The Influence of QoS Routing on the Achievable Capacity in TDMA-based Ad hoc Wireless Networks*

S. Sriram, T. Bheemarjuna Reddy, B. S. Manoj, and C. Siva Ram Murthy†
Department of Computer Science and Engineering
Indian Institute of Technology, Madras, India 600036
sriram@dcs.cs.iitm.ernet.in, {arjun, bsmanoj}@cs.iitm.ernet.in, murthy@iitm.ernet.in

Abstract— **The issue of providing QoS guarantees in an Ad hoc wireless network is a challenging problem. Irrespective of the nature of the routing and reservation protocol used in the QoS scheme, there is an inherent limitation on the kind of QoS guarantees that can be provided. Unlike existing studies which analyze the transport capacity, we focus on the achievable capacity. The framework that we assume is that of a TDMA-based network. In this paper, we investigate the achievable capacity and the influence of routing protocols on it. The metrics that we consider are the call acceptance probability and the system saturation probability. We derive general bounds on the call acceptance and the system saturation for the case of multiple-classes of users in the network. These bounds indicate the number of calls of the highest priority class that can be admitted into the network.**

Simulation studies were performed to study the effect of load, hopcount, and the routing protocol on the call acceptance. The increase of the call acceptance with the introduction of load-balancing highlights the importance of load-balancing in enhancing the system performance.

I. INTRODUCTION

With their widespread deployment, Ad hoc wireless networks now need to support applications that generate real-time traffic. Such traffic requires the network to provide guarantees on the QoS of the connection. The important aspects in the process of providing such guarantees are the routing protocols that establish paths that can satisfy the QoS requirements and the reservation mechanisms that reserve the necessary resources along the path. A problem of considerable interest in this regard is that of theoretically estimating the nature of the guarantees that can be provided by a QoS scheme. These estimates allow us to gauge how far existing schemes are from the ideal limit.

In this work, we consider the problem of QoS routing in a TDMA-based Ad hoc wireless network, where the QoS constraint on the calls is that of bandwidth. Our focus is the **achievable capacity** which is an upper-bound on the number of calls that can be supported by the network and depends on the call acceptance probability. The calls arriving in the network belong to different classes based on which the requirements of the calls are prioritized. Thus, the parameters that we focus on are: the probability of call acceptance and the system saturation probability.

We model the network at the level of the transmission range of each node. The range of a node is analyzed as a Markov process where the calls are the entities to be serviced. The reservation of slots for the call in the transmission range constitutes the service of the call. The modeling of a wireless network as a collection of Markov processes is unique in that, due to the local broadcast nature of the channel, the reservation of slots in the transmission range of a node affects the status of the slots in the neighboring regions. Capturing this property of wireless networks is essential to model the characteristics of the network accurately. Such a modeling must also be able to reflect the characteristics of the routing protocol used. We begin by analyzing a general case of a network that can support multiple-classes of calls where preemption of calls does not exist. We then provide a closed-form estimate of the call acceptance probability and the saturation probability for the case of a single-class of users. We compare the call acceptance probabilities of shortest-path routing and a routing protocol that attempts load-balancing.

II. RELATED WORK

In [1] it was shown that even under the optimal conditions, the transport capacity (bit-distance product) of an N-node network over a W bps channel is $\theta(W\sqrt{N})$ bit-meters per second, for protocol model considered. It means that the throughput obtained by each user is $\theta(\frac{W}{\sqrt{N}})$ which diminishes to zero as N increases, and may be not acceptable to users present in the network. In [2], the authors have tried to address issues related to the capacity of wireless networks without making preconceived assumptions about how networks operate. In [3] the problem of maximizing the transport capacity of a single-transmitter Ad hoc network in a Gaussian power law channel was addressed. In [4] the stability and capacity problems of wireless networks were studied using probability density functions that determine the packet reception probabilities. These studies analyze the transport capacity of the network. In this work, our focus is on the achievable capacity: a measure

* This work was supported by the iNautix Technologies India Private Limited, Chennai, India and the Department of Science and Technology, New Delhi, India.

† Author for correspondence.

of the number of calls with end-to-end bandwidth reservation that can be supported by the network. The routing scheme employed influences the achievable capacity of the network. We investigate the achievable capacity and the influence of shortest-path and load-balanced routing protocols on it.

III. Our Work

We consider an Ad hoc wireless network comprising N nodes randomly distributed in a terrain of area A. The transmission range of each node is R. We assume the presence of a slotted TDMA mechanism at the MAC layer. The total number of slots available is B. The bandwidth of a call is measured in terms of the number of slots used for transmission. A call is setup by reserving slots along the path of the call. A node may transmit or receive in a particular slot. A slot is said to be free at a node j, $(1 \leq j \leq N)$, if it is neither transmitting nor receiving during that slot. For a node j to transmit in a particular slot, the slot must be free at j and none of the neighbors of j must be receiving in that slot. For a node j to receive in a particular slot, the slot must be free at j.

A. System Model

Consider a network $NW = \{1, \ldots, N\}$ of N nodes that can support K classes of calls where class i calls have a higher priority than class j $(j > i)$ calls and can preempt the latter. We would like an estimate of how many calls of a particular class can be supported. This implies that we can definitely support such a number of class 1 calls.

Calls of a particular class-k arrive at each node distributed according to a *poisson* process of mean λ_k. The duration of a call is exponentially distributed with mean duration $\frac{1}{\mu_k}$. We assume that the routing algorithm is such that for any path found by the algorithm, for any intermediate node on the path, the number of other nodes on the path that lie within its transmission range is not greater than some constant c. In the absence of such an assumption, it is possible to construct a scenario (Fig. 1) where a single call needs to use all the slots in the system. Hence, it would be difficult to provide a bound on the number of calls that can be admitted. This property is satisfied with $c = 2$ for protocols that ensure that if a path is to be set up from node A to node C, the path used is the link(A, C) rather than links (A, B) and (B, C), where A, B, and C are nodes such that each can listen to the other two. This can be done by using an appropriate forwarding of the route request packets.

B. Theoretical Analysis

We assume that the calls of all classes have equal bandwidth requirements: each call requires a single slot. For the purpose of the current analysis, we also assume that preemption does not occur.

Consider a node j and the region spanned by its transmission range $R(j)$. Any call passing through $R(j)$ uses up some number of slots. A slot is said to be **free** in $R(j)$ if no nodes in $R(j)$ are either transmitting or receiving in that slot. We can view $R(j)$ as a server of slots for which the calls contend.

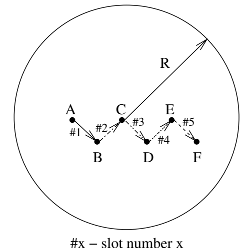

Fig. 1. An example scenario. Each of the nodes on the path is in the transmission range of the other nodes. If we were to consider a P-hop path with the nodes in the configuration given, the number of slots used would be P.

Although the distribution of call arrivals of a particular class at each node is known to be poisson, the distribution of calls arriving at $R(j)$ is not poisson due to the splitting of the poisson streams. We make use of Kleinrock's Independence Assumption, according to which, for moderately heavy call arrival at each node, the net call arrival at the region $R(j)$ can be regarded as poisson. Thus, calls of a class-k arrive at $R(j)$ according to a poisson distribution with mean:

$$\lambda_k(j) = \sum_{i=1}^{i=N} f_k(i, j)\lambda_k \tag{1}$$

where $f_k(i, j)$ is the fraction of class-k calls originating in node i that pass through the region $R(j)$. The parameter $f_k(i, j)$ is **dependent** on the routing protocol. For a protocol such as shortest-path routing which leads to heavy loads in the center of the network, $f_k(i, j)$ would be high for nodes j $(1 \leq j \leq N)$ located near the center. For protocols that implement load-balancing, the value of $f_k(i, j)$ should be fairly uniform across the nodes.

The state of the system $R(j)$ is given by the number of calls of each class being served by $R(j)$. We thus model $R(j)$ as a K-dimensional discrete-time Markov process $X(t) = (n_1, \ldots, n_K)$, where n_k denotes the number of class-k calls being served by $R(j)$ at time t [5].

Denote:
$P((n'_1, \ldots, n'_K)|(n_1, \ldots, n_K))$=$P(X(t + \Delta t) = (n'_1, \ldots, n'_K)|X(t) = (n_1, \ldots, n_K))$ as the probability that the system $R(j)$ is in the state $(n'_1, \ldots, n'_K)$ at time $t + \Delta t$ given it is in the state $(n_1, \ldots, n_K)$ at time t.

$$P((n_1, \ldots, n_k + 1, \ldots, n_K)|(n_1, \ldots, n_k, \ldots, n_K)) = \lambda_k(j)\Delta t \tag{2}$$

$$P((n_1, \ldots, n_k - 1, \ldots, n_K)|(n_1, \ldots, n_k, \ldots, n_K)) = n_k\mu_k\Delta t, \ n_k > 0 \tag{3}$$

The Markov process has a unique steady-state probability distribution [5]. From (2) and (3), the probability that the system is in a particular state $(n_1, \ldots, n_K)$ is:

$$P((n_1, \ldots, n_K)) = \frac{1}{G(j)} \prod_{k=1}^{k=K} \frac{\rho_k(j)^{n_k}}{n_k!} \tag{4}$$

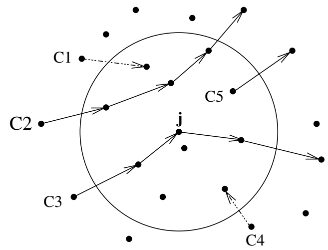

Fig. 2. In the region $R(j)$, C1 and C4 are type-U calls; C2, C3, and C5 are type-V calls. For each type-V call, we see that at least one slot that has not been used so far in $R(j)$ must be used. For the type-U calls, slot reuse is possible in some cases.

$\rho_k(j) = \frac{\lambda_k(j)}{\mu_k}$ and $G(j) = \sum_{0 \le n_1 +, \ldots, + n_K \le B} \prod_{k=1}^{k=K} \frac{\rho_k(j)^{n_k}}{n_k!}$.

We would now like to extend this markov process to distinguish between calls that terminate in a node in $R(j)$ (call them type-U calls) and those that do not (type-V calls). Let us say that a fraction f of the calls terminate in some node in $R(j)$. If the destination were to be chosen randomly, then $f = \frac{|N(j)|+1}{N}$. The state of the system is now given by: $(n_{1,U}, n_{1,V}, n_{2,U}, n_{2,V}, \ldots, n_{K,U}, n_{K,V})$, where $n_{k,U}$ and $n_{k,V}$ are the number of class-k calls in $R(j)$ that are type-U calls and type-V calls, respectively.

The probability that the system is in a state $(n_{1,V}, n_{2,V}, \ldots, n_{K,V})$ is: $P((n_{1,V}, \ldots, n_{K,V})) = \frac{1}{H(j)} \prod_{k=1}^{k=K} \frac{\rho_{k,V}(j)^{n_{k,V}}}{n_{k,V}!}$ where $\rho_{k,V}(j) = (1-f)\rho_k(j)$ and $H(j) = \sum_{0 \le n_{1,V} +, \ldots, + n_{K,V} \le B} \prod_{k=1}^{k=K} \frac{\rho_{k,V}(j)^{n_{k,V}}}{n_{k,V}!}$.

1) Call Acceptance Probability: In this section, we are going to derive the call acceptance probability of both single-hop and multi-hop cases for a non-preemptive system.

Single-hop case: Consider a single-hop call from node j to its neighbor node l. For the call to be accepted, at least one slot must be free in the region $R(j)$. For every type-V call, at least one free slot in the region $R(j)$ must be used (Fig. 2). Thus:

$$\begin{aligned} B &\ge Number\ of\ used\ slots \\ &\ge Number\ of\ type - V\ calls \end{aligned} \quad (5)$$

Then P_{Acc} (probability of a single-hop call is accepted) is:

$$\begin{aligned} P_{Acc} &= P(Number\ of\ free\ slots \ge 1) \\ &= P(Number\ of\ used\ slots \le B-1) \\ P_{Acc} &\le P(Number\ of\ type-V\ calls \le B-1) \\ &\le 1 - P(Number\ of\ type-V\ calls > B-1) \end{aligned}$$

From (5),

$$\begin{aligned} P_{Acc} &\le 1 - P(Number\ of\ type-V\ calls = B) \\ &\le 1 - \sum_{\sum_{i=1}^{i=K} n_{i,V} = B} \frac{1}{H(j)} \prod_{k=1}^{k=K} \frac{\rho_{k,V}(j)^{n_{k,V}}}{n_{k,V}!} \end{aligned} \quad (6)$$

For the case of a single-class of calls, (6) reduces to

$$P_{Acc} \le 1 - \frac{1}{H(j)} \frac{\rho_{1,V}(j)^B}{B!} \quad (7)$$

Multi-hop case: We set the constant $c = 2$. Consider a $(M-1)$-hop call $(M \ge 3)$ setup along the nodes $(p_1, \ldots, p_M)$. For the call to be accepted, the minimum requirements are that at each of the regions $R(p_1), \ldots, R(p_{M-1})$, at least one slot must be free. In addition, when a slot is reserved for transmission between p_1 and p_2, the total number of free slots at $R(p_2)$ decreases by 1 (since the slot cannot be used for transmission from p_2 to p_3). Thus, the total number of slots available at $R(p_2)$ can be considered as $B-1$. When slots have been reserved between p_1 and p_2, and between p_2 and p_3, the number of free slots at $R(p_3)$ decreases by 2 so that the total number of slots at $R(p_3)$ can be regarded as $B-2$. The number of slots, for the regions $R(p_3), \ldots, R(p_{M-1})$, is thus effectively, $B-2$. Thus:

$$P_{Acc} \le \prod_{i=1}^{i=M-1} P(Number\ of\ free\ slots \ge 1\ in\ R(p_i))$$

$$\begin{aligned} P_{Acc} \le & \\ & (1 - \sum_{\sum_{k=1}^{K} n_{k,V} = B} \frac{1}{H(p_1)} \prod_{k=1}^{k=K} \frac{\rho_{k,V}(p_1)^{n_{k,V}}}{n_{k,V}!}) \times \\ & (1 - \sum_{\sum_{k=1}^{K} n_{k,V} = B-1} \frac{1}{H'(p_2)} \prod_{k=1}^{k=K} \frac{\rho_{k,V}(p_2)^{n_{k,V}}}{n_{k,V}!}) \times \\ \prod_{i=3}^{i=M-1} & (1 - \sum_{\sum_{k=1}^{K} n_{k,V} = B-2} \frac{1}{H''(p_i)} \prod_{k=1}^{k=K} \frac{\rho_{k,V}(p_i)^{n_{k,V}}}{n_{k,V}!}) \end{aligned} \quad (8)$$

where $H'(j) = \sum_{0 \le n_{1,V} +, \ldots, + n_{K,V} \le B-1} \prod_{k=1}^{k=K} \frac{\rho_{k,V}(j)^{n_{k,V}}}{n_{k,V}!}$ and $H''(j) = \sum_{0 \le n_{1,V} +, \ldots, + n_{K,V} \le B-2} \prod_{k=1}^{k=K} \frac{\rho_{k,V}(j)^{n_{k,V}}}{n_{k,V}!}$. For the case of a single-class of calls, (8) reduces to

$$\begin{aligned} P_{Acc} \le & [1 - \tfrac{1}{H(p_1)} \tfrac{\rho_{1,V}(p_1)^B}{B!}] \times [1 - \tfrac{1}{H'(p_2)} \tfrac{\rho_{1,V}(p_2)^{B-1}}{(B-1)!}] \times \\ & \prod_{i=3}^{i=M-1} [1 - \frac{1}{H''(p_i)} \frac{\rho_{1,V}(p_i)^{B-2}}{(B-2)!}] \end{aligned} \quad (9)$$

The RHS of (7) and (9) are hard to solve for in a closed-form. For moderate-to-heavy traffic, $\rho > 1$ and the inequality remains valid if we replace $\rho_{1,V}(p_j), 1 \le j \le M-1$ by ${\rho_{1,V}}^{Max}$ (the maximum value of $\rho_{1,V}(p_j)$ across all the regions). Let $H = \sum_{b=0}^{b=B} \frac{{\rho_{1,V}}^{Max^b}}{b!}$. Denoting the RHS as P_{Acc}^{Max}:

$$P_{Acc}^{Max} = 1 - \frac{1}{H} \frac{{\rho_{1,V}}^{Max^B}}{B!} \quad \text{for single} - \text{hop calls} \quad (10)$$

$$P_{Acc}^{Max} = \left[1 - \frac{1}{H}\frac{{\rho_{1,V}^{Max}}^B}{B!}\right] \times \left[1 - \frac{1}{H'}\frac{{\rho_{1,V}^{Max}}^{B-1}}{(B-1)!}\right] \times \left[1 - \frac{1}{H''}\frac{{\rho_{1,V}^{Max}}^{B-2}}{(B-2)!}\right]^{M-3} \text{for multi-hop calls} \quad (11)$$

where $H' = \sum_{b=0}^{b=B-1} \frac{{\rho_{1,V}^{Max}}^b}{b!}$ and $H'' = \sum_{b=0}^{b=B-2} \frac{{\rho_{1,V}^{Max}}^b}{b!}$.

2) System Saturation: For the case of a single-class of calls, the probability that the network is saturated *i.e.,* no further calls can be accepted is given by P_{Sat}. If the number of type-V calls in a region is B, then this would require at least B slots to be used, and no further calls can be accepted. $P(Saturation\ in\ R(j)) = P(B\ slots\ are\ used)$

$$\begin{aligned} &P(Saturation\ in\ R(j)) \\ &\geq P(Number\ of\ type-V\ calls\ at\ R(j) = B) \\ &\geq \frac{1}{H(j)}\frac{\rho_{1,V}(j)^B}{B!} \end{aligned} \quad (12)$$

$$P_{Sat} \geq \prod_{i=1}^{i=N} \frac{1}{H(i)}\frac{\rho_{1,V}(i)^B}{B!} \quad (13)$$

$$P_{Sat} \geq \left[\frac{1}{H}\frac{{\rho_{1,V}^{Max}}^B}{B!}\right]^N \quad (14)$$

3) The Case of Preemption: The analysis so far has been done under the assumption that high-priority calls cannot preempt lower-priority ones. However, a realistic scenario may require that high-priority calls are ensured high probability of call acceptance. This may require introduction of preemption into the system. The analysis of the steady-state probabilities of a preemptive Markov process is a difficult problem. The stationary distribution of the highest priority calls can be easily obtained since these calls effectively ignore the presence of other low-priority calls. Thus, the stationary distribution of the class-1 calls is the same as that of the single-class system given in (10) and (11).

IV. Simulation Studies

The analysis tells us that the parameters: the call acceptance probability and the system saturation probability depend on the load on the network, the hopcount of the path, and the routing protocol. We first look at the effect of the routing protocol.

A. Routing Protocols

The routing protocol is related to the call acceptance and the system saturation probability through the factor $f_k(i,j)$ specified in (1). To study the effect of the routing protocol, we consider the following two routing strategies:

- Shortest-path routing: Shortest-path routing computes the shortest-path between the source and the destination where the distance refers to the Euclidean distance between the nodes.
- Ring-based routing: Ring-based routing ensures that the load is distributed evenly across the network. A node is identified as the center of the network and each node is assigned to a ring concentric about the center based on the hopcount of the shortest path from the center. The scheme makes use of heuristics to balance the load.

The load balancing heuristic that we use is a Preferred Outer Ring routing Scheme (PORS) [6]. In this strategy, traffic generated in a node in $Ring_i$ and destined for a node in $Ring_j$ must not go beyond the rings enclosed by $Ring_i$ and $Ring_j$. Further, the packets must be preferentially routed through the outer of the two rings. Thus, for nodes belonging to the same ring, packets must be preferentially transferred in the same ring. For nodes belonging to different rings, all angular transmissions must preferentially take place in the outer of the two rings while the radial transmissions transfer packets across the rings. Thus, PORS affects the hopcount while at the same time moving most of the load away from the center. For a discussion on the determination of the center and the rings, and an example of PORS, see [6].

B. Simulation

To study the actual behavior of the parameters of interest, we built an Ad hoc wireless network simulator in C++. The network is static and TDMA-based. Slot allocation for a particular call is done in a greedy manner. If at any intermediate node, the number of free slots is found to be inadequate, the call is rejected. Calls are generated at each node according to a poisson process and the accepted calls have an exponentially distributed call duration. We simulated a network of 50 nodes, each with a range of 300 m, in a 1000 m×1000 m terrain having a 32-slot TDMA mechanism. The average call duration was 30 sec while the simulation was run for a duration of 200 sec.

For the simulation studies, we vary the load by varying the call arrival rate at each node. We compare the call acceptance probabilities for varying values of the ratio $\rho = (Average\ Call\ Arrival\ Rate) \times (Average\ Call\ Duration)$.

In order to compare the theoretical values and the experimental results, we need to translate the ρ value to the $\rho_{1,V}^{Max}$ value which denotes the average ratio of calls arriving at the region $R(j)$. Thus, we also measure the average fraction of calls that pass through a region. This factor is an indication of the nature of the routing protocol used. Since the estimate of the call acceptance probability is dependent on the hopcount of the path setup, the simulator ensures that for a given hopcount value, only calls whose paths are of the specified hopcount are admitted. This ensures that we can easily compare the performance of the routing schemes against the theoretical estimates which depends on the hopcount value, by fixing the hopcount of the paths for the simulation.

V. Simulation Results

A. Probability of Call Acceptance

We have compared the probabilities of call acceptance of shortest-path routing, PORS and the theoretical upper bound

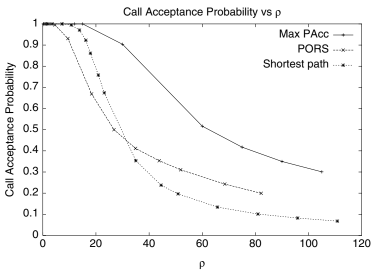

Fig. 3. Variation of Call Acceptance vs ρ for Single-hop calls.

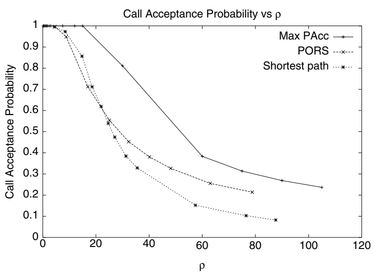

Fig. 4. Variation of Call Acceptance vs ρ for 2-hop calls.

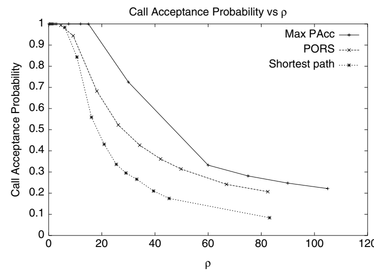

Fig. 5. Variation of call Acceptance vs ρ for 3-hop calls.

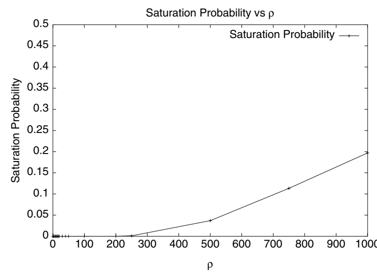

Fig. 6. Variation of Saturation Probability vs ρ for 3-hop calls.

at different values of load (in terms of the ratio ρ). We have also studied the acceptance probability for hopcount values of 1, 2, and 3 (Figs. 3, 4, and 5). The call acceptance probability value decreases with an increase in the network load, as expected. Both the routing schemes have lower call acceptance probabilities compared to the theoretical upper bound. This could be due to two reasons: firstly, both the routing protocols are sub-optimal, and secondly, the bound derived is not tight enough. On the other hand, the call acceptance probability of PORS is higher than that of shortest-path routing especially for the 2-hop and the 3-hop cases. This increase in the call acceptance probability of PORS as compared to shortest-path routing indicates the importance of load-balancing in ensuring better throughput in terms of call acceptance. In fact, load-balancing seems to be an important method of approaching the theoretical upper bound.

B. Probability of System Saturation

The variation of the probability of system saturation with load is shown in Fig. 6. This metric remains near zero for moderate-to-heavy loads, and takes on an appreciable value only at very high values of load. This indicates that system saturation is a rare occurrence for the common values of load. For common values of load, it is always possible to ensure that some fraction of the calls are guaranteed acceptance. This fraction depends on the probability of call acceptance.

VI. Conclusion

A realistic analysis of the nature of QoS guarantees is crucial in the design of new protocols and the improvement of existing ones to handle the growing diversity of demands on networks. In this paper, we have analyzed a TDMA-based Ad hoc wireless network and derived an upper bound on the probability of call acceptance and a lower bound on the probability of system saturation. Our analysis takes into consideration the behavior of the routing protocol and the inter-dependence of resources (time-slots) of neighboring nodes. Further, our simulation studies indicate that the protocols tested fall short of the established bounds. Amongst the two protocols compared, the one that incorporated load-balancing out-performed the shortest-path routing based protocol. This clearly indicates the importance of load-balancing in the provision of better QoS guarantees.

The experimental studies in this work were performed with a single-class of calls. The next step would involve studying the effect of introducing multiple classes of calls. Further, we are working on extending the analysis to handle the case of call preemption, and on obtaining tighter estimates. The experimental studies also need to be extended to compare other protocols to infer the essential properties in attaining optimal-behavior. This will also serve as a guideline in the design of protocols to meet specific QoS guarantees.

References

[1] P. Gupta and P. R. Kumar, "The Capacity of Wireless Networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388-404, March 2000.

[2] L. L. Xie and P. R. Kumar, "A Network Information Theory for Wireless Communication: Scaling Laws and Optimal Operation," unpublished.

[3] A. Reznik and S. Verdu, "On the Transport Capacity of a Broadcast Gaussian Channel," *in proceedings of Communications in Information and Systems*, vol. 2, no. 2, pp. 157-190, September 2002.

[4] G. Mergen and L. Tong, "Stability and Capacity of Wireless Networks with Probabilistic Receptions: Part 1–General Topologies," unpublished.

[5] D. Bertsekas and R. Gallager, "Data Networks," *Prentice-Hall of India Private Limited*, 2nd edition, 2000.

[6] G. Bhaya, B. S. Manoj, and C. Siva Ram Murthy, "Ring-Based Routing Schemes for Load Distribution and Throughput Improvement in Multi-hop Cellular, Ad hoc, and Mesh Networks," *in proceeding of HiPC 2003*, pp. 152-161, December 2003.