PAPER *Special Section on Mobile Multimedia Communications*

A Priority-Based QoS Routing for Multimedia Traffic in Ad Hoc Wireless Networks with Directional Antenna Using a Zone-Reservation Protocol

Tetsuro UEDA[†a)], **Shinsuke TANAKA**[†], ***Members***, **Siuli ROY**[††], **Dola SAHA**[††],
and **Somprakash BANDYOPADHYAY**[††], ***Nonmembers***

SUMMARY Quality of Service (QoS) provisioning is a new but challenging research area in the field of Mobile Ad hoc Network (MANET) to support multimedia data communication. However, the existing QoS routing protocols in ad hoc network did not consider a major aspect of wireless environment, i.e., mutual interference. Interference between nodes belonging to two or more routes within the proximity of one another causes *Route Coupling*. This can be avoided by using *zone-disjoint routes*. Two routes are said to be *zone disjoint* if data communication over one path does not interfere with the data communication along the other path. In this paper, we have proposed a scheme for supporting priority-based QoS in MANET by classifying the traffic flows in the network into different priority classes and giving different treatment to the flows belonging to different classes during routing so that the high priority flows will achieve best possible throughput. Our objective is to reduce the effect of coupling between routes used by high and low priority traffic by *reserving zone of communication*. The part of the network, used for high priority data communication, i.e, *high priority zone*, will be avoided by low priority data through the selection of a different route that is maximally zone-disjoint with respect to high priority zones and which consequently allows contention-free transmission of high priority traffic. The suggested protocol in our paper selects shortest path for high priority traffic and diverse routes for low priority traffic that will minimally interfere with high priority flows, thus reducing the effect of coupling between *high* and *low priority routes*. This adaptive, priority-based routing protocol is implemented on Qualnet Simulator using directional antenna to prove the effectiveness of our proposal. The use of directional antenna in our protocol largely reduces the probability of radio interference between communicating hosts compared to omni-directional antenna and improves the overall utilization of the wireless medium in the context of ad hoc wireless network through Space Division Multiple Access (SDMA).

key words: *ad hoc networks, quality of service, route coupling, directional antenna, space division multiple access, priority based routing*

1. Introduction

The rising popularity of multimedia applications and potential commercial usage of Mobile Ad hoc Network (MANET) clearly indicates that Quality of Service (QoS) support in MANETs, is an unavoidable task. Specially, the thrust is on the development of QoS networks that will guarantee the delivery of time sensitive multimedia data. Essentially, Multimedia traffic should get preference over conventional data traffic during communication through some kind of priority-based resource reservation to assure a timely and guaranteed delivery of multimedia data. Numerous solutions to the QoS problems have been proposed in MANET [1]–[7]. However, the existing QoS routing protocols in MANET did not consider a major aspect of wireless environment, i.e., mutual interference. QoS between any source and destination does not depend only on the congestion characteristics of the nodes in that path. Pattern of communication in the neighborhood region will also contribute to the performance. Thus, even two routes don't share the same node, routing performance will deteriorate by the interference from the neighbors.

Our objective is to devise a priority based routing scheme to reduce the effect of mutual interference between routes used by high and low priority traffic by *reserving high priority zone of communication*. The part of the network, used for high priority data communication will be temporarily reserved as *high priority zone*, which will be avoided by low priority data through the selection of a different route that is maximally zone-disjoint with respect to the reserved high priority zone and which consequently allows contention-free transmission of high priority traffic.

Several researchers have already studied and addressed the issues related to QoS in MANET. QoS support in MANET includes QoS models, QoS Resource Reservation Signaling, QoS Routing and Medium Access Control (MAC) [1], [2]. Basic concept of QoS Models currently available for the wired network like IntServ/RSVP and Diffserv are not suitable for MANET because they produce a huge overhead at each mobile host. Moreover, in IntServ the amount of flow-based state information to be kept in each host increases proportionally with the number of flows that raises scalability problems. The scalability issue in QoS Internet can be solved with DiffServ. Unlike IntServ, interior routers in DiffServ keep aggregated flow based state information instead of per-flow state information [1]–[3]. So DiffServ may be a potential QoS model for MANET. RSVP, the QoS signaling protocol used in Internet, is not suitable for MANET as the control message for resource reservation will contend with data packets. Due to the unpredicted mobility of the hosts, the traditional meaning, that some performance metrics must be guaranteed once a request is accepted in a QoS enabled network, is no longer true in MANET. The RSVP kind of resource reservation is difficult in an ad hoc mobile environment as mobility may

---

Manuscript received August 24, 2003.

Manuscript revised October 23, 2003.

[†] The authors are with ATR Adaptive Communications Research Laboratory, Kyoto-fu, 619-0288 Japan.

[††] The authors are with Indian Institute of Management, Calcutta, DH Road, Calcutta 700104, India.

a) E-mail: teueda@atr.jp

cause link break or may change link quality and thus the existing path may fail to provide QoS requirement. So the QoS models, signaling and routing protocols of wired network cannot be directly mapped to MANET [4]. A proper blend of IntServ and DiffServ is essential to achieve an Assured Service for an application requiring better reliability than Best Effort Service.

A Flexible QoS Model for MANET (FQMM) is proposed in [5] that uses the merits of both IntServ and DiffServ model. In [6] a comprehensive study on several issues related to QoS provisioning in ad hoc network is done. The authors have also discussed few proposals for QoS provisioning at different layers and inter-layer QoS framework for MANET in this survey. Xavier Pallot et al. have proposed in [7] that limited bandwidth of the mobile radio channel prevents giving every class of traffic the same QoS except when the network is very lightly loaded. So, some means for providing each class a different QoS must be implemented by assigning priority to one class over another class in terms of allocating resources. Thus, linkage between *QoS* and *Priority* is a common one in the literature, and the two terms are almost synonym [7]. So, QoS provisioning through priority based service is an interesting idea that is worth exploring in MANET.

Several efforts have been made to support QoS in MANET by changing the size of contention window (CW) according to the priority of traffic in MAC layer and modifying *backoff algorithm* accordingly [11]. But it does not guarantee that high priority packet will always get a contention free access to the medium for data communication. There is a chance that low priority traffic may choose a contention window size smaller than that of high priority traffic and get a chance to transmit low priority data even if high priority data is present in the network. Moreover, two high priority flows contending for the medium may not always get guaranteed fair access of the medium in these schemes.

None of the existing priority-based QoS routing protocols [1]–[7] consider the effect of mutual interference between routes in wireless medium during routing, though coupling is one of the major causes for degradation in network performance. *Route Coupling* is a phenomenon of wireless medium that occurs when two routes are located physically close enough to interfere with each other during data communication [8], [9]. The nodes on those interfering routes will constantly contend to access the wireless medium they share. Eventually, average end-to-end delay for a packet will increase, which will drastically reduce the throughput. So, end-to-end delay and throughput, which are two major concerns in supporting priority-based QoS in MANET, are heavily dependent on route coupling. In order to get rid of the effect of coupling between routes during data communication the notion of *zone-disjoint routes* is proposed in [10]. If *zone-disjoint routes* are used then data communication along one path will not interfere with data communication along other paths and simultaneous communication will be possible.

Let us assume that, S_1-N_1-N_2-D_1 and S_2-N_3-N_4-D_2 are two node-disjoint paths used by S_1 and S_2 to communicate with D_1 and D_2 respectively as shown in Fig. 1.

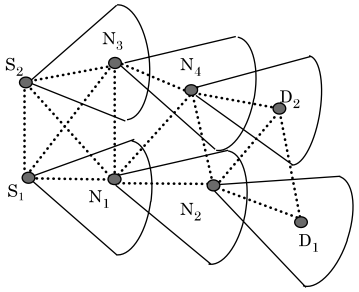


Fig. 1 Zone disjoint communication between S_1-D_1 and S_2-D_2 with directional antenna.

Since they are node-disjoint, so they are apparently contention-free paths. If S_1 and S_2 are within the omnidirectional transmission range of each other (as shown in dotted line), then they cannot communicate simultaneously. Similarly, N_1, N_3 and N_2, N_4 in Fig. 1 cannot communicate simultaneously as well. So, even if node-disjoint routes are used for communication between S_1-D_1 and S_2-D_2, the inherent route coupling among the node-disjoint routes will not allow them to communicate simultaneously and the routing performance in wireless environment degrades substantially. So, it is evident that, in order to provide priority-based QoS, effect of route coupling should be minimized in case of high priority traffic. High priority traffic should select a non-interfering path, i.e, a path free from interference caused by other low priority traffic to get a contention-free access to the medium. On the other hand, low priority packets may choose a route, which is *zone disjoint* with respect to high-priority traffic. Thus, low priority traffic *should get repelled* from the high-priority communication region as far as possible to minimize the effect of route coupling with active high-priority communications present in the network.

Our objective is to exploit the advantage of zone-disjointness and use it to calculate diverse routes for low priority flows, which will minimally interfere with zone containing high priority traffic. But, getting zone-disjoint or even partially zone disjoint paths using omni-directional antenna is difficult since transmission zone of omni-directional antenna covers all directions. Directional antenna has a reduced transmission zone width compared to omni-directional antenna. So, two interfering routes can be easily decoupled using directional antenna [10].

It has been shown earlier that the use of directional antenna would largely reduce radio interference, thereby improving the utilization of wireless medium and consequently the network throughput [10], [12], [13]. In our earlier work, we have developed a MAC and routing protocol using directional ESPAR antenna [14], [15]. In this paper, we propose to investigate the use of directional antenna in priority-based routing using zone reservation.

As discussed earlier, the probability of getting zone-disjoint paths is comparatively high using directional antenna than omni-directional antenna, which in turn improves the overall utilization of the wireless medium in the context of ad hoc wireless network. Figure 1 illustrates that it is possible to decouple two node-disjoint routes S_1-N_1-N_2-D_1 and S_2-N_3-N_4-D_2 with directional antenna, which would not be possible if omni-directional antenna were used in this case.

In this paper, we have proposed a scheme for supporting priority-based QoS in MANET by classifying the traffic flows in the network into different priority classes and giving different treatment to the flows belonging to different classes during routing, so that, high priority flows will achieve best possible throughput. The suggested protocol selects shortest path for high priority traffic and reserves the zone along the path of high-priority communication. For low priority flows, it selects zone-disjoint diverse route that will minimally interfere with high priority flows and thus reduces the effect of coupling between *high* and *low priority routes*. This scheme uses some kind of "capture" of the selected part of the wireless medium through adaptive *reservation of zone* by the high priority traffic.

The rest of the paper is organized as follows. Section 2 starts with a few definitions used in this paper. As the proposed routing protocol is proactive and requires network status information (i.e., current network topology information and active communication information) to select *maximally zone-disjoint path,* this information percolation mechanism through the network is briefly discussed in Sect. 3. Section 4 describes the concept and mechanism of selection of maximally zone disjoint routes in general. Using this notion, a priority-based service differentiation scheme for providing QoS in MANET through adaptive zone reservation for high priority traffic is presented in Sect. 5. Effectiveness of our proposal is evaluated on QualNet Network Simulator in Sect. 6 and the simulation result shows a clear and desired difference in the performance of the high priority flows in comparison to that of the low priority flows. Section 7 concludes the paper.

2. Some Important Definitions

Definition 1. When a node n forms a transmission beam at an angle α and a beam-width β with a transmission range R, the coverage area of n at an angle α is defined as transmission_zone$_n$ (α, β, R) (Fig. 2) of node n. It implies that if a node $m \in N$ is within the transmission_zone$_n$ (α, β, R) and m is in receive mode, then, whenever n transmits a message at that transmission angle α with respect to n and beam-width β and transmission range R, it will be received by m. When node m moves out of the transmission_zone$_n$ (α, β, R), the connectivity between n and m is lost. Since transmission beam-width β and transmission range R are fixed here in our study, we will refer transmission_zone$_n$ (α, β, R) as transmission_zone$_n$ (α) in subsequent discussions.

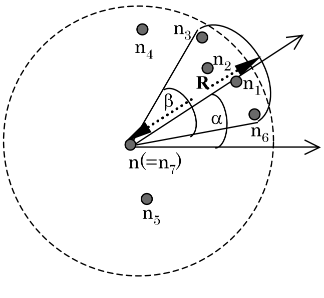


Fig. 2 Transmission zone$_n$ (α, β, R) [in solid line] and omni-directional transmission range [in dotted lines] showing directional and omni-directional neighbors.

Definition 2. High priority zone is the transmission_zone$_n$ (α) formed by any node n that is involved in high priority communication. If there is an ongoing high priority communication $n \rightarrow n_1$, then transmission_zone$_n$ (α) shown in Fig. 2 is the high priority zone.

Definition 3. We define neighbors of n $(G^n) \in N$ as a set of nodes within the omni-directional transmission range R of n. Hence, in Fig. 2, n_1 to n_6 are six neighbors of n.

Definition 4. A subset of G^n, $G^n_\alpha \in G^n$, is defined as the directional neighbors of n, where the nodes in G^n_α lie within its transmission_zone$_n$ (α). Hence only n_1, n_2, n_3, and n_6 are directional neighbors of n at an angle α, in Fig. 2.

Definition 5. Communication-id c is a unique flow-id that specifies a source-destination pair for which the communication is on.

Definition 6. Reserved-Node-List $[RNL(t)]$ is a set of nodes at an instant of time t where each node is either a sender or a receiver in any high priority communication process or a directional neighbor of any of this sender or receiver node mentioned above. Each node in the list is associated with a set C of communication-ids for which it is reserved.

Definition 7. Reserved-Directional-Neighbors of node n at transmission_zone$_n$ (α) $[RG^n_\alpha(t)]$ is a set of nodes within the transmission_zone$_n(\alpha)$ that are reserved for high priority communication at that instant of time (i.e. belongs to $RNL(t)$). So, $RG^n_\alpha(t) = G^n_\alpha(t) \cap RNL(t)$. In Fig. 2, if only n_2, n_3 and n_4 are found to be reserved nodes from $RNL(t)$, then $RG^n_\alpha(t)$ are n_2, n_3.

Definition 8. Correlation factor of node n_i in a path P for Communication-id c $[\eta^{n_i}_c(P)]$, where n_j is the next-hop from n_i in path P and $\alpha(n_i \rightarrow n_j)$ is the transmission zone formed by n_i towards n_j in order to communicate with n_j, is defined as the sum of the number of communication-ids C handled by each reserved-directional-neighbor of node n_i at transmission_zone$_{n_i}$ $(\alpha(n_i \rightarrow n_j))$ excluding the communication-id c. So, $\eta^{n_i}_c(P) = \sum_{\forall_n \in RG - n_i - \alpha(n_i \rightarrow n_j)(t)} (|C - c|)$. Informally speaking, correlation factor of a node along a specified path measures the degree of coupling of that node with respect to its directional neighbors that are reserved for high priority communication in that given communication direction. For example, if the correlation factor of node n_i in path P along transmission zone $\alpha(n_i \rightarrow n_j)$ is zero, it implies that node n_i can transmit to n_j at transmission zone $\alpha(n_i \rightarrow n_j)$ without affecting any other high priority communications.

This corresponds the case that all nodes within $\alpha(n_7 \to n_1)$, i.e., n_1, n_2, n_3, and n_6, are not reserved nodes in Fig. 2. On the other hand, if n_i has 2 reserved-directional-neighbors at transmission zone $\alpha(n_i \to n_j)$: one is reserved for 2 high priority communications and the other is reserved for another high priority communication and if one of them is having communication-id c, then $[\eta_c^{n_i}(P)]$ will be $2 + 1 - 1 = 2$. This implies that node n_i can transmit to n_j at transmission zone $\alpha(n_i \to n_j)$ but will affect other high priority communications with a degree of coupling 2. This is equivalent to the situation in Fig. 2 that n_2, n_3 are reserved-directional-neighbors at $\alpha(n_7 \to n_1)$: n_3 has 2 high priority communications and n_2 is engaged for another high priority communication and that n_3 deals communication-id c. It is to be noted that, if a reserved-directional-neighbor of a node n_i is reserved with its own communication-id c, then n_i ignores the reserve-status of that node for that communication-id c for calculating $\eta_c^{n_i}(P)$. Thus, in Fig. 2, when n_1 is reserved node and when n_1 is retained for communication-id c of n_7, then n_7 doesn't include the reserved-status of n_1 for calculating $\eta_c^{n_7}(P)$.

Definition 9. Correlation factor η of path P for Communication-id c $[\eta(P)]$ is defined as the sum of the correlation factors of all the nodes in path P. So, $\eta(P) = \sum_{\forall n_i \in P} (\eta_c^{n_i}(P))$. When $\eta(P) = 0$, any communication along path P will not disturb any high priority communication process at that instant of time. Otherwise, the path P is η—related with other high-priority paths. Correlation factor is used to measure route coupling [8], [9].

3. Network Status Information Percolation

The purpose of an information percolation mechanism is to make each node aware of the *approximate topology* and the *communication events going on in the network* [15]. The objective here is to get accurate local perception, but approximate global perception of the network information. This approximate network awareness would be helpful to implement both MAC and an adaptive routing protocol, as discussed in [14], [15]. In order to track the direction of its neighbor, each node n periodically *collects* its directional neighborhood information so that a node can determine the best possible direction of communication with any of its neighbor. Each node n in the network maintains the following two network-status information, namely Reserved Node List and Global Link State Table:

Reserved Node List (RNLn): It contains the perception of node n about high-priority communication activities in the entire network. As mentioned earlier, it is a set of nodes at an instant of time t where each node is either a sender or a receiver in any high priority communication process or a directional neighbor of any of this sender or receiver node. Each node in the list is associated with a set C of communication-ids for which it is reserved.

Formation of RNL: Each node in the network maintains a timer and sets it with a predefined time interval. If a node starts a high priority communication (n in Fig. 2), or receives a packet of high priority flow as the intended receiver (n_1 in Fig. 2), or just overhears a high priority data transmission being within a high priority transmission zone (n_2, n_3 and n_6 in Fig. 2) then, it starts its timer and set itself as "Reserved." This information is recorded in the RNL of that reserved node. If the above situation occurs again before the timer expires then the timer will be set again with the same predefined time interval. If the timer expires before the occurrence of the above situation then the node sets itself as "Unreserved" and removes its entry from RNL. Each node in the network periodically broadcasts this Reserved Node List. On receiving periodic RNL from different nodes, each node combines them to form revised RNL and waits for a periodic interval to broadcast it to its neighbors.

At each node, RNL first gets updated by Neighborhood Reserved Node List of that node. So, initially when the network commences, all the nodes are just aware of the reservation status of their own neighbors and are in a *don't-know-state* regarding the other nodes in the system. Periodically, each node broadcasts its RNL as update to its neighbors. With this periodic update messages from its neighbors about their neighbors, the nodes slowly get reservation status information about the other nodes in the network. Thus, each node updates its own RNL based on received update messages from other nodes.

A major aspect underlying the infiltration of network status information into mobile nodes is that the information carried must be recognized with a degree of correctness. Since the propagation of updates from different nodes is asynchronous, it becomes imperative to introduce a concept of recency of information [16], [17]. For example, let us assume two RNL packets A_1 and A_2 arrive at node n, both of them carrying information about node m which is multihop away from node n. In order to update the information at node n about node m, there has to be a mechanism to find out who carries the most recent information about node m: A_1 or A_2?

To implement this, we have used the concept of recency and a mechanism to increment it appropriately. If two update messages have a set of data concerning the same node, say node n, then the update message carrying the higher recency value of node n has more current information about it. The structure of RNL at a node n is given in Table 1.

Table 1 The structure of RNL.

Nodes	n_1	n_2	…	n_N
Recency	R_1	R_2	…	R_N
State	S_1	S_2	…	S_N
Communication-id	C_{11}	C_{21}	…	C_{N1}
	C_{12}	C_{22}	…	C_{N2}
	C_{13}	C_{23}	…	C_{N3}

Here, R_i is the recency of node n_i in a network of N nodes and S_i denotes the corresponding reservation status of each node, which can be either 0 (not-reserved) or any positive value (reserved). Any value of S_i greater than 0 indicates the node to be reserved for S_i number of communication. C_{ij} denotes the communication-id of j—the communication for which i-th node is active. Only first three communication-ids are propagated for each reserved node. If the node i is inactive, $C_{ij} \forall j,\ j \in \{1 \ldots 3\}$ is null.

Global Link State Table (GLST) and its Formation: Each node maintains a Global Link State Table (GLST) to capture network connectivity information. Initially when a network commences, all the nodes are just aware of their own neighbors and are in a *don't-know-state* regarding the other nodes in the system. Periodically, each node broadcasts its GLST as update to its neighbors. With this periodic update messages from its neighbors about their neighbors, the nodes slowly get information about the other nodes and their neighbors. Thus, each node updates its own GLST based on received update messages from other nodes. It is to be noted that by controlling the periodicity of updates, it is possible to control the update-traffic in the network and the accuracy of network status information stored in each of the node. For example, if the propagation of update messages is too frequent, the control traffic will increase but the accuracy of network status information stored in each node will also be better. However, the network would never get flooded with propagation of updates. The maximum number of update packets in the network at any point of time is always less than the number of nodes in the network. In this case also, we need to implement the concept of recency as explained in the context of RNL propagation. This implies that if two GLST update messages have a set of data concerning the same node, say node *n*, then the update message carrying the higher recency token value of node *n* has more current information about it.

As and when a node *n* receives GLST from other nodes, it updates its GLST. In order to do that, the recency tokens of all the nodes stored in the GLST of n and the recency tokens of all the nodes stored in the recently arrived update packet are compared. If the recency token of any node, say *X*, in GLST of *n* happens to be less than that in the update packet, then it is obvious that the update packet is carrying more recent information about node *X*. So, the entire information about node *X* in the GLST of node *n* is overwritten by the received information of *X* in the update packet. This step is performed asynchronously for all the update packets as they arrive at that host node *n*. This step helps the node *n* to acquire all the recent information that it can gather from the update packets.

It is to be noted that the mechanism does not guarantee that each node would know the exact status of the network. It is merely an *awareness* that helps each node to figure out the approximate status of the network. It helps to maintain accurate status information about the immediate neighborhood of a node, with progressively less accurate details as the distance increases. The structure of GLST at any node *n* is given in Table 2.

Table 2 The structure of GLST.

Nodes	***Recency***	***Neighbors***
n_1	R_1	→ {…………}
n_2	R_2	→ {…………}
…	…	→ {…………}
n_i	R_i	$\rightarrow\{<n_j,\alpha(n_i,n_j)><n_k,\alpha(n_i,n_k)>\ldots\}$
…	…	→ {…………}
n_N	R_N	→ {…………}

Here, R_i is the recency of node n_i in a network of N nodes and $\langle n_j,\ \alpha(n_i, n_j)\rangle$ denotes that n_j is a neighbor of n_i where $\alpha(n_i, n_j)$ indicates the transmission beam-angle α at which n_i can best communicate with n_j.

So, it is evident from the above discussion that RNL and GLST can be maintained easily by using some simple timers with each node.

Using this RNL and GLST, a node calculates route avoiding the zones containing reserved nodes as far as possible. Thus, it seems that all the nodes in the RNL have reserved a part of the network, which is referred as *high priority zone*. Other low priority traffic are not allowed to use that zone. So they are selecting a suitable diverse route to achieve best effort QoS. A new high priority flow will consult the existing RNL to select a route that is maximally zone-disjoint with respect to other on-going high priority communication, as will be illustrated later.

4. Maximally Zone-Disjoint Routing

We propose to use two metrics in our route selection criteria: correlation factor and propagated hop count. As explained earlier, correlation factor of a route is inversely related to zone disjointness of that route with respect to other active routes. At the same time, hop count of the selected diverse route is also another concern in this context. Otherwise, under some communication scenario, it may so happen that, for a particular destination, each intermediate node tries to select a route avoiding the existing communication zones and ultimately ends up traversing the entire network in search of a zone-disjoint route. So, minimization of both correlation factor and propagated hop count will give rise to maximally zone disjoint shortest path. Each node in the network uses its current network status information (approximate topology information and ongoing high priority communication information) to calculate the *suitable next hop* for reaching a specified destination such that the interference with the nodes already busy in some communication gets minimized.

Initially, when a packet is transmitted from a source it gives preference to the zone-disjoint path selection criteria. But, if a packet reaches an intermediate node after traversing multiple hops then *progressively shorter hop route* towards the destination will be selected. So this adaptive route calculation mechanism guarantees the convergence of the proposed routing algorithm. We have used the following function to calculate the link-weight that will ensure the selection of lower η path for low propagated hop count and selection of lower hop path for higher propagated hop count. Dijkstra's shortest path algorithm has been modified to select a path having smallest link-weight, i.e., total link-weight of all the links on that selected path will be minimum.

Link-cost (n_i, n_j) *during the current communication having Communication Id* $c = \alpha + \eta_c^{n_i} + \gamma H$ where,

α = Initial link-weight (.01 in our case; $\alpha \ll \eta_c$ and $\alpha \ll H$, as will be explained in the following section)

$\eta_c^{n_i}$ = The sum of the total number of communications (excepting the current communication c) handled by each reserved directional neighbor in the directional zone $(n_i \rightarrow n_j)$ (as explained in Sect. 2).

H = propagated hop-count of the current packet for which route is being calculated.

γ =diversity factor.

When H and η is zero, α is the initial link-weight assigned to each link to find out the shortest path. γ is to be adjusted in such a way that a low priority flow always selects a diverse path. But at the same time, progressively shorter hop route should get preference over η-driven route to ensure convergence. Otherwise, low priority throughput will be unacceptable. For high priority flow, more emphasis is given on hop count rather than η. This will be illustrated in the next section.

5. Priority-Based Service Differentiation through High-Priority Zone Reservation

Our protocol basically assigns a path to a high priority flow that is shortest as well as maximally zone-disjoint with respect to other high-priority communications. Low priority flows will take an adaptive zone-disjoint path avoiding all *high priority zones,* reserved by on-going high priority communications. For low priority flows, a shortest path criterion is not a predominant metric. Thus, our proposal ensures the fact that even if several low priority communications are present in the network, the high priority traffic will get least disturbed by the low priority traffic. As a result, in a network having multiple flows with different priorities, the high priority throughput will remain almost same as if the network is handling only high priority flows. If more than one high priority flows are present in the network, then each of them will try to take a path which will maximally zone disjoint with respect to other high priority flows.

Each flow in the network is identified with a unique id and belongs to either HIGH or LOW priority category. Whenever a packet sent by high priority flow comes to a node for a particular destination, the node simply selects the *lowest-cost path* towards that destination and transmits the packet to the immediate next-hop on the selected path.

The *lowest-cost path for each priority flow* is calculated as follows, according to the formula shown in Sect. 4:

- Each link $(n_i,\ n_j)$ in the network is initialized with a constant value α.
- If any pair of nodes n_i and n_j are involved in some high priority communication, then all the nodes in the directional transmission zone of the sender n_i towards receiver n_j will set their activity status as HIGH to indicate that they should sit idle to support high priority transmission. They are treated as reserved nodes and are updated in RNL. RNL of a node is transmitted periodically to all other nodes so that they can update their RNLs.
- Link-weight of each link connected to a reserved node (nodes belong to RNL) is set to a value based on the formula described above. Accordingly, each high priority flow will try to take a path which will maximally zone disjoint with respect to other high priority flows. And when a source of a low priority flow calculates its route, our path selection algorithm will automatically avoid the links with high values during path selection process. So, in this way, the entire zone around the path taken by high priority flows gets avoided.
- If a reserved node, with HIGH activity status does not receive high priority packet for a considerable period of time then it will set itself as *unreserved* so that other communications may select paths through it, if required.

This is a kind of adaptive zone reservation and release process used by the high priority nodes during route selection that allows us to achieve desired throughput for the different priority flows. Diversity factor γ is taken as 0.5 in our high priority case and 0.2 for low priority. Low value of diversity factor implies that the selected route will be longer than route selected with high value of diversity factor. That means, low priority traffic will select a longer but diverse route to avoid high priority zone where as high priority traffic will select shorter diverse route with respect to other high priority routes to reduce interference among multiple high priority flows.

6. Performance Evaluation

The proposed routing protocol is implemented on QualNet simulator using ESPAR antenna with 12 overlapping patterns at 30 degree intervals [14], [15] as our directional antenna pattern to prove the effectiveness of our proposal. The simulation environment specifications and parameters used are described in Table 3.

We have selected six random source-destination pairs (Flow1 to Flow6). We have considered the following scenarios and initially observed the throughput of Flow1 (Fig. 3) when: (i) Only Flow 1 is present in the network and selects a shortest path (shown in Fig. 3 as "Flow1 As Single Flow"); (ii) Flow1 is communicating in presence of 5 other Flows (Flow2–Flow6) in the network and no priority scheme is used. Thus, all of them select shortest paths that may be

Table 3 Parameters used in simulation.

Parameters	***Value***
Area	1500 x 1500 m
Number of nodes	100
Transmission Power	10 dBm
Receiving Threshold	-81.0 dBm
Sensing Threshold	-81.0 dBm
Data Rate	2Mbps
Packet Size	512 bytes
CBR Packet Arrival Interval	5 ms
No. of simultaneous comm.	6
Simulation Time	5 minutes
Mobility Model	Random Way-point
Topology	Random
RNL Periodicity	500msec
GLST Periodicity	10sec

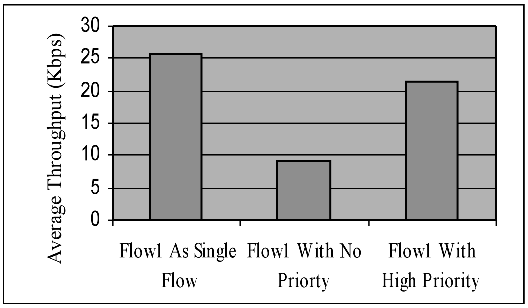


Fig. 3 Behavior of a particular flow (Flow1) with different priority assignments in a scenario of 6 communications.

interfering with each other (shown in Fig. 3 as "Flow1 With No Priority"); Finally in (iii) A priority-based service differentiation scheme is employed. Flow1 is assigned *high priority*, thus takes the shortest path. Moreover, Flow1 reserves a directional *zone* around each node on its route so that 5 other low priority flows will eventually select adaptive paths avoiding the zone reserved by Flow1. The throughput of Flow1 in this scenario is shown in Fig. 3 as "Flow1 With High Priority."

It is observed that, in the first case, throughput of Flow1 is maximum, which is an obvious outcome of the fact that no other flow is causing any disturbance to it. In second case, as all the flows are using shortest path, existence of route coupling among those routes reduces the throughput of Flow1 drastically. But, in the third case, as soon as high priority is assigned to Flow1 and routes are selected according to our protocol, throughput of Flow1 shows a remarkable improvement, which is almost same as the throughput in the first case.

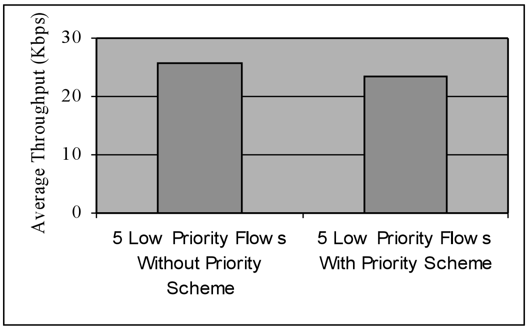


Fig. 4 Behavior of 5 low priority flows with and without assigning priority to one flow (Flow1) in a scenario of 6 communications.

We have also observed the average throughput of the 5 low priority flows (i.e., Flow2–Flow6) under the situation described above. Figure 4 illustrates that if high priority is assigned to flow1 then average throughput of 5 low priority communications reduces a little bit in comparison to the corresponding average throughput when no priority is assigned to Flow1. The low priority flows avoid high priority zones and selects route away from this zone. So, low priority flows may have to travel larger hops to reach destination, but will experience less interference. This explains the small reduction in throughput of average low priority flow in Fig. 4.

It is expected that each flow will behave in similar fashion, if no priority scheme is applied. So, average throughput of 5 low priority flows (Flow2 to Flow6), captioned as "5 Low Priority Flows Without Priority Scheme" in Fig. 4, should be approximately equal to that of Flow1 when no priority has been assigned among 6 flows, captioned as "Flow1 With No Priority" in Fig. 3. But Figs. 3 and 4 are not reflecting this fact due to following reasons. We have intentionally taken a "coupled" flow as Flow1 to show the performance improvement with priority scheme. Our intention is to establish the effectiveness of the proposed priority based zone-reservation scheme by decoupling the coupled flow (Flow1) and reserving a zone around the flow (Flow1) which in turn will improve the throughput performance of the flow (Flow1). On the other hand, all other 5 flows i.e., Flow2 to Flow6 may not be coupled with each other since source-destination pairs are randomly chosen. So, there are one or more flows amongst Flow2 to Flow6, which are not facing any kind of interference. Those uncoupled flows among Flow2 to Flow6 are producing higher throughput than the flows coupled with each other. Such flows are responsible for resultant higher average throughput of low priority flows (flow2 to 6) than flow1 when no priority scheme is adopted. The main objective of this scheme is to improve high priority throughput with minimally disturbing the low priority throughput, which is depicted in Fig. 4. Earlier without any priority scheme, flow1 was producing approx. 10 kbps, which was approximately doubled if priority scheme is adopted (approx. 20 kbps.). Flow1 may not yield better throughput than the average throughput of

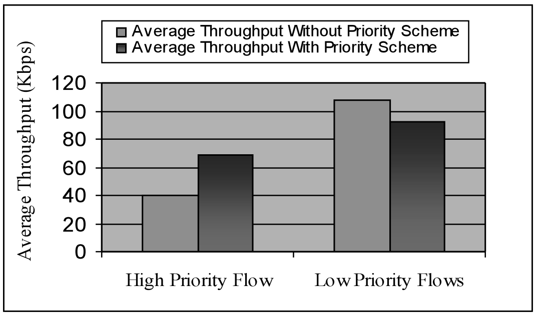


Fig. 5 Comparison of average throughput of high and low priority flows under high mobility (0–15 mps) in a scenario of 6 communications.

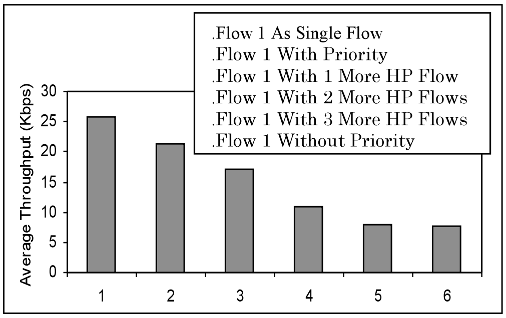


Fig. 6 Behavior of Flow1 with increasing number of high priority communications in a scenario of 6 communications.

the low priority flows in some scenarios due to the reason described above.

The priority scheme is also studied under a high mobility (0–15 meters/sec for each node) to find out whether the performance improvement can still be guaranteed. Figure 5 shows that, in a scenario of 6 communications, if Flow1 is assigned high priority, it performs better than the case when no priority assignment is made to Flow1. Similarly, in case of low priority traffic the result shows a small degradation in average throughput of 5 low-priority flows (Flow2–Flow6) if priority scheme is adopted.

Lastly, we have studied the capacity of the network in terms of handling high priority traffic, i.e, how many high priority flows may be efficiently handled by the network? In order to examine that, we have progressively increased the number of high priority communications from 1 to 4 in a network handling a total of 6 communications. The throughput of a particular high priority communication (Flow1) shows progressive reduction in improvement as we increase number of high priority flows (Fig. 6) and if 4 high priority communications are allowed in the network, result shows that average throughput degrades even if priority is assigned. This result clearly indicates that with more number of flows, unless we block the low priority flows, it will be forced to go through high priority reserved zones and we will loose the advantage of zone reservation.

Number of high priority communications that can be efficiently handled by a network basically depends on the following factors—(i) the transmission beam-width β and the transmission range R of the directional beam *(ref. Definition 1)* and (ii) the number of simultaneous high and low priority communications in the network.

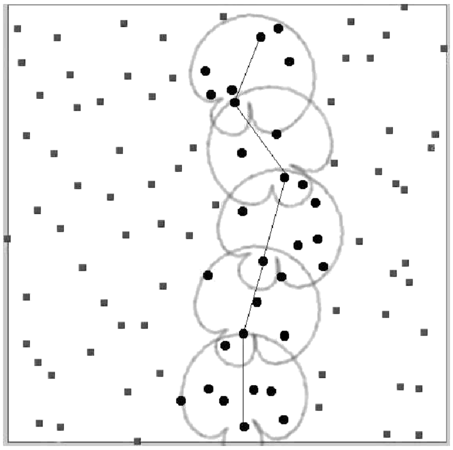


Fig. 7 A snapshot showing zone-reservation by a single high priority communication [Reserved nodes: •].

A zone reserved by a high priority flow in each hop along its route essentially depends on the beam-width β and the transmission range R of a directional beam. So, if we assume that the average number of hops taken by each high priority flow is H, then, the maximum number of non-overlapping high priority zones that can be formed in a given area A is equal to $2 \cdot A/\beta \cdot R^2 \cdot H$.

Taking $\beta = 60°$, $R = 300$ mts. (for ESPAR Antenna), average number of hops taken by high priority flows i.e., $H = 5$ and area $A = 1500 \times 1500$ sq. mts., maximum number of non-overlapping high-priority zones theoretically possible = 10 (approx.). But, due to the presence of side-lobes and irregular shape of the directional beam as shown in Fig. 7, the zone actually covered by ESPAR antenna in each hop is much wider than the area covered by a 60° geometric beam. Figure 7 shows zone reservation by a typical high priority in a random topology of 100 nodes in 1500 × 1500 sq. mt. area using ESPAR antenna pattern. As a result, it is not possible to accommodate 10 non-overlapping high priority zones (as obtained by the above formula) with such a wide beam pattern. This explains why maximum 3 high priority transmissions can be efficiently accommodated in the network on top of other low priority flows in Fig. 6. By reducing the beam-width and the transmission range of the directional beam, more number of high-priority zones can be accommodated in a given area. Otherwise, with more number of high priority communications, overlapping zones will be created which will interfere with each other and the performance of priority-based scheme will suffer.

Moreover, as the number of high priority flows increase in the network, it becomes difficult for the low priority flows to find routes avoiding high-priority zones. As a result, low priority flows will be forced to take route through high priority zone, causing interference. This may be controlled by temporarily blocking such low priority flows in the system, but we have not implemented such call blocking.

Additionally, throughput of multiple high priority flows may suffer since they actually take maximally zone-disjoint path with respect to each other instead of shortest path. So, end-to-end delay for high priority flows may increase, causing degradation in high priority throughput.

7. Conclusion

In this paper, we have suggested a zone-reservation based mechanism towards prioritised routing with the objective of providing a interference-free communication to high priority traffic. However, with the current beam pattern of ESPAR antenna, it is not possible to accommodate multiple number of non-overlapping high priority zones. So by controlling the beam width and the transmission power of directional antenna, it is possible to increase the number of high priority flows in the network.

Acknowledgments

This research was supported in part by the Telecommunications Advancement Organization of Japan.

References

[1] K. Wu and J. Harms, "QoS support in mobile ad hoc networks," Crossing Boundaries—An Interdisciplinary Journal, vol.1, no.1, Fall 2001.

[2] S. Chakrabarti and A. Mishra, "QoS issues in ad hoc wireless networks," IEEE Commun. Mag., vol.39, no.2, pp.142–148, Feb. 2001.

[3] Z.-Y. Demetrios, "A Glance at Quality of Service in mobile Ad Hoc Networks," (http://www.cs.ucr.edu/˜csyiazti/cs260.html) Final Research Report for cs260—Seminar on Mobile Ad Hoc Networks, Fall 2001.

[4] S. Chen and K. Nahrstedt, "Distributed quality-of-service routing in ad hoc networks," IEEE J. Sel. Areas Commun., vol.17, no.8, pp.1488–1505, Aug. 1999.

[5] H. Xiao, K. Chua, W. Seah, and A. Lo, "A flexible quality of service model for mobile ad-hoc networks," Proc. IEEE VTC 2000-Spring, Tokyo, Japan, May 2000.

[6] P. Mohapatra, J. Li, and C. Gui, "QoS in mobile ad hoc networks," Special Issue on QoS in Next-Generation Wireless Multimedia Communications Systems in IEEE Wireless Commun. Mag., June 2003.

[7] X. Pallot and L.E. Miller, "Implementing message priority policies over an 802.11 based mobile ad hoc network," Proc. MILCOM 2001, Washington, Oct. 2001.

[8] K. Wu and J. Harms, "On-demand multipath routing for mobile ad hoc networks," EPMCC 2001, Vienna, Feb. 2001.

[9] M.R. Pearlman, Z.J. Haas, P. Sholander, and S.S. Tabrizi, "On the impact of alternate path routing for load balancing in mobile ad hoc networks," MobiHOC 2000, pp.3–10, Aug. 2000.

[10] S. Bandyopadhyay, S. Roy, T. Ueda, and K. Hasuike, "Multipath routing in ad hoc networks with directional antenna," Proc. IFIP TC6/WG6.8 Conference on Personal Wireless Communications (PWC 2002), Singapore, Oct. 2002.

[11] S.-S. Kang and M.W. Mutka, "Provisioning service differentiation in ad hoc networks by the modification of backoff algorithm," International Conference on Computer Communication and Network (ICCCN) 2001, Scottsdale, Arizona, Oct. 2001.

[12] R.R. Choudhury, X. Yang, N.H. Vaidya, and R. Ramanathan, "Media access control for ad hoc networks: Using directional antennas for medium access control in ad hoc networks," Proc. Eighth Annual International Conference on Mobile Computing and Networking, Sept. 2002.

[13] R. Ramanathan, "On the performance of ad hoc networks with beamforming antennas," ACM MobiHoc, Oct. 2001.

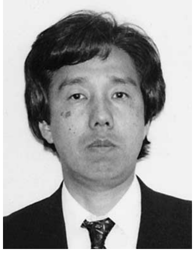
[14] T. Ueda, S. Tanaka, D. Saha, S. Roy, and S. Bandyopadhyay, "A rotational sector-based, receiver-oriented mechanism for location tracking and medium access control in ad hoc networks using directional antenna," Proc. IFIP Conference on Personal Wireless Communications PWC 2003, Venice, Italy, Sept. 2003.

[15] S. Roy, D. Saha, S. Bandyopadhyay, T. Ueda, and S. Tanaka, "A network-aware MAC and routing protocol for effective load balancing in ad hoc wireless networks with directional antenna," ACM MobiHoc,2003, Maryland, USA, June 2003.

[16] R.R. Choudhury, S. Bandyopadhyay, and K. Paul, "A distributed mechanism for topology discovery in ad hoc wireless networks using mobile agents," Proc. First Annual Workshop on Mobile Ad Hoc Networking & Computing (MOBIHOC 2000), Boston, Massachusetts, USA, Aug. 2000.

[17] R.R. Choudhury, K. Paul, and S. Bandyopadhyay, "An agent-based connection management protocol for ad hoc wireless networks," Journal of Network and System Management, Dec. 2002.

Tetsuro Ueda received his B.E. degree in electrical and communications engineering and M.E. degree from Tohoku University in 1986 and 1988, and joined NEC. He has researched channel allocation schemes and worked on the IMT-2000 standardization. He jointed KDDI in 1997, and moved to ATR Adaptive Communications Research Laboratories in 2001. His research interest is wireless ad hoc network.



Shinsuke Tanaka received his B.E. degree in material engineering and M.E. degree from Kyoto University in 1978 and 1980, and joined KDD. He has researched magneto-optical recording medium and semiconductor optical devices for optical fiber communications, and received his Ph.D degree in 1988. In 2002, he was appointed manager of the First Research Department, ATR Adaptive Communications Research Laboratories. His research interests are wireless ad hoc networks and communication quality.

Siuli Roy is a MCA (Master of Computer Aplipcations) from Jadavpur University and currently working as Senior Research Engineer at Indian Institute of Management Calcutta in ADHOCNET Project. She is currently working towards her PhD Degree on MAC and Routing Protocols for Ad Hoc Wireless Networks using Smart Antennas.

Dola Saha is a B. Tech in Information Technology from Netaji Subhas College of Engineering and currently working as Research Engineer at Indian Institute of Management Calcutta in ADHOCNET Project. She is currently working towards her Ph.D. Degree on priority based service provisioning in Ad Hoc Wireless Networks using Smart Antennas.

Somprakash Bandyopadhyay is a Ph.D. in Computer Science from Jadavpur Universityand B.Tech in Electronics and Electrical Communication Engineering from Indian Institute of Technology, Kharagpur (1979). He is currently Associate Professor of MIS and Computer Science Group of Indian Institute of Management, Calcutta. He is also acting as the Project Director of ADHOCNET Project at Indian Institute of Management Calcutta.