AODV routing protocol, one source–destination pair usually uses only one active route for all the packet flows. And when the network traffic is heavy, this single route becomes heavily loaded, causing packets to be delayed or dropped, and hence increases the packet delay a lot. On the other hand, the QoS routing protocol

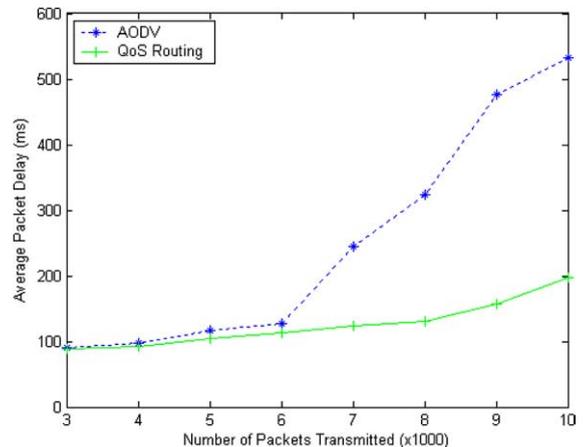


Fig. 7. Average packet delay.

tries to find and use routes satisfying bandwidth constraints for different flows, even between the same pair of source and destination. So when a particular route does not have enough bandwidth, a new route is discovered. Thus, the network load is balanced among different routes by the QoS routing protocol. And the average packet delay increases slowly when the network load increases.

We also compare the throughput of AODV and the QoS routing protocol under different network traffic load. Node mobility also affects the throughput. Fig. 8 reports the packet throughputs of the two routing protocols for two different maximum node speeds: 5 and 20 m/s.

We observe from Fig. 8, under light traffic load, the throughputs are close for AODV and the QoS routing protocol. The advantage of QoS routing protocol becomes apparent when traffic becomes heavy. For example, when the network load is 10^4 packets, the number of received packet for the QoS routing protocol is 9×10^3 , while the number for AODV is only about 6.5×10^3 . This is because of the same reason for the difference in packet delay. The QoS routing protocol distributes traffic load over different routes. So the throughput of the QoS routing protocol does not decrease much as the network load becomes heavy. While the best effort AODV only uses one route for multiple flows, and as the network load getting heavy, the throughput drops significantly. Also, from Fig. 8,

we can see that the throughput decreases as the node mobility increases for both routing protocols. This is because larger node mobility causes more broken links, and more packets dropped.

From Fig. 8, we also find out that the throughput performance for the QoS routing protocol at maximum mobility of 20 m/s is better than AODV at maximum mobility of 5 m/s. This demonstrates the good performance of the QoS routing protocol.

The above simulations demonstrate that the QoS routing protocol has better performance than AODV, especially when the network traffic load is heavy. The QoS routing protocol builds different QoS routes for different flows, even between the same source and destination. Packets transmitted on QoS routes are guaranteed of bandwidth. When an area of the network is congested, the QoS routing protocol will find a new route to go around the area, and balance the load in the network.

For real-time multimedia traffic (like audio, video), we can expect that the QoS routing protocol will perform much better than the best effort routing protocols. Because the best effort routing protocols do not provide any QoS guarantee, many flows will be blocked due to insufficient bandwidth, and causes severe performance degradation. This will be our future work. In next section, we will show that the probability of having one B -node in each grid is very large.

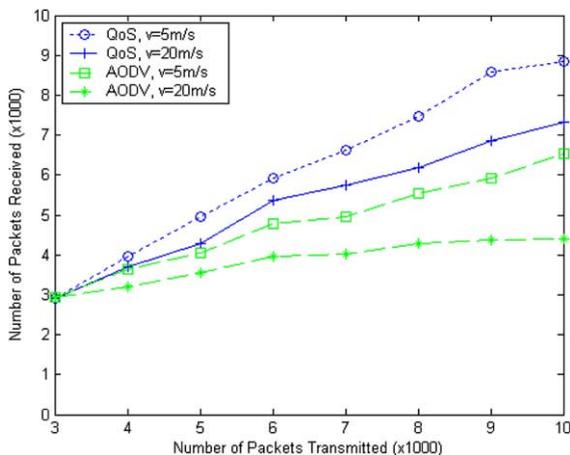


Fig. 8. Comparison of throughputs.

6. The probability of having B -node in one grid

To ensure the QoS routing protocol work well, an important aspect is to have at least one B -node in each grid. If there is no B -node in one grid, then more flooding is needed and it will cause larger routing overhead.

We divide the routing area into several small squares—grids. Let F be the set of all grids. The Vapnik–Chervonenkis dimension of F is 3. And we will use the following VC-dimension Theorem to obtain the probability of one grid having a B -node.

Theorem 1 (The Vapnik–Chervonenkis Theorem). *If F is a set of finite VC-dimension $VC - d(F)$, and $\{X_i\}$ is a sequence of i.i.d. random*

variables with common probability distribution P , and denote E as a element of F , $E \in F$, then for every $\varepsilon, \delta > 0$, we have

$$P\left(\sup_{E \in F} \left| \frac{1}{n} \sum_{i=1}^n I(X_i \in E) - P(E) \right| \leq \varepsilon\right) > 1 - \delta$$

when

$$n > \max \left\{ \frac{8 \times \text{VC} - d(F)}{\varepsilon} \log \frac{16e}{\varepsilon}, \frac{4}{\varepsilon} \log \frac{2}{\delta} \right\}.$$

Assume the nodes are uniformly randomly distributed in the routing area. Then we can use uniform convergence in the weak law of large numbers. Assume there are a^2 grids in the routing area. And there are totally n B -nodes. Then we have, $P(E)$ = probability (having a B -node in one grid) = n/a^2 . We choose $\varepsilon(n) = \delta(n) = 20 \times \log n/n$.

Denote $n_1 = \max((8 \times \text{VC} - d(F))/\varepsilon) \log 16e/\varepsilon = (24/\varepsilon) \log 16e/\delta$, and $n_2 = (4/\varepsilon) \log 2/\delta$. We plot the curve of n, n_1 and n_2 in Fig. 9. And we can see from Fig. 9, n is larger than both n_1 and n_2 when n is larger than 10. Which means $n > \max\{((8 \times \text{VC} - d(F))/\varepsilon) \log 16e/\varepsilon, (4/\varepsilon) \log 2/\delta\}$ is satisfied when $n > 10$. So we have when $n > 10$,

$$P\left(\sup_{E \in F} \left| \frac{1}{n} \sum_{i=1}^n I(X_i \in E) - P(E) \right| \leq \varepsilon\right) > 1 - \delta,$$

i.e.,

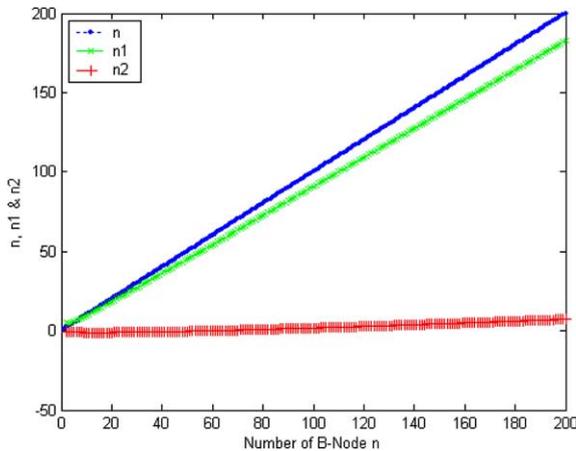


Fig. 9. The curve of n, n_1 and n_2 .

$$P\left(\sup_{E \in F} \left| \frac{\text{Number of nodes in } E}{n} - \frac{n}{a^2} \right| \leq \varepsilon\right) > 1 - \delta \Rightarrow$$

$$P\left(\text{Number of nodes in } E \geq n\left(\frac{n}{a^2} - \varepsilon\right)\right) > 1 - \delta.$$

Recall that we choose $\varepsilon(n) = \delta(n) = 20 \times \log n/n$. If the number of B -nodes n is around a^2 —the number of grids, we assume $n = a^2$ for simplicity, then we can plot the curve of $n_0 = n(n/a^2 - \varepsilon) = n(1 - 20 \log n/n)$ in Fig. 10(a). And the curve shows that n_0 is always larger than 1 when n is larger than 40.

We also plot the curve of $1 - \delta$ in Fig. 10(b). As we can see that $1 - \delta$ is always greater than zero

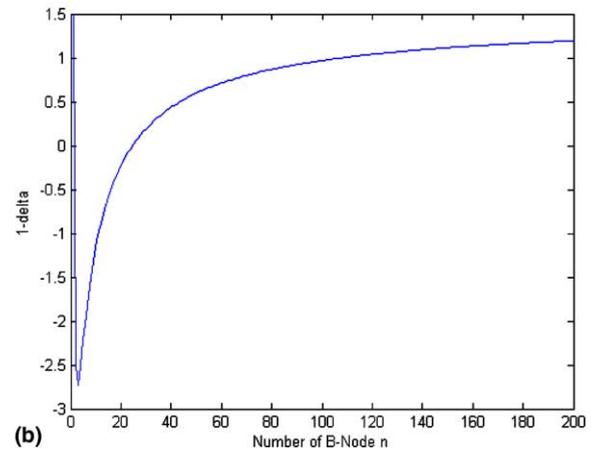
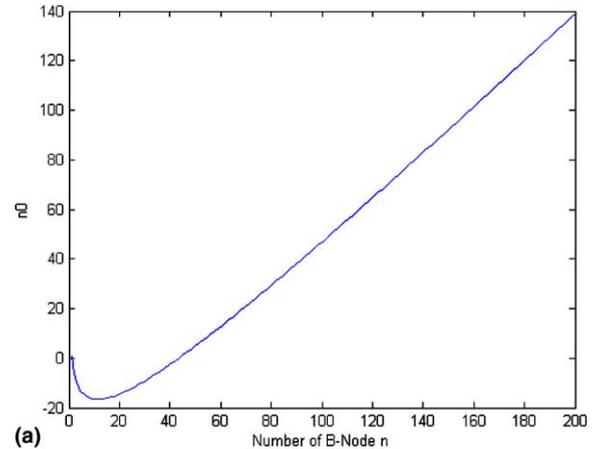


Fig. 10. (a) The curve of n_0 . (b) The curve of $(1 - \delta)$ —the lower bound of having B -node in a grid.

when n is larger than 25, and it is greater than 0.5 when n is larger than 40. The value of $1 - \delta$ goes to 1 as n increases. This can easily be verified from $1 - \delta(n) = 1 - 20 \log n/n \rightarrow 1$ as $n \rightarrow \infty$.

So we have the following proposition.

Proposition 1. *When the number of B-nodes n is equal to or larger than the number of grids a^2 , and a^2 is greater than 40, the probability of having at least one B-node in each grid is larger than $1 - 20 \log n/n$. And this probability goes to 1 as $n \rightarrow \infty$, i.e., when n is sufficient large, each grid will have at least one B-node with very high probability.*

Proof. The proof follows from the above discussion. \square

We also use simulations to measure the probability of having no B-node in a grid. The simulations are run with two different setting. One routing area has 16 grids and the other has 25 grids. We measure the probability in the following way. We record the time t_j of each grid j having no B-node during the whole simulation process. Then we add all the t_j together, and divided by the product of number of grids and the simulation time, i.e., $P(\text{no B-node}) = \sum t_j / (a^2 T)$, where a^2 is the number of grids, and T is the simulation time. And we measure the chance of having no B-node in a grid when the number of B-nodes varies from a^2 (the number of grids) to $2.5a^2$.

The test results are given in Fig. 11. The maximum node speed is set as 15 m/s. And the pause time is set to 0, which means the nodes continuously move during the simulation. In Fig. 11, the top curve is the probability of having no B-nodes in a grid for the 16-grid case, and the bottom curve is for the 25-grid case. The Figure shows when the number of B-nodes is in proportional to the number of grids, the network with larger number of grids has smaller probability of no B-node in a grid. This is because larger number of grids means larger number of B-nodes in the entire routing area. This is why the curve of 16 grids in on top of the curve of the 25 grids. Also from Fig. 11, the probability of having no B-node in one grid decreases when the total number of B-nodes increases. When the total number of B-nodes is

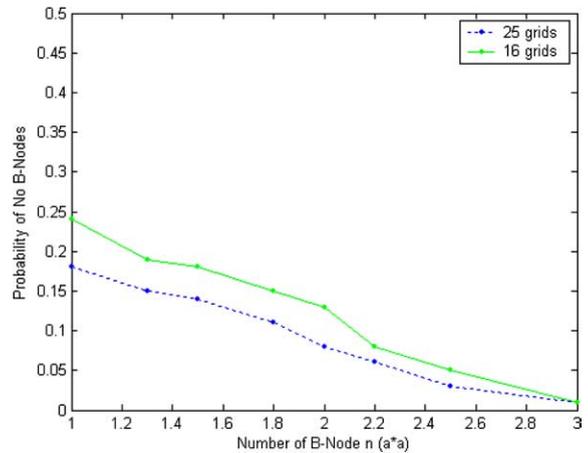


Fig. 11. The probability of having no B-nodes vs. the number of total B-nodes.

larger than 2.5 times the number of grids, the probability of no B-node is very low, lower than 0.05. This means with a very high probability, there is a B-node in each grid.

7. Conclusions

In this paper, we propose a new QoS routing protocol that contains bandwidth calculation and slot reservation for mobile ad hoc networks. The QoS routing protocol takes advantage of the different transmission capability of multi-class nodes. Most of the links of the QoS route are among B-nodes, which have large bandwidth and large transmission range. Large bandwidth greatly increases the chance to meet the QoS requirement. And large transmission range reduces the number of hops in the route, and thus reduces the routing overhead. The QoS routing protocol has very large chance to setup QoS routes successfully. In the performance experiments, traffic flows with different QoS types are considered. Simulation results show the performance advantage of the QoS routing protocol. We also compare the performance of the QoS routing protocol with a popular routing protocol—AODV. Our simulation tests show that the QoS routing protocol works better than AODV. The QoS routing protocol balances load among different routes and it has lower

packet delay and higher throughput than AODV. By using VC-dimension theorem, we show that the probability of having one B -node in each grid is large, given a reasonable number of B -nodes. We also demonstrate it by our simulations.

References

- [1] C. Zhu, M. Corson, QoS routing for mobile ad hoc networks, in: Proc. IEEE INFOCOM 2002, pp. 958–967.
- [2] C.R. Lin, On-demand QoS routing in multihop mobile networks, in: Proc. IEEE INFOCOM 2001, pp. 1735–1744.
- [3] C.R. Lin, J.-S. Liu, QoS routing in ad hoc wireless networks, *IEEE J. Select. Areas Commun.* 17 (8) (1999) 1426–1438.
- [4] E. Gelenbe, R. Lent, Z. Xu, Towards networks with cognitive packets, in: Proc. IEEE MASCOTS Conference, San Francisco, CA, 2000, pp. 3–12.
- [5] E. Gelenbe, R. Lent, Z. Xu, Measurement and performance of cognitive packet networks, *Comput. Networks* 37 (2001) 691–701.
- [6] E. Gelenbe, R. Lent, A. Montuori, Z. Xu, Cognitive packet networks: QoS and performance, in: Proc. IEEE MASCOTS Conference, Fort Worth, TX, 2002, pp. 3–12.
- [7] K. Xu, X. Hong, M. Gerla, An ad hoc network with mobile backbones, in: Proc. IEEE ICC 2002, vol. 5, pp. 3138–3143.
- [8] E. Gelenbe, P. Liu, J. Lainé, Cognitive packet network routing with genetic algorithms, in: Proc. SPECTS'03, Summer Simulation Multiconference, Society for Computer Simulation, 20–24 July 2003, Montreal, Canada, pp. 519–524.
- [9] Z. Ye, S. Krishnamurthy, S. Tripathi, A framework for reliable routing in mobile ad hoc networks, *IEEE INFOCOM 2003*, San Francisco, CA, April 2003.
- [10] Y.-B. Ko, N.H. Vaidya, Location-aided routing (LAR) in mobile ad hoc networks, *ACM/IEEE Int. Conf. on Mobile Comp. Net.*, 1998, pp. 66–75.
- [11] E. Gelenbe, E. Seref, Z. Xu, Towards networks with intelligent packets, in: Proc. IEEE-ICTAI Conference on Tools for Artificial Intelligence, 1999, pp. 47–60.
- [12] E. Gelenbe, M. Gellman, P. Su, Using loss and delay as QoS goals in cognitive packet networks, in: Proc. SPECTS'03, Summer Simulation Multi conference, Society for Computer Simulation, 20–24 July 2003, Montreal, Canada, pp. 508–513.
- [13] R. Sivakumar et al., CEDAR: a core-extraction distributed ad hoc routing algorithm, in: Proc. INFOCOM 1999, pp. 202–209.
- [14] B. Karp, H.T. Kung, GPSR: greedy perimeter stateless routing for wireless networks, in: Proc. 6th Annual Int. Conf. on Mobile Computing and Networking (MobiCom 2000), Boston, MA, USA, 2000, pp. 243–254.
- [15] Y.-C. Hsu, T.-C. Tsai, Bandwidth routing in multihop packet radio environment, in: Proc. 3rd Int. Mobile Computing Workshop, 1997.
- [16] S. Basagni et al., A distance routing effect algorithm for mobility (DREAM), *ACM/IEEE Int. Conf. on Mobile Comp. Net.*, 1998, pp. 76–84.
- [17] C. Perkins, E. Royer, Ad hoc on-demand distance vector routing, in: Proc. of the 2nd IEEE Workshop on Mobile Computing Systems and Applications, New Orleans, LA, February 1999, pp. 90–100.
- [18] J. Li et al., A scalable location service for geographic ad hoc routing, in: Proc. of MobiCom 2000, pp. 120–130.
- [19] E. Gelenbe, A. Nunez, R. Lent, Traffic balancing via smart packets, in: Proc. 2003 IEEE Workshop on IP Operations and Management, 1–3 October 2003, Kansas City, MO, pp. 15–21.
- [20] E. Gelenbe, Review of experiments in self-aware networks, in: Proc. ISCS XVIII, Lecture Notes in Computer Science, vol. 2869, Springer, Berlin, 2003, pp. 1–8.
- [21] E. Gelenbe, M. Gellman, P. Su, Self-awareness and adaptivity for quality-of-service, in: Proc. 8th IEEE Int. Symp. on Computers and Communications, Antalya, July 2003, pp. 3–19.
- [22] G. Xue, Minimum cost QoS multicast and unicast routing in communication networks, *IEEE Trans. Commun.* 51 (2003) 817–824.
- [23] S. Chakarabarti, A. Mishra, QoS issues in ad hoc wireless networks, *IEEE Comm. Mag.* 39 (2) (2001) 142–148.
- [24] E. Elmallah et al., Supporting QoS routing in mobile ad hoc networks using probabilistic locality and load balancing, in: Proc. IEEE GLOBECOM 2001, vol. 5, pp. 2901–2906.
- [25] H.F. Salama, D.S. Reeves, Y. Viniotis, A distributed algorithm for delay-constrained unicast routing, in: Proc. IEEE INFOCOM'97, pp. 84–91.
- [26] I. Cidon, R. Rom, Y. Shavitt, Multi-path routing combined with resource reservation, in: Proc. IEEE INFOCOM'97, pp. 92–100.
- [27] S. Chen, K. Nahrstedt, Distributed quality-of-service in ad-hoc networks, *IEEE J. Select. Areas Commun.* 17 (8) (1999) 1488–1505.
- [28] C.R. Lin, M. Gerla, Adaptive clustering for mobile wireless networks, *IEEE J. Select. Areas Commun.* 15 (7) (1997) 1265–1275.
- [29] S. Lee, A.T. Campbell, INSIGNIA: in-band signalling support for QoS in mobile ad hoc networks, in: Proc. of the 5th Int. Workshop on Mobile Multimedia Communication, 1998.
- [30] J. Tsai, T. Chen, M. Gerla, QoS routing performance in multihop, multimedia, wireless networks, in: Proc. of IEEE ICUPC, 1997.