QoS Routing Performance in Multihop, Multimedia, Wireless Networks *

T.-W. Chen, J. T. Tsai and M. Gerla[†]
Computer Science Department
University of California, Los Angeles
Los Angeles, CA 90024

Abstract

In this paper, we propose an approach to QoS (Quality of Service) routing in a multimedia, multihop, wireless network. The wireless net can be either stand alone, or connected to the wired net. The main focus of the paper is the QoS routing procedure which can inform the source of the bandwidth and quality of service available to any destination in the wireless network. This knowledge enables the establishment of QoS connections within the wireless network and the efficient support of real time, multimedia traffic. In addition, it enables more effective call acceptance control. In the case of ATM interconnection, QoS information permits to extend the ATM virtual circuit service to the wireless network, with possible renegotiation of QoS parameters at the gateway. Simulation experiments show the efficiency of QoS routing in selected multihop, mobile radio network scenarios.

1 Introduction

Currently, most of the existing wireless network proposals and prototypes are based on the single hop, cellular model supported by a wired infrastructure [1, 2, 3, 4, 5]. The main design challenges in this environment are the support of QoS connections in the wireless segment and the graceful handoff of mobile hosts from one base station to the next. Consequently, most of the research has been directed to: (1) multimedia traffic integration within the wireless cell, and; (2) timely rerouting of backbone virtual circuits from one cell to next with minimal disruption of service [6, 7].

In addition to the traditional, cellular based applications, a number of new applications have emerged recently, which require wireless "multihopping" between remote users, without relying on the fixed, wired infrastructure. These applications include disaster (flood, fire, earthquake) recovery, search and rescue and ad hoc networking among nomadic users. These multihop wireless networks can be stand alone, or may be connected to a wired infrastructure, which is typically a few hops away. For example, in a search and rescue operation, one or more base stations, mounted on vehicles, may be connected to ATM via point to point satellite access links. The members of the rescue team communicate with each other and with the base station via multihopping.

The problem of interconnecting to the wired backbone acquires a new dimension when the wireless segment is multihop. Namely, one must guarantee the Quality of Service (QoS) not only over a single hop, but over an entire wireless multihop path. In turns, this requires the propagation of QoS information within the network. Key to the support of QoS reporting is QoS routing, which provides path QoS information at each source. Prior research efforts in multihop radio systems have not fully addressed this problem. For example, the ARPA supported Packet Radio project [8] has provided an efficient solution to datagram support, but has not adequately solved the support of real time traffic with Quality of Service.

In this paper, we address the problem of supporting multimedia, multihop, mobile communications in a wireless environment using QoS routing. In section 2, we review the requirements of multihop, wireless networking, and propose a cluster based approach. In section 3, we introduce a routing scheme, derived from DSDV (Destination Sequenced Distance Vector), which provides QoS information. In section 4, we demonstrate the efficiency of QoS routing in several representative simulation experiments. Section 5 concludes the paper.

2 Multihop, Multimedia, Wireless Networking

2.1 Requirements

The following requirements can be identified for a multihop, wireless network which support multimedia traffic and/or is interconnected to a wired backbone (such as ATM) offering quality of service guarantee:

- **VC, bandwidth reservation:** The wireless network must implement VCs in order to support real time connections and allocate bandwidth to them at call setup time. VC bandwidth reservation in a mobile network is a major challenge, since the path changes dynamically.

- **QoS routing:** Traditional distance vector (e.g. Bellman-Ford) routing algorithms are not adequate. To support QoS for real time traffic we need to know not only the minimum delay path to destination, but also the bandwidth available on it. For example, at call setup time, a VC should be accepted only if there is bandwidth available. Otherwise, it would disrupt the existing VCs. Routing with QoS indication is thus needed in order to efficiently manage bandwidth resources.

- **Congestion control:** Even if we know how to manage the acceptance of the VCs at call setup time using QoS routing, we can experience network congestion due to the dynamics of mobility and of traffic patterns. Some of the concepts suc-

cessfully used in ATM, i.e. selective packet dropping and input rate control, can be applied here.

- **Mobility:** The inter-working of mobility and VC maintenance is a critical issue, especially when the wireless network interconnects to an wired backbone.

2.2 Proposed Approach

A multihop, dynamically-reconfigured multi-cluster network for supporting multimedia traffic has been presented in [9]. Spread-Spectrum radios are utilized to permit code division multiple access (CDMA). The transmit power is controlled by a distributed power control algorithm DPC-ALP [10, 11] to reduce power interference and to increase channel capacity.

The distributed clustering algorithm [9] partitions the multihop network into *clusters* (see Fig. 1). Neighboring clusters use
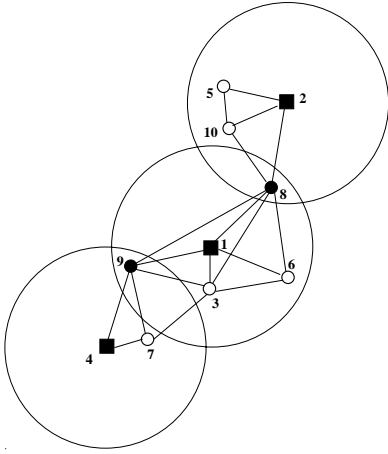

Figure 1: Example of clustering

different spreading codes in order to reduce interference and to enhance spatial reuse of channels. Within each cluster, the MAC (Medium Access Control) layer is implemented using a time slotted frame. Namely, the transmission time scale is organized in frames, each containing a fixed number of time slots. The entire network is synchronized on a frame and slot basis using an elastic synchronization scheme [12]. The frame is divided into two phases, namely, control phase and information phase. The control phase, see Fig. 2 supports the communications needs of all
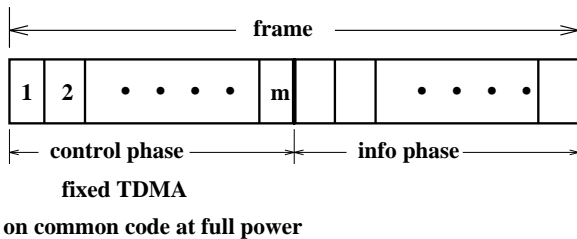

Figure 2: Channel Access Frame

control functions, such as slot and frame synchronization, clustering, routing, power measurement and control code assignment,

VC reservation, ACKs, etc. Each real time connection is assigned to a VC. The VC is an end to end path along which slots have been reserved. The number of slots per frame assigned to a VC is determined by the bandwidth required by the VC. The path and slots of a VC may change dynamically during the lifetime of a connection.

The information phase supports both virtual circuit and datagram traffic. Since real time traffic (which is carried on a VC) needs guaranteed bandwidth during its active period, slots must be allocated to the VC at call setup time. The remaining slots (free slots) can be accessed by datagram traffic using a random access scheme, for example, slotted ALOHA. While VC packets are not ACKed nor retransmitted (because of real time constraints), datagrams are retransmitted if an ACK is not received.

The conventional VC setup scheme is not suitable for a mobile, wireless network. The time required to set up a new VC (i.e., acquire bandwidth before data transmission) may be comparable to the interval between path changes. To handle mobility, we have developed a VC setup based on fast reservations. In the fast reservation scheme, each packet in the VC stream is routed individually, based on its destination address. The first packet in the VC stream, upon successfully capturing a slot in the info subframe, will reserve it for all subsequent frames. If the slot remains unused for a certain number of frames, it is declared free by the cluster head and is returned to the free slot pool. This scheme, which was inspired by the PRMA protocol [13], allows the VC stream to dynamically select a new path to the destination when the old path fails. Our solution is more akin to the "soft state" connection and flow ID concept proposed in RSVP [14]. Of course, each path change will cause some disruption (i.e., possible out of sequencing, delay to acquire a slot on the new path, lack of free slots, etc.). However, the disruption is much less severe than if a new path had to be entirely reconstructed end to end each time.

The disruption caused by rerouting is mitigated by exploiting the rate-adaptive, hierarchically encoded voice/video compression schemes presented in [15]. Namely, "least significant" packets within a hierarchically encoded video or voice stream are dropped during rerouting. This will alleviate congestion and reduce reservation delay if bandwidth on the new path is scarce, albeit at the expense of temporary signal quality degradation.

QoS can be guaranteed to already established VCs only if strict access control is exercised on incoming calls. Note that the conventional approach (used in the wired network) of probing various alternate paths until a feasible one is found is not acceptable here, especially if fast VC reservations are used. Call setup packets cause much more overhead in the wireless network than the call setup signalling does in, for example, a wired ATM network. Thus, QoS information is needed at the origin, in order to exercise effective access control. In the next section, we show how QoS information can be conveyed by the routing algorithms.

3 DSDV QoS Routing and VC Management

In the first stage of developing QoS routing algorithms, we consider only "bandwidth" as the quality of service (thus omitting Signal to Interference Ratio (SIR), packet loss rate, etc). This is because bandwidth guarantee is the most critical requirement for real time application. "Bandwidth" in a time-slotted network sys-

tem is measured in terms of number of "free" slots. "Free" here is opposed to "reserved" in our clustering VC approach described in the previous section.

Usually, in a wireless network, the transmitter and receiver will not transmit and receive at the same time because they use the same frequency band. With this constraint, the computation of available bandwidth along a path is non-trivial. In fact, it is an NP-complete problem [16]. Let us illustrate this by an example. Referring to Fig. 3, let us consider the path (a, b, c, d) with free
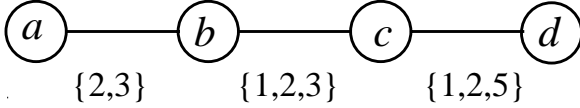

Figure 3: Bandwidth computation example

slot pools $\{2,3\},\{1,2,3\},\{1,2,5\}$ on links ab, bc, cd, respectively. The numbers refer to the free data slot positions in the TDMA frame. The size of the set means the available bandwidth on the link. Let $BW(\gamma)$ denote bandwidth on the link or path γ. I.e. $BW(link\ ab) = 2$, $BW(link\ bc) = 3$, $BW(link\ cd) = 3$.

It will be easily noticed that the available bandwidth along the path from a to d is not equal to the minimum of all link bandwidths, i.e.:

$$BW(path\ ad) \neq min\ (BW(link\ ab),\ BW(link\ bc),\ BW(link\ cd))$$

Also,

$$BW(path\ ad) \neq min\ (BW(link\ ab),\ BW(path\ bd))$$

In fact, $BW(path\ ad) = 1$, and $BW(path\ bd) = 2$. This is because the intermediate nodes along the path must spend part of the time receiving and part of the time transmitting. For example, if the same free slots are common to both upstream and downstream links along the path, the bandwidth is only one half of the free slots.

In general, to compute the available bandwidth for a path in a time-slotted network, one not only needs to know the available bandwidth on the links along the path, but also needs to determine the scheduling of free slots. 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 [16].

Fortunately, in the clustering infrastructure, the computation of available bandwidth is easier than for the general case. This is because the "gateway" nodes which connect neighboring clusters must switch to different codes in different time slots. The code scheduling of time slots for gateway nodes automatically excludes the probability of having common free slots on both upstream and downstream links. As an example, in Fig. 3, let us assume that node b is a gateway. Upon receiving the QoS update message from node b, node a can easily compute the bandwidth as follows:

$$BW(path\ ad) = min\ (BW(link\ ab),\ BW(path\ bd))$$

where nodes a and b are in the same cluster. The computation of available bandwidth on the link is simple, since the scheduling of time slots in a cluster is controlled by the "clusterhead" [9].

The available bandwidth computation is carried out independently at each node and is piggybacked on the routing algorithm. In the cluster/TDMA architecture, routing is based on a loop-free, destination sequence distance vector (DSDV) scheme [17]. As well known, a distance vector routing scheme, such as Bellman-Ford, is subject to looping and counting-to-infinite problems. DSDV provides loop-freedom while making minimum modifications to the B-F routing structure. DSDV works in the following way. Each routing table entry carries hop distance and next hop for all available destinations (as in B-F). In addition, each entry is tagged with a sequence number which originates from the destination station. The routing information is advertised by broadcasting periodically and incrementally. Upon receiving the routing information, routes with more recent sequence numbers are preferred as the basis for making forwarding decisions. Of the paths with the same sequence number, those with the shortest hop distance will be used. That information (i.e. next hop and hop distance) is entered in the routing table, along with the associated sequence number tag.

When the link to the next hop has failed, any route through that next hop is immediately assigned an ∞ hop distance and its sequence number is updated. When a node receives a broadcast with an ∞ metric, and it has a more recent sequence number to that destination, it triggers a route update broadcast to disseminate the important news about that destination. In [17], the DSDV protocol is proven to guarantee loop-free paths to each destination at all instants.

4 Simulation Results

QoS routing performance was evaluated using Maisie, a C based, parallel simulation language, developed at UCLA [18]. The simulated environment consists of 20 mobile stations randomly roaming in a 1000×1000 feet square. Two mobiles can communicate with each other if their distance is within 500 feet. Fig. 4 shows a snapshot of the simulator graphical interface, displaying clusters and connectivity. The routing update period is
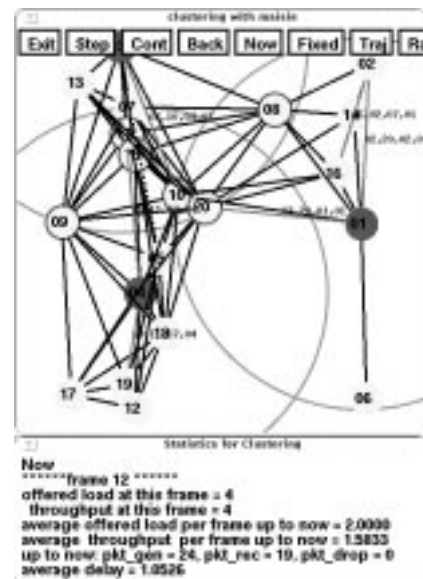

Figure 4: Maisie simulator snapshot

$100\ ms$. After every routing update period, a mobile moves x feet with probability P. We consider two cases: $(x = 2, P = 0.5)$ and $(x = 10, P = 0.5)$, with average speeds 11 km/hr and 55 km/hr, respectively. In this section, we evaluate the performance of this algorithm when implemented in the Cluster TDMA scheme. As a term of reference, we also evaluate the performance of the Packet Radio network.

4.1 VC Traffic Performance

We assume that VC calls between arbitrary source/destination pairs are generated according to a Poisson distribution. Call duration is exponentially distributed. We monitor total packet throughput versus offered traffic load as stations move at various speeds. In the first experiment, we measure the performance of the cluster TDMA architecture. Then, we repeat the experiment for the PRNET, the Packet Radio Network developed under ARPA support in the 70's and 80's [8]

4.1.1 Cluster TDMA In this experiment, the TDM frame is 100ms (same as the routing update period), with 10 data slots per frame. Real time traffic is modeled by a continuous source which generates 1 packet per time frame. Because of real time constraints, packets will be dropped if their "age" in the network exceeds 10 frames (i.e. 1second). Each real time connection is mapped into a VC. As described before, with QoS routing a call is accepted only when available bandwidth is above the required QoS. Otherwise, it is rejected. When an area becomes congested because of VC rerouting due to node movements, and there is no sufficient bandwidth to support the initial QoS, low priority packets are dropped first. Only high priority packets are transmitted and QoS is gracefully degraded. If the congested area cannot carry even the high priority stream, the VC experiences a temporary "blackout".

node speed	11 km/hr		55 km/hr	
routing	DSDV+Q	B-F	DSDV+Q	B-F
packet received	88.4%	76.5%	88.2%	73.5%
packet dropped (low priority)	10.2%	15.4%	10.0%	18.3%
packet dropped (high priority)	1.4%	8.1%	1.8%	8.2%
offered load (pkt/frame)	6.19	11.80	5.66	11.80
throughput (pkt/frame)	5.47	9.02	4.99	8.68
packet delay (msec/packet)	189	229	207	232
packets in loop	0%	0.2%	0%	0.2%
out-of-seq pkts	11.2%	11.3%	13.3%	17.8%

Table 1: Cluster TDMA: Traffic Performance

An experiment with Cluster TDMA with Bellman-Ford routing (instead of DSDV+QoS) is also run for comparison, to assess the improvement introduced by loop free operation and QoS reporting. In the B-F case, call acceptance relies only on local rather end to end available bandwidth information. The through-put vs offered load results are summarized in Table 1. Offered load corresponds to the load offered by the calls which are actually accepted by the network. As expected, B-F achieves higher throughput, because of the more aggressive call acceptance policy. It suffers, however, a much higher packet drop rate. Detailed statistics are reported in Table 1. With QoS routing, the low priority packet loss rate is at least 50% less than with B-F routing. More importantly, the high priority packet drop in QoS routing is less than 2% as compared to the 8% in B-F routing. Thus, blackouts are much less frequent. This is due to the more accurate call acceptance control enabled by the QoS information.

The fraction of packets caught in a loop in the B-F scheme is surprisingly low (note: there are no loops in DSDV). This suggest that the most important payoff of DSDV+QoS is its QoS reporting rather than loop suppression. At 11 km/hr, the fraction of packets received out of sequence (yet, within the 1 second constraint) is about the same for the two schemes. Apparently, B-F looping has only negligible impact at this speed. At the hight speed, however, the out-of-sequence effect of B-F looping is more evident. Average packet delay is again comparable for the two schemes. The higher delay experienced by B-F can be attributed to the higher operating load.

4.1.2 PRNET The PRNET simulator is based on the protocols described in [8]. Since there is no notion of slot and frame in the PRNET, we chose the parameters carefully such that the channel capacity in both experiments is comparable. Of course, PRNET has a built in disadvantage since it can use only one channel (with spatial reuse) while cluster TDMA has better separation due to the multiple CDMA channels. Note that there is no notion of VC setup management in PRNET, so all real time traffic is allowed into the network with no call acceptance control.

The followings describe the PRNET protocol features implemented in the simulator: (a) CSMA: negligible propagation time and rx-to-tx switch over time; immediate carrier sensing (no acquisition time); if channel is busy, wait for "randomization delay" until channel becomes idle; (b) ACK: "echo ack" for intermediate nodes, and "active ack" at destination node; (c) flow control: pacing (re)transmission protocol; (d) alternate routing: requesting alternate route if # of transmission exceeds 3 times for datagram, and 1 for real time traffic; packets will be dropped if total # of transmissions exceeds 6 times for datagram, and 3 for real time traffic; (e) routing: B-F routing scheme; no loop suppression; no loss/collision incurred by routing update packets; (f) channel: perfect capture, i.e. packet is received successfully if no collision.

In addition, "duct routing" is used, as described in [19]. Essentially, with duct routing multiple paths can be used to carry multiple copies of a real time packet, in order to improve reliability. Duct routing follows automatically from point (d) above, where alternate routes are used simultaneously for a real time packet.

The results are summarized in Table 2. The poor performance (with respect to cluster TDMA) is mainly due to the lack of acceptance control of the real time calls. The network is overloaded with packets which cannot be delivered. Duct routing does not help in this respect. On the contrary, in heavy load situations, duct routing tends to generate excessive number of "requests for alternate routes", aggravating network congestion, and dramatically reducing system throughput. These results clearly confirm

node speed	11 km/hr	55 km/hr
packet generated	100%	100%
packet received	11.7%	15.6%
packet delay (msec/pkt)	231	266
offered load (pkt/frame)	14.16	9.46
throughput (pkt/frame)	1.65	1.48

Table 2: PRNET: Traffic Performance

that, for real traffic support in a packet radio network we need a structure which provides bandwidth allocation and reservation (like cluster TDMA, for example) and call acceptance (possibly, with QoS information).

5 Conclusion and Future Works

In this paper we evaluate the performance of a QoS routing scheme for multihop packet radio networks, which is the "augmentation" of an existing loop free routing algorithm, DSDV. This QoS routing scheme can be effectively used in the support of real time traffic in multihop wireless network, and find applications in two important scenarios: (a) standalone, multimedia, multihop networks, and; (b) wireless, multihop extensions of an ATM network. In either case, the QoS feature of the routing algorithm is instrumental in establishing and maintaining virtual circuits. In the ATM interconnection case, the QoS routing information can be used to assist in the handoff of the mobile host between different ATM base station.

The simulation results show that the DSDV+QoS algorithm, in spite of some delay and inaccuracy in tracking station movements and traffic fluctuations, leads to substantial improvements over the original B-F routing scheme. In this regard, QoS information seems to be more critical than loop suppression. A comparison with the original PRNET packet radio architecture confirms the importance of QoS routing and more generally, call acceptance control for real time traffic support.

Future research plans include: the inclusion of additional metrics (eg, SIR, Signal to Interference Ratio) in the QoS scheme; the support of multipath routing (and the solution of the associated out-of-order deliver problem) and; QoS multicasting in the wireless network.

References

[1] D. Raychaudhuri and N. D. Wilson. Atm-based transport architecture for multiservices wireless personal communication networks. *IEEE JSAC*, 12(8):1401–1414, October 1994.

[2] J. Porter and A. Hopper. An atm based protocol for wireless lans. Technical report, Olivetti Research Limited, April 1994.

[3] S. Bush, S. Jogannath, J.B. Evans, and V. Frost. Mobile atm orderwire and network configuration. In *ICC*, 1996.

[4] P. Agrawal et al. Swan: A mobile multimedia wireless network. In *IEEE Personal Communications*, pages 18–33, April 1996.

[5] K. Y. Eng et al. Bahama: A broadband ad-hoc wireless atm local-area network. In *Proc. IEEE ICC '95*, pages 1216–1223, June 1995.

[6] A. S. Acampora and M. Naghshineh. An architecture and methodology for mobile executed handoff in cellular atm. In *IEEE Journal on Selected Areas in Communications*, pages 1365–1375, October 1994.

[7] K. Keeton, B. A. Mah, S. Seshan, R. H. Katz, and D. Ferrari. Providing connection-oriented network services to mobile hosts. In *Proceedings of USENIX ASSOCIATION Mobile and Location-Independent Computing System*, pages 83–102, August 1993.

[8] J. Jubin and J. D. Tornow. The darpa packet radio network protocols. *Proceedings of IEEE*, 75(1), January 1987.

[9] M. Gerla and J. T. Tsai. Multicluster, mobile, multimedia radio network. *ACM-Baltzer Journal of Wireless Networks*, 1(3):255–265, 1995.

[10] M. Gerla, J. T. Tsai, N. Bambos, and S. C. Chen. A distributed, mobile wireless infrastructure for multimedia applications. In *Fifth WINLAB Workshop on Third Generation Wireless Networks*, New Jersey, April 1995.

[11] Nicholas Bambos, Shou Chen, and Gregory Pottie. Radio link admission algorithms for wireless networks with power control and active link quality protection. In *IEEE INFOCOM*, pages 97–104, 1995.

[12] Y. Ofek. Generating a fault tolerant global clock using high-speed control signals for the metanet architecture. *IEEE Trans. on Communications*, 42(5):2179–2188, 1994.

[13] D. J. Goodman and S. X. Wei. Efficiency of packet reservation multiple access. In *IEEE Transactions on Vehicular Technology*, volume 40, February 1991.

[14] L. Zhang, S. Deering, D. Estrin, S. Shenker, and D. Zappala. Rsvp: A new resource reservation protocol. *IEEE Network*, pages 8–18, September 1993.

[15] A. Alwan, R. Bagrodia, N.Bambos, M. Gerla, L. Kleinrock, J. Short, and J. Villasenor. Adaptive mobile multimedia networks. In *IEEE PCS Magazine*, 1996.

[16] C. Ferguson. *Routing in a Wireless Mobile CDMA Radio Environment*. PhD thesis, University of California, Computer Science Department, Los Angeles, 1996.

[17] C.E. Perkins and P. Bhagwat. Highly dynamic destination-sequenced distance-vector routing (dsdv) for mobile computers. In *ACM SIGCOMM'94*, pages 234,244, 1994.

[18] R. L. Bagrodia and Wen-Toh Liao. Maisie: A language for the design of efficient discrete-event simulations. In *IEEE Transactions on Software Engineering*, volume 20, pages 225–238, April 1994.

[19] N. Shacham, E. J. Craighill, and A. A. Poggio. Speech transport in packet-radio networks with mobile nodes. *IEEE Journal of Selected Areas in Communications*, 1(6), December 1983.