A Cross-Layer Design Approach to Multicast in Wireless Networks

Weiyan Ge, *Student Member, IEEE,* Junshan Zhang, *Member, IEEE,* and Sherman Shen, *Senior Member, IEEE*

***Abstract*— We study rate optimization for multicast communications at the media access control (MAC) layer, and explore transport layer erasure coding to enhance multicast reliability in wireless networks. We start with investigating network models with single-input-single-output (SISO) links. For *Threshold*-T based multicast policies, we characterize the optimal transmission rates that maximize the throughput in stable networks and in saturated networks, respectively. We investigate the tradeoff between stability and throughput therein. We then generalize our study to network models with multiple-input-multiple-output (MIMO) links and non-i.i.d. channel links, and investigate the optimal transmission rate. In addition, to ensure multicast reliability while no retransmission is required at the MAC layer, we propose to use transport layer erasure coding for reliability enhancement, where the problem boils down to jointly optimizing the transmission rate and the multicast threshold. We provide a solution to this optimization problem accordingly.**

***Index Terms*— Rate optimization, multicast, erasure coding, reliable multicast.**

I. Introduction

MULTICAST is an efficient mechanism to transmit data to multiple receivers in wireless networks. Since wireless communication is broadcast in nature, one information packet can be received by many receivers through one transmission. This property, called the *wireless multicast advantage* [22], can enhance bandwidth efficiency and reduce transmission power consumption considerably compared to unicast communications, where a packet has to be transmitted on each link separately. In particular, multicast can be used for audio-video conferencing, disaster recovery, and military operations.

Most existing work on wireless multicast focuses on network layer multicast strategies, e.g., energy efficient multicast routing protocols. A properly designed medium access control (MAC) layer multicast protocol would significantly improve the network performance. However, MAC layer multicast is not well understood. Another important issue is the reliability of multicast transmissions. Needless to say, it is of great interest to design a cross-layer protocol to provide reliable and efficient multicast transmissions. In this paper, we first address the rate optimization of MAC layer multicast, and then propose to use erasure coding at the transport layer to enhance the reliability of multicast transmissions.

Consider a single-hop network where one transmitter desires to send packets simultaneously to multiple destinations. For a given transmission rate, some receivers may not receive the packet correctly if the corresponding channels experience deep fading. Accordingly, only some but not all receivers would be ready to receive the packet. One approach is to transmit only when all the receivers are ready to receive. This may result in large delays, making the network unstable. On the other hand, if the transmitter sends the packet without any knowledge of the channels, as in IEEE 802.11, a severe packet loss may occur, making the network unreliable. In a nutshell, the throughput may be poor in both cases. A more plausible policy is to send packets when some receivers are ready [2], and this mechanism can be combined with transport layer erasure coding ([13]) to achieve reliable transmissions.

We consider multicast with a pre-determined threshold $T > 0$, i.e., the transmitter sends packets if at least T receivers are ready (namely the *Threshold-T policy* [2]). Then, the throughput is a function of the transmission rate R. Intuitively, for a larger R, the actual transmission time would be decreased, but the channel outage probability may be increased. On the other hand, a lower transmission rate would reduce the waiting duration but increase the actual transmission time. For those two extreme cases, namely $R = 0$ and $R = \infty$, it is clear to see that the corresponding throughput is zero.

In general, the *Threshold-T* policy is not reliable since only some but not all receivers receive the packet successfully in each transmission. One standard solution is to use retransmission. However, it may not be efficient for multicast communications, simply because the lost packets may vary from receiver to receiver, i.e., each retransmission benefits a subset of the receivers only. An alternative approach is to use transport layer erasure coding (also known as *Digital Fountain* [1]). The basic idea is that k original data packets are encoded to n packets with $n = k + h$, where $h > 0$ is the number of redundant packets. A receiver can recover all the original packets as long as enough encoded packets are successfully received. A main advantage of this scheme is that one redundant packet can be used by different receivers to recover different lost packets. It is shown in [11] that the approximate decoding time of this scheme is $O(k)$, which makes it practically appealing.

In this paper, we first investigate the MAC layer multicast, and characterize the optimal transmission rates that maximize

Manuscript received June 24, 2005; revised October 18, 2005, February 24, 2006, and June 29, 2006; accepted August 17, 2006. The associate editor coordinating the review of this paper and approving it for publication was J. Hou. This research is supported in part by National Science Foundation through the CAREER award ANI-0238550, and Office of Naval Research through the YIP grant N00014-05-1-0636.

W. Ge and J. Zhang are with the Department of Electrical Engineering, Arizona State University, Tempe, AZ 85287 USA (email: {junshan.zhang, weiyan.ge}@asu.edu).

S. Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario N2L 3G1, Canana (email: xshen@bbcr.uwaterloo.ca).

Digital Object Identifier 10.1109/TWC.2007.05468.

the throughput in stable networks and in saturated networks, respectively. Then we address the reliable multicast transmission through transport layer erasure coding and optimize the multicast threshold T. To the best of our knowledge, our work is the first cross-layer scheme that combines the transport layer erasure coding with MAC layer multicast policy to provide reliable and efficient multicast transmissions. Our contributions can be summarized as follows. 1) We analyze rate optimization in SISO multicast networks. 2) We extend our results to the networks with MIMO links and non-i.i.d. channel links. 3) We propose a cross layer design approach to provide reliable and efficient multicast communications.

The remainder of the paper is organized as follows. We discuss related work in Section II. The analysis of MAC layer rate optimization is presented in Section III. In Section IV, we investigate the reliable multicast problem. Conclusions and future work are given in Section V.

II. Related Work

The main objective of MAC layer multicast is to provide efficient channel access to resolve channel contention and maximize network throughput. The most popular MAC layer multicast strategy is perhaps the *Threshold-0* scheme used in IEEE 802.11, where a packet is transmitted without any knowledge of the channel. To improve the performance of the *Threshold-0* strategy, a *Threshold-1* scheme has been proposed in [19], where the transmitter broadcasts the request-to-send (RTS) packet first and then transmits the packet if at least one clear-to-send (CTS) packets are received. In [20], a unicast based multicast has been proposed. The basic idea is to reliably transmit each packet to each neighbor in a round-robin fashion. It does not exploit the broadcast nature of wireless medium. A two threshold transmission policy *(T,q)-policy* has been studied in a recent interesting work [2], in which the threshold is set to be a constant T with probability q or $T+1$ with probability $1-q$. It has been shown that this policy is $\epsilon-optimal$ subject to stability conditions or loss constraints.

Recently, the throughput-delay tradeoff in cellular multicast has been investigated in [4], in which the transmission rate is set such that a fixed fraction of the receivers is able to decode the packet. Then, the scaling behavior of throughput and the delay are characterized as the number of receivers grows. In contrast, our study examines a wireless network with finite number of receivers and the number of ready receivers at each transmission is random.

There is a great deal of interest in reliable multicast transmissions ([1], [13]). The early work on combining the Reed-Solomon codes with automatic repeat request (ARQ) to provide reliable multipoint transmission has been shown in [10]. A quality-of-service (QoS)-based adaptive *Forward Error Correction* (FEC) scheme for multicast communication has been proposed to dynamically control coding parameters [12] . In [14], the erasure coding scheme has also been used in wireless security design to guarantee the reliability of multicast authentication.

There also have been extensive studies on network layer multicast with focus on how to establish efficient routing protocols (e.g., [7], [16]). Energy efficiency is an important

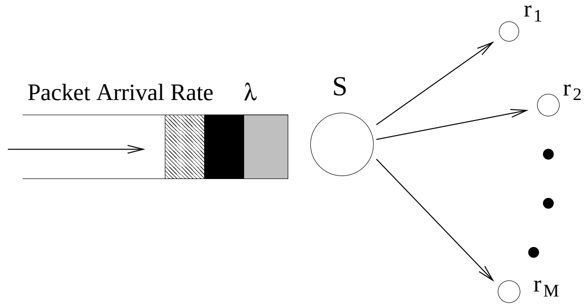


Fig. 1. Single transmitter with M receivers and average packet arrival rate λ.

issue for energy-limited networks. In [22], a protocol named *broadcast incremental power* (BIP) has been proposed to construct the energy-efficient multicast routing tree. A revised BIP protocol has been proposed in [23] to show the impact of bandwidth on the performance of the scheme. Both protocols are sub-optimal since it has been shown in [8] that the construction of minimum-energy multicast tree in wireless networks is NP-hard.

III. Rate Optimization in Single Hop Networks

A. Basic Setting

Consider a single hop network with one transmitter S and totally M potential receivers $\{r_1, r_2, \cdots, r_M\}$, as shown in Fig. 1. Each node is equipped with an omni-directional antenna. Let s denote the transmitted symbols. The received symbols at the i-th receiver, denoted y_i, is given by

$$y_i = \sqrt{P} h_i s + n_i, \quad (1)$$

where P is the total transmission power, h_i is the channel gain between the transmitter and the i-th receiver, which is assumed to be a complex Gaussian random variable with zero mean, and n_i is the additive white Gaussian noise with $\mathcal{CN}(0, N_0)$. Then the corresponding capacity of the channel is given by

$$C_i = \log(1 + \frac{P}{N_0}|h_i|^2). \quad (2)$$

Definition 3.1: The i-th receiver is **ready to receive** if the capacity of the corresponding channel is no less than the transmission rate R, i.e., $C_i \geq R$ [5].
We assume the *block fading* channel model [17], i.e., $\{h_i\}$ are i.i.d. for different time slots, but remain constant during one time slot. Let p_i denote the probability that the i-th receiver is ready to receive. Then

$$\begin{aligned} p_i &= Pr\{C_i \geq R\} \\ &= \exp\left(\frac{(1-2^R)N_0}{P}\right), \ i = 1, 2, \cdots, M. \end{aligned} \quad (3)$$

In this section, we assume that the channel gains $\{h_i\}$ have the same distribution. Hence, the probabilities defined in (3) are identical for all links, i.e.,

$$p_1 = p_2 = \cdots p_M \triangleq p. \quad (4)$$

In the next section, we will extend the studies to non-i.i.d. cases.

Suppose that the packets arrive at the transmitter in accordance with a Poisson rate λ, as shown in Fig. 1. We assume that each packet is transmitted only once, i.e., no retransmission is done at the MAC layer protocol (multicast reliability is guaranteed through the transport layer erasure coding, to be discussed in Section IV). Due to channel fading, the transmitter may have to wait for a long duration if it is required that all potential receivers are ready, making the system unstable. On the other hand, if the transmitter sends a packet without any knowledge of the channel, the packets loss may be high, making the network unreliable. Moreover, it may happen that the channel condition is good but the buffer of the transmitter is empty. Therefore, it is plausible to set a pre-determined threshold $T > 0$. Before each transmission, the transmitter first queries the channel by exchanging control packets and transmits the packet as long as at least T of M receivers are ready. Otherwise, it would back off for a random duration and query the channel again. As in [2], we assume that after each transmission, the transmitter would also back off for a random duration before querying the channel again, so as to allow other transmitters to use the shard media. The querying overhead would increase with the number of receivers M. In this study, M is relatively small, and the overhead incurred by querying is not significant. A more accurate model that includes the overhead will be investigated in future studies.

Clearly, the transmission rate R plays a key role in the multicast scheme. Intuitively, a higher transmission rate decreases the packet transmission time, but it may increase the back-off durations. On the other hand, if the transmission rate is decreased, it would decrease the waiting time but possibly increase the packet transmission time. Thus motivated, we investigate the optimal transmission rate to maximize the MAC layer throughput in the following.

We assume that the duration of the querying plus random back-off and that of packet transmission time are comparable. Let X_i denote the duration of the i-th querying and back-off, and B denote the total time for querying and back-off before one transmission. It follows that

$$\mathbf{E}[B] = \mathbf{E}\left[\sum_{i=1}^{K} X_i\right] = \frac{\mathbf{E}[X]}{P_t}, \tag{5}$$

where K is a geometric random variable with parameter P_t, which is the probability that at least T receivers are ready to receive. Then, the average service time for each packet is given by

$$\begin{aligned}\mathbf{E}[S] &\triangleq \mathbf{E}[\text{service time}] \\ &= \mathbf{E}[\text{transmission time}] \\ &\quad +\mathbf{E}[\text{querying} + \text{backoff duration}] \\ &= \frac{\mathbf{E}[V]}{R} + \frac{\mathbf{E}[X]}{P_t},\end{aligned} \tag{6}$$

where $\mathbf{E}[V]$ is the average packet length.[1]

[1]Strictly speaking, the service time in (6) is an approximation when a head-of-queue packet arrives in between a slot, although the difference is negligible.

B. The Stability Region

As is standard, the traffic load ρ is given by

$$\rho \triangleq \lambda \mathbf{E}[\text{service time}] = \lambda\left(\frac{\mathbf{E}[X]}{P_t} + \frac{\mathbf{E}[V]}{R}\right). \tag{7}$$

Definition 3.2: A system is **stable** if the busy period of the transmitter is finite, i.e., the traffic load ρ is less than 1 [15].

Definition 3.3: The **stability region** is the region $[0, \lambda_{max}(R)]$, where $\lambda_{max}(R)$ is the maximum value of packet arrival rate λ that makes the network stable with transmission rate R.

It is clear that

$$\lambda_{max}(R) < 1/\left(\frac{\mathbf{E}[X]}{P_t} + \frac{\mathbf{E}[V]}{R}\right). \tag{8}$$

For any given multicast threshold T, the transmission probability P_t in (5) is given by

$$P_t = Pr\{\text{at least T receivers are ready}\} = \sum_{i=T}^{M} q_i, \tag{9}$$

where q_i is the probability that exactly i receivers are ready. Note that

$$\begin{aligned}\sum_{i=T}^{M} q_i &= \sum_{i=T}^{M} \binom{M}{i} p^i [1-p]^{M-i} \\ &= I_{(p)}(T, M-T+1),\end{aligned} \tag{10}$$

where p is defined in (4), and $I_{(P)}(a, b)$ is the *incomplete Beta function* [18]. The next key step is to find the optimal transmission rate that supports the maximum arrival rate, i.e.,

$$R_1^* = \arg\max_{R>0}\{\lambda_{max}(R)\}. \tag{11}$$

Therefore, the optimal transmission rate to maximize the stability region is given by

$$R_1^* = \arg\min_{R>0}\left\{\frac{\mathbf{E}[X]}{I_{(p)}(T, M-T+1)} + \frac{\mathbf{E}[V]}{R}\right\}. \tag{12}$$

Fig. 2 illustrates the maximum packet arrival rate λ_{max} as a function of transmission rate. It is not surprising that $T = 1$ has the largest stability region.

C. Optimal Rates for Throughput Maximization Subject to Stability

Following [2], we define the MAC layer throughput as follows.

Definition 3.4: The **MAC layer throughput** is the expected number of successful received packets per unit time.

By definition, if the network is stable, the throughput is the product of the packet arrival rate λ and the expected received packets per transmission.

Let $b(t)$ denote the number of queries till time t, $b_i(t)$ the number of queries with i ready receivers till t, and $p(t)$ the number of transmitted packets till t. Then, by the *Law of Large Numbers*,

$$\lim_{t\to\infty} \frac{b_i(t)}{b(t)} = q_i \qquad \text{w.p. } 1. \tag{13}$$

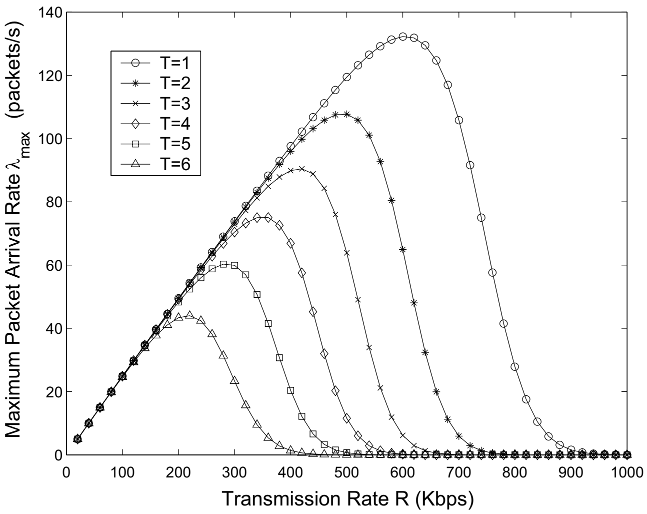


Fig. 2. λ_{max} as a function of transmission rate ($M = 6$).

and

$$\lim_{t\to\infty}\frac{p(t)}{b(t)} = Pr\{\text{at least T receivers are ready}\} = \sum_{i=T}^{M} q_i \quad \text{w.p. 1.} \tag{14}$$

If the network is stable,

$$\lim_{t\to\infty}\frac{p(t)}{t} = \lambda \quad \text{w.p. 1.} \tag{15}$$

By definition, the throughput is given by

$$\begin{aligned} Th(R) &= \lim_{t\to\infty}\frac{\sum_{i=T}^{M} ib_i(t)}{t} \\ &= \sum_{i=T}^{M} i \lim_{t\to\infty}\frac{b_i(t)}{b(t)}\frac{b(t)}{p(t)}\frac{p(t)}{t} \\ &= \frac{\lambda}{\sum_{i=T}^{M} q_i}\sum_{i=T}^{M} iq_i \quad \text{w.p. 1.} \end{aligned} \tag{16}$$

Note that the derivation of average throughput is valid only for stable networks. (The throughput in saturated cases will be discussed in the next section.) Thus the throughput maximization problem can be stated as

$$\max_{R>0} \quad Th(R) = \frac{\lambda}{\sum_{i=T}^{M} q_i}\sum_{i=T}^{M} iq_i, \tag{17}$$

subject to the stability condition:

$$\rho = \lambda\left(\frac{\mathbf{E}[X]}{\sum_{i=T}^{M} q_i} + \frac{\mathbf{E}[V]}{R}\right) < 1. \tag{18}$$

To have a more concrete understanding of $Th(R)$, define

$$g(R) \triangleq \frac{1}{\sum_{i=T}^{M} q_i}\sum_{i=T}^{M} iq_i. \tag{19}$$

We have the following results:

Lemma 3.1: For any $T < M$, $\sum_{i=T+1}^{M} q_i / \sum_{i=T}^{M} q_i$ is a monotonically decreasing function of R.
The proof is relegated to Appendix A.

Theorem 3.1: For any $T < M$, the function

$$g(R) = \frac{1}{\sum_{i=T}^{M} q_i}\sum_{i=T}^{M} iq_i \tag{20}$$

is a monotonically decreasing function of R .

Proof: Observe that

$$\begin{aligned} g(R) &= \frac{1}{\sum_{i=T}^{M} q_i}\sum_{i=T}^{M} iq_i \\ &= T + \frac{1}{\sum_{i=T}^{M} q_i}\left[\sum_{i=T+1}^{M} q_i + \cdots + \sum_{i=M}^{M} q_i\right]. \end{aligned} \tag{21}$$

Next, we prove by induction that each term in (21) is a monotonically decreasing function of R, i.e., for any $T < M$, and $1 \le m \le M - T$, $\sum_{i=T+m}^{M} q_i / \sum_{i=T}^{M} q_i$ is a monotonically decreasing function of R.
When $m = 1$, by Lemma 3.1, $\frac{\sum_{i=T+1}^{M} q_i}{\sum_{i=T}^{M} q_i}$ is a monotonically decreasing function of R.
Let $m = l$, $\frac{\sum_{i=T+l}^{M} q_i}{\sum_{i=T}^{M} q_i}$ is a monotonically decreasing function of R. Then for $m = l + 1$,

$$\frac{\sum_{i=T+l+1}^{M} q_i}{\sum_{i=T}^{M} q_i} = \frac{\sum_{i=T+l+1}^{M} q_i}{\sum_{i=T+l}^{M} q_i} \times \frac{\sum_{i=T+l}^{M} q_i}{\sum_{i=T}^{M} q_i}, \tag{22}$$

which is also monotonically decreasing of R. Hence the summation in (21) is also a monotonically decreasing function of R. This completes the proof. ■

Theorem 3.1 implies that for a given λ, the network achieves high throughput with low transmission rate. Hence, the transmitter should transmit the packet with the smallest rate as long as the stability condition (18) is satisfied. Our intuition is as follows. When the network is stable, the average number of transmitted packets is decided only by the arrival rate λ, instead of the transmission rate. Therefore, decreasing the transmission rate would only increase the ready probability for each receiver. As a consequence, the average number of successfully received packets would be increased. In summary, the maximum throughput is obtained when the transmitter multicasts packets with the smallest transmission rate that satisfies the stable condition in (18).

Therefore, given λ, the optimal transmission rate R_2^* maximizing the throughput in a stable network is given by

$$R_2^* = \arg\min_{R>0}\left\{\frac{\mathbf{E}[X]}{\sum_{i=T}^{M} q_i} + \frac{\mathbf{E}[V]}{R} \le 1/\lambda\right\}. \tag{23}$$

D. Optimal Rates for Throughput Maximization in Saturated Networks

If the network is saturated, (13) and (14) still hold. Since the idle period of the transmitter is zero, (15) is no longer valid and should be replaced by the following condition:

$$t = \sum_{i=1}^{p(t)} \frac{V_i}{R} + \sum_{j=1}^{b(t)-p(t)} X_j. \tag{24}$$

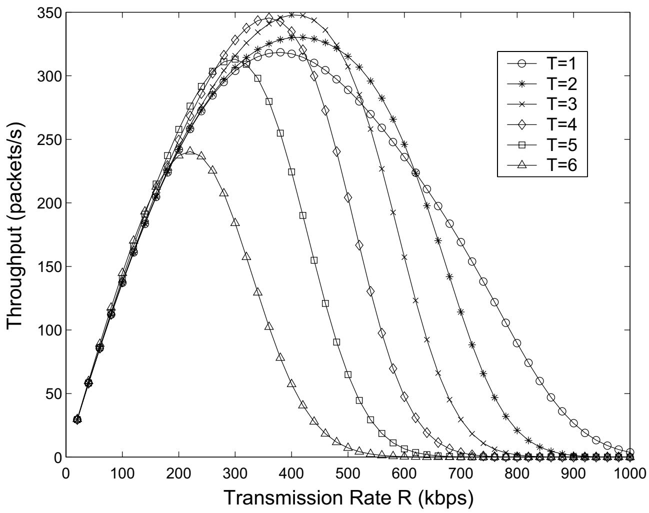


Fig. 3. Throughput in saturated networks as a function of transmission rate.

It follows that

$$\begin{aligned} Th(R) &= \lim_{t\to\infty} \frac{\sum_{i=T}^{M} ib_i(t)}{t} \\ &= \sum_{i=T}^{M} iq_i \frac{1}{\sum_{i=T}^{M} q_i} \lim_{t\to\infty} \frac{p(t)}{\sum_{i=1}^{p(t)} \frac{V_i}{R} + \sum_{j=1}^{b(t)-p(t)} X_j} \\ &= \frac{R\sum_{i=T}^{M} iq_i}{(\mathbf{E}[V] - R\mathbf{E}[X])\sum_{i=T}^{M} q_i + R\mathbf{E}[X]} \quad \text{w.p. 1.} \end{aligned} \tag{25}$$

The optimal transmission rate that maximizes the throughput in a saturated network can be found via numerical methods

$$R_3^* = \arg\max_{R>0} \left\{ \frac{R\sum_{i=T}^{M} iq_i}{(\mathbf{E}[V] - R\mathbf{E}[X])\sum_{i=T}^{M} q_i + R\mathbf{E}[X]} \right\}. \tag{26}$$

The throughput in a saturated network, as a function of the transmission rate, is plotted in Fig. 3. We should emphasize that the scheme in Fig. 3 may not be reliable without using higher-layer recovering schemes. In Section IV, we propose a reliable cross-layer scheme that jointly optimizes the transmission rate R and the multicast threshold T.

E. Tradeoff between Stability and Throughput

In unicast cases, maximizing the throughput is equivalent to maximizing the stability region. However, in multicast cases, the number of successfully received packets per transmission can be any integer between T and M [2]. As a result, the equivalence may not hold.

Rewrite the optimal transmission rate R_1^* defined in (12) that maximizes the stability region:

$$\begin{aligned} R_1^* &= \arg\min_{R>0} \left\{ \frac{\mathbf{E}[X]}{I_{(P)}(T, M-T+1)} + \frac{\mathbf{E}[V]}{R} \right\} \\ &\triangleq \arg\min_{R>0} \{f(R)\}. \end{aligned} \tag{27}$$

Recall that from (17), the optimal transmission rate R_4^* that maximizes the throughput is given by

$$\begin{aligned} R_4^* &= \arg\max_{R>0} \left\{ \lambda \frac{\sum_{i=T}^{M} iq_i}{\sum_{i=T}^{M} q_i} \right\} \\ &= \arg\max_{R>0} \left\{ \frac{1}{\frac{\mathbf{E}[X]}{I_{(P)}(T,M-T+1)} + \frac{\mathbf{E}[V]}{R}} \times \frac{\sum_{i=T}^{M} iq_i}{\sum_{i=T}^{M} q_i} \right\} \\ &= \arg\max_{R>0} \left\{ \frac{g(R)}{f(R)} \right\}. \end{aligned} \tag{28}$$

Proposition 3.1: For any $T < M$, $R_1^* > R_4^*$.

Proof: It is clear that for any $T < M$, $R_1^* \neq R_4^*$. Assume $R_1^* < R_4^*$. Then by Theorem 3.1, $g(R_1^*) \geq g(R_4^*)$. By (27), $f(R_1^*) < f(R_4^*)$. Thus, $f(R_1^*)/g(R_1^*) < f(R_4^*)/g(R_4^*)$. This contradicts the definition of R_4^* in (28), where R_4^* minimizes $f(R)/g(R)$. Hence $R_1^* > R_4^*$. ■

Proposition 3.1 reveals that maximum stability region and maximum throughout cannot be achieved simultaneously when $T < M$. Interestingly, when $T = M$, throughput becomes a linear function of λ_{max}. Thus the stability and the throughput can both reach maximum at the same time.

F. Some Generalizations

1) Networks with MIMO Links: Next we extend the study on rate optimization to network models with MIMO links. Assume that the transmitter has T_s transmit antennas and each receiver has T_r receive antennas. Let $\mathbf{H}_i$ denote the $T_r \times T_s$ MIMO channel matrix, and all the entries in $\mathbf{H}_i$ are independently complex Gaussian with zero mean. Then the capacity of the channel between the transmitter and the i-th receiver with a given $\mathbf{H}_i$ is [21]

$$C_i = \log\det(\mathbf{I}_M + \frac{P}{T_s N_0}\mathbf{H}_i{}^*\mathbf{H}_i). \tag{29}$$

In general, since $\mathbf{H}_i$ is a random variable, the capacity is also random. It is shown in [5] that when T_s or T_r (or both) is large, the distribution of C_i approaches to Gaussian distribution. Furthermore, even when T_s and T_r are small, the Gaussian approximation is still accurate. The mean and variance of the distribution are only decided by T_s, T_r, and P/N_0. Then the outage probabilities can be easily predicted and computed by this approximation.

It is easy to see that the capacity C_i defined in (29) is positive. Therefore, C_i can be approximated more accurately by a truncated Gaussian distribution. Let $\bar{C}_i$ denote the "un-truncated" Gaussian random variable $\bar{C}_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ and $\hat{p}_i$ denote the probability that the i-th receiver is ready to receive in MIMO link, then

$$\hat{p}_i = Pr\{\bar{C}_i \geq R | \bar{C}_i > 0\} = \frac{Q(\frac{R-\mu_i}{\sigma_i})}{1 - Q(\frac{\mu_i}{\sigma_i})},\ i = 1, \cdots, M, \tag{30}$$

where $Q(\cdot)$ is defined as

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt. \tag{31}$$

Thus the rate optimization in MIMO networks can be solved by replacing the ready probability p in (12) and (23) with $\hat{p}$.

2) Networks with Non-identical Links: In the above, we assume that the links between the transmitter and each receiver are i.i.d.. In general, however, they may not be identical because of the different distances between transmitter and receivers. To be more accurate, we assume that $\{C_i\}$ are independent but non-identical distributed random variables. Then (9) can be rewritten as ([3])

$$P_t = \sum_{i=T}^{M} q_i = \sum_{i=T}^{M} \sum_{S_i} \prod_{l=1}^{i} p_{j_l} \prod_{l=i+1}^{M} [1 - p_{j_l}], \tag{32}$$

where the summation S_i extends over all permutations $(j_1, j_2, \cdots, j_M)$ of $1, 2, \cdots, M$ for which $j_1 < \cdots < j_i$ and $j_{i+1} < \cdots < j_M$. And p_{j_l} is defined in (3). In light of the complexity of (32), we present a bound on P_t in the following proposition.

Proposition 3.2: For all T,

$$I_{(p_{min})}(T, M - T + 1) \le P_t \le I_{(p_{max})}(T, M - T + 1),$$

where $p_{min} = \min\{p_i,\ i = 1, 2, \cdots, M\}$, $p_{max} = \max\{p_i,\ i = 1, 2, \cdots, M\}$, and $I_p(a, b)$ is the incomplete Beta function.

The proof is relegated to Appendix B.

We then can apply the derivation in Section III-B to the non-identical case, and give a boundary to the optimal transmission rate.

3) Channels with Memory: In the studies above, channels are assumed to be independent for each time slot. In general, the channel conditions are correlated. In this section, we use the concept of *level crossing* to derive the optimal transmission rate for correlated fading channels.

Based on [5], we approximate the channel between the transmitter and the i-th receiver as a random process $\{C_i(t) \ge 0\}$ with mean μ_i and covariance function $r_i(\tau)$. It follows that

$$\begin{aligned} Pr(\text{receiver } i \text{ is ready}) &= Pr(C_i \ge R) \\ &= V_R \times \mathbf{E}[u] \quad \text{w.p. } 1, \end{aligned} \tag{33}$$

where V_R is the up-crossing rate and $\mathbf{E}[u]$ is the average length between one up-crossing and its successive down-crossing. If the process is stationary and ergodic, by [6],

$$\mathbf{E}[u] = \frac{2}{a} Pr(C_i(0) \ge R), \tag{34}$$

where $a = \mathbf{E}v(0, 1]$, and $v(0, t]$ denotes the crossing times between 0 and t. Then

$$V_R = \frac{1}{2} \mathbf{E}v(0, 1]. \tag{35}$$

Equation (33) can be rewritten as

$$Pr(C_i \ge R) = Pr(C_i(0) \ge R). \tag{36}$$

which is exactly the same as (3). It is not surprising since the average behavior of the random process is not related to the covariance function. Hence all the conclusions in previous sections can be used in the correlated channel case.

IV. Enhancing Multicast Reliability via Erasure Coding

In the previous sections, we have studied the rate optimization for MAC layer multicast. A multicast threshold T was set to achieve the tradeoff between stability and reliability. However, since retransmission is not done at the MAC layer protocol, the *Threshold-T policy* is not reliable, in the sense that if $T < M$, some receivers may miss some packets needed. In this section, we study the erasure coding at transport layer to enhance multicast reliability.

Suppose that the coding scheme produces $n = h + k$ encoded packets from k original packets, where $h > 0$ is the degree of redundancy. A receiver can reconstruct the original k data packets once it successfully receives at least k encoded packets [14]. The encoding/decoding algorithms with this property can be found in [1], [9].

In *Threshold-T* based wireless multicast networks, the reliable transmission can be realized by the following scheme. Before each transmission, the transmitter first queries the channel and sends an encoded packet when at least T receivers are ready ([2]). It would keep transmitting until all receivers successfully receive k encoded packets. Then each receiver can decode the data independently based on its locally received data only. It is clear that T would affect the performance of the coding scheme. Thus, in this section, we will characterize the optimal multicast threshold T that maximizes the performance of the reliable multicast scheme.

Let $W(M)$ denote the number of ready receivers in totally M receivers. Then for a given T, when the transmitter sends a packet, a receiver can successfully receive the packet if it is ready to receive and there are at least $T - 1$ ready receivers among other $M - 1$ receivers. Thus, the probability that a particular receiver receives a packet successfully under a transmission is given by

$$\begin{aligned} p_r &= Pr\{\text{A receiver receives a packet successfully} \\ &\quad |\textit{At least } T \textit{ receivers are ready}\} \\ &= pPr\{W(M-1) \ge T-1\}/Pr\{W(M) \ge T\} \\ &= \frac{pI_p(T-1, M-T+1)}{I_p(T, M-T+1)}. \end{aligned} \tag{37}$$

where p is defined in (4). Let n_j denote the number of transmissions when user j receives its kth encoded packets, and $F(i)$ denote the probability that $n_j \le i$, which is given by

$$\begin{aligned} F(i) &\triangleq Pr\{n_j \le i\} \\ &= \sum_{m=k}^{i} \binom{i}{m} p_r^m [1 - p_r]^{i-m} \\ &= I_{(p_r)}(k, i - k + 1). \end{aligned} \tag{38}$$

Then, the total number of transmissions to provide reliable multicast is given by $n_{max} = \max\{n_1, n_2, \cdots, n_M\}$. Appealing to (3.3.4) in [3], we have

$$\mathbf{E}[n_{max}] = \sum_{i=k}^{\infty} \left\{1 - [I_{(p_r)}(k, i - k + 1)]^M\right\} + k. \tag{39}$$

Since redundant packets are generated to provide fully reliable multicast, the actual number of transmitted packets is increased. Therefore, the equivalent packet arrival rate λ' should be higher, which is given by

$$\begin{aligned} \lambda' &= \lambda \frac{\mathbf{E}[n_{max}]}{k} \\ &= \frac{\lambda}{k} \sum_{i=k}^{\infty} \left\{ 1 - [I_{(p_r)}(k, i-k+1)]^M \right\} + \lambda, \quad (40) \end{aligned}$$

and the stability condition should be rewritten as

$$\lambda' \le 1 / \left(\frac{\mathbf{E}[X]}{P_t} + \frac{\mathbf{E}[V]}{R} \right). \quad (41)$$

Hence, the optimization problem can be revised as

$$\begin{aligned} \max_{(T,\ R)} \quad & Th(T,R) = \frac{\lambda}{\sum_{i=T}^{M} q_i} \sum_{i=T}^{M} i q_i, \quad (42) \\ \text{s.t.} \quad & \lambda' \left(\frac{\mathbf{E}[X]}{\sum_{i=T}^{M} q_i} + \frac{\mathbf{E}[V]}{R} \right) \le 1, \\ & T \in \mathcal{N}, \quad R > 0. \end{aligned}$$

It can be seen that the equivalent packet arrival rate λ' is a function of T. The corresponding cross-layer optimization boils down to a joint rate and multicast threshold optimization problem. In what follows, we propose to use a two-step algorithm to find an optimal (T, R) pair for (42):

Algorithm 1

1: For any $T = 0, 1, \cdots, M$, its corresponding optimal rate $R^*(T)$ can be found by

$$R^*(T) = \arg\min_{R>0} \left\{ \frac{\mathbf{E}[X]}{\sum_{i=T}^{M} q_i} + \frac{\mathbf{E}[V]}{R} \le 1/\lambda' \right\}. \quad (43)$$

2: Among all elements in the set $\{(T, R^*(T))\}$, find the optimal one (T^*, R^*) that has the largest throughput defined in (42).

The performance of the proposed scheme is shown in Fig. 4. However, in a practical setting, continuous rate selection may not be possible, and the rate is quantized to discrete values, denoted by $\{U_i, i = 1, 2, \cdots, L\}$ with $U_1 < U_2 < \cdots < U_L$. Since both T and R are discrete values, one can do exhaustive search to find the optimal pair (T^*, R^*).

Note that in the above packet level erasure coding scheme, the number of coded packets may grow unbounded. In a practical setting, however, the number of coded packets is bounded, say by D_0. Then, for any given $D_0 \ge k$, the outage probability is given by

$$\begin{aligned} Pr\{n_{max} > D_0\} &= 1 - Pr\{n_{max} \le D_0\} \\ &= 1 - F(D_0)^M \\ &= 1 - \left[I_{(p_r)}(k, D_0 - k + 1) \right]^M. \end{aligned}$$

It is clear that the outage probability decreases fast as D_0 increases, especially when T and p_r are reasonably large, as illustrated in Fig. 5. In case if $n_{max} > D_0$, some receives may not receive enough encoded packets to decode the original data packets successfully, resulting in degradation in reliability.

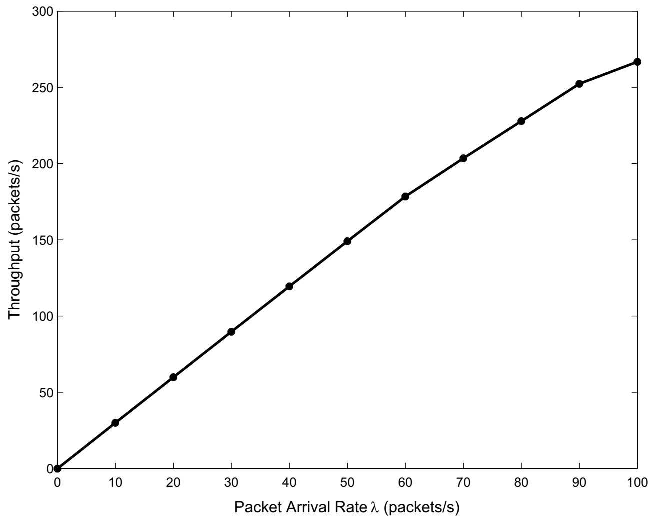


Fig. 4. Throughput for the two-threshold reliable multicast ($M = 6$).

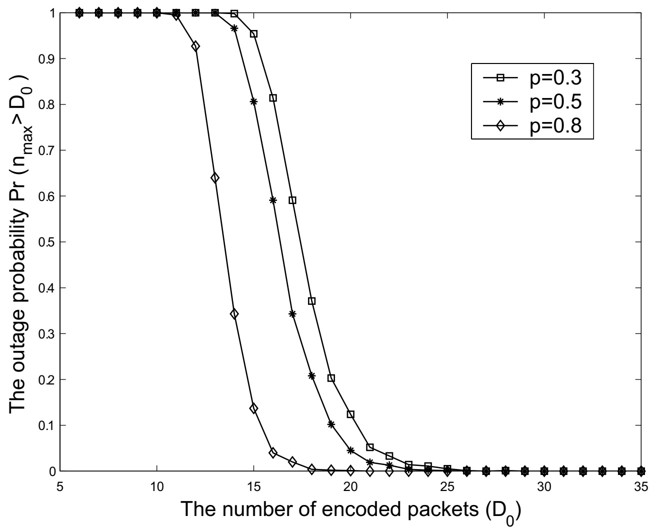


Fig. 5. Outage probability vs. D_0 ($M = 6$, $k = 10$).

Fig. 6 depicts the performance degradation if the transport layer FEC coding scheme is absent, where the MAC curve shows the performance of the scheme by retransmitting lost packets to guarantee reliability. We observe that the cross-layer scheme proposed in this paper always outperforms the conventional feedback-retransmission scheme.

We note that the proposed protocol is sub-optimal in the sense that it studies single multicast threshold policies. If a two threshold policy (T, q) is considered (e.g., [2]), i.e., the threshold is set to be T with probability q, and $T + 1$ with probability $1 - q$, the corresponding scheme should lead to better performance.

V. Conclusions and Future Work

In this paper, we proposed a cross-layer optimization scheme for wireless multicast networks. We first studied the optimal rates for throughput maximization at the MAC layer multicast. We examined the stability region for multicast policies with a pre-determined multicast threshold, and characterized the optimal transmission rates that maximize

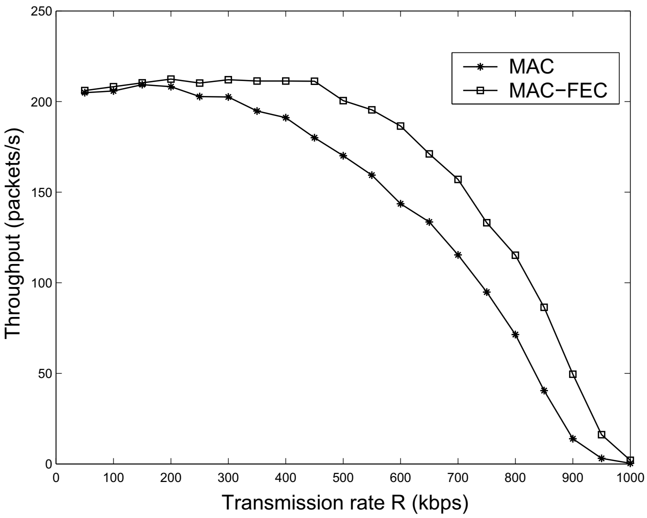


Fig. 6. Performance degradation without transport layer FEC coding.

the throughput in stable networks and in saturated networks, respectively. Then, we analyzed the tradeoff between the stability and the throughput. Furthermore, we extended our studies to non-i.i.d. link cases and MIMO link cases. To meet the requirement that no retransmission is needed at the MAC layer, we used transport layer erasure coding to enhance the reliability and provided an optimal solution to the corresponding cross-layer optimization problem.

We note that the proposed cross-layer approach could be useful for the downlink of a cellular network. More work is needed to generalize the proposed scheme to take into account channel contention in multi-hop wireless networks. We are currently investigating on energy efficient multicast for battery-operated wireless networks, and the security issue therein.

APPENDIX I
PROOF OF LEMMA 3.1

By (10) and [18], we have

$$\begin{aligned}\frac{\sum_{i=T+1}^{M} q_i}{\sum_{i=T}^{M} q_i} &= \frac{I_p(T+1, M-T)}{I_p(T, M-T+1)} \\ &= \frac{B_p(T+1, M-T)\Gamma(M+1)}{\Gamma(T+1)\Gamma(M-T)} \\ &\quad \times \frac{\Gamma(T)\Gamma(M-T+1)}{B_p(T, M-T+1)\Gamma(M+1)} \\ &= \frac{B_p(T+1, M-T)(M-T)}{TB_p(T, M-T+1)}, \end{aligned} \quad \text{(I.1)}$$

where $B_p(a, b)$ is defined as

$$\begin{aligned} B_p(a, b) &\triangleq \frac{I_p(a, b)\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \\ &= \int_0^p t^{a-1}(1-t)^{b-1} dt. \end{aligned} \quad \text{(I.2)}$$

Then

$$\begin{aligned} &\frac{\partial \frac{\sum_{i=T+1}^{M} q_i}{\sum_{i=T}^{M} q_i}}{\partial R} \\ &= \left[\frac{p^T(1-p)^{M-T-1} B_p(T, M-T+1)}{[B_p(T, M-T+1)]^2} \right. \\ &\quad \left. - \frac{p^{T-1}(1-p)^{M-T} B_p(T+1, M-T)}{[B_p(T, M-T+1)]^2} \right] \\ &\quad \times \frac{M-T}{T} \frac{\partial p}{\partial R} \\ &= \frac{pB_p(T, M-T+1) - (1-P)B_p(T+1, M-T)}{[B_p(T, M-T+1)]^2} \\ &\quad \times \frac{M-T}{T} \frac{\partial p}{\partial R} p^{T-1}(1-p)^{M-T-1}. \end{aligned} \quad \text{(I.3)}$$

By the property of $B_P(a, b)$ in [18] that

$$B_p(a, b) = B_p(a+1, b) + B_p(a, b+1), \quad \text{(I.4)}$$

we have

$$\begin{aligned} & pB_p(T, M-T+1) - (1-p)B_p(T+1, M-T) \\ =\ & pB_p(T, M-T) - B_p(T+1, M-T) \\ =\ & p\int_0^p t^{T-1}(1-t)^{M-T-1} dt - \int_0^p t^T(1-t)^{M-T-1} dt \\ >\ & 0. \end{aligned} \quad \text{(I.5)}$$

Recall the definition of p in (3) and (4), we have

$$\frac{\partial p}{\partial R} < 0. \quad \text{(I.6)}$$

Hence, (I.3) is always negative, which implies that $\frac{\sum_{i=T+1}^{M} q_i}{\sum_{i=T}^{M} q_i}$ is a monotonically decreasing function of R. ∎

APPENDIX II
PROOF OF PROPOSITION 3.2

$$\begin{aligned} \sum_{i=T}^{M} q_i &= Pr\{\text{at least T receivers are ready to receive}\} \\ &= Pr\{C_{(T)} \geq R\}, \end{aligned} \quad \text{(II.1)}$$

where $C_{(1)} \geq C_{(2)} \geq \cdots \geq C_{(M)}$ is the ordered random variates of the independent, non-identical distributed random variates $\{C_i\},\ i = 1, 2 \cdots M$. Since

$$\begin{aligned} Pr\{C_{(1)} \geq R\} &= 1 - Pr\{\text{no receiver is ready}\} \\ &= 1 - \prod_{i=1}^{M}(1-p_i) \\ &\leq 1 - (1-p_{max})^M \\ &= Pr\{X_{(1)} \geq R\}, \end{aligned} \quad \text{(II.2)}$$

where $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(M)}$ is the ordered i.i.d. random variables with $Pr\{C > R\} = p_{max}$. It follows that

$$C_{(1)} \leq_{st} X_{(1)}, \quad \text{(II.3)}$$

where '$\leq_{st}$' stands for *stochastically smaller*. By Theorem 5.2.2 of [3], we have

$$C_{(r)} \leq_{st} X_{(r)}, \quad r = 1, 2, \cdots, M. \quad \text{(II.4)}$$

Hence

$$\begin{aligned}\sum_{i=T}^{M} q_i &= Pr\{C_{(T)} \geq R\} \\ &\leq Pr\{X_{(T)} \geq R\} \\ &= \sum_{i=T}^{M} \binom{M}{i} p_{max}^{i} [1 - p_{max}]^{M-i} \\ &= I_{(p_{max})}(T, M - T + 1), \qquad \text{(II.5)}\end{aligned}$$

and equality is achieved if $\{C_i\}$ are i.i.d.. Similarly, we can show that

$$\sum_{i=T}^{M} q_i \geq I_{(p_{min})}(T, M - T + 1), \qquad \text{(II.6)}$$

and equality is achieved if $\{C_i\}$ are i.i.d.. ∎

References

[1] J. Byers, M. Luby, and M. Mitzenmacher, “A digital fountain approach to asynchronous reliable multicast,” *IEEE J. Sel. Areas Commun.*, vol. 20, no. 8, pp. 1528-1540, Oct. 2002.

[2] P. Chaporkar and S. Sarkar, “Wireless multicast: theory and approaches,” *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1954-1972, June 2005.

[3] H. A. David and H. N. Nagaraja, *Order Statistics, Third Edition*. Wiley Series in Probability and Statistics, 2003.

[4] P. K. Gopala and H. E. Gamal, “On the throughput-delay tradeoff in cellular multicast,” in *Proc. WirelessCom’05*, vol. 2, pp. 1401-1406, June 2005.

[5] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, “Multi-antenna channel-hardening and its implications for rate feedback and scheduling,” *Eur. Trans. Telecom.*, vol. 10, pp. 585-595, Mar. 2004.

[6] I. N. Kovalenko, N. Yu. Kuznetsov, and V. M. Shurenkov, *Models of Random Processes*. CRC Press, 1996.

[7] S. Lee, M. Gerla, and C. Chiang, “On-demand multicast routing protocol,” in *Proc. IEEE WCNC’99*, pp. 1298-1302.

[8] F. Li and I. Nikolaidis, “On minimum-energy broadcasting in all-wireless networks,” in *Proc. LCN’01*, pp. 193-202.

[9] F. MacWilliams and N. Sloane, *The Theory of Error Correcting Codes*. New York: North-Holland, 1977.

[10] J. Metzner, “An improved broadcast retransmission protocol,” *IEEE Trans. Commun.*, vol. 32, no. 6, pp. 679-683, June 1984.

[11] M. Mitzenmacher, “Digital fountains: a survey and look forward,” in *Proc. IEEE Information Theory Workshop 2004*, pp. 271-276.

[12] N. Nikaein, H. Labiod, and C. Bonnet, “MA-FEC: a QoS-based adaptive FEC for multicast communication in wireless networks,” in *Proc. ICC 2000*, pp. 954-958.

[13] J. Nonnenmacher, E. Biersack, and D. Towsley, “Parity-based loss recovery for reliable multicast transmission,” *IEEE/ACM Trans. Networking*, vol. 6, no. 4, pp. 349-361, Aug. 1998.

[14] J. Park, “Efficient multicast stream authentication using erasure codes,” *ACM Trans. Inf. Syst. Sec.*, vol. 6, no. 2, pp. 258-285, May 2003.

[15] S. Ross, *Introduction to Probability Models, Seventh Edition*. Harcourt Academic Press, 2000.

[16] E. Royer and C. Perkins, “Multicast operation of the ad hoc on-demand distance vector routing protocol,” in *Proc. MOBICOM’99*, pp. 207-218, Aug. 1999.

[17] D. Shiu, G. J. Foschini, and M. J. Gans, “Fading correlation and its effect on the capacity of multielement antenna systems,” *IEEE Trans. Commun.*, vol. 48, no. 3, March 2000.

[18] J. Spanier and K. B. Oldham, *An Atlas of Functions*. Hemisphere Publishing Corporation, 1987.

[19] K. Tang and M. Gerla, “MAC layer broadcast support in 802.11 wireless networks,” in *Proc. MILCOM’00*, vol. 1 pp. 544-548.

[20] K. Tang and M. Gerla, “MAC reliable broadcast in ad hoc networks,” in *Proc. MILCOM’01*, vol. 2 pp. 1008-1013.

[21] I. E. Telatar, “Capacity of multi-antenna Gaussian channels,” *Eur. Trans. Telecom.*, vol. 10, pp. 585-595, Nov. 1999.

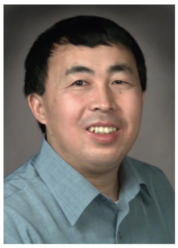
[22] J. Wieselthier, G. Nguyen, and A. Ephermides, “On the construction of energy-efficient broadcast and multicast trees in wireless network,” in *Proc. INFOCOM’00*, pp. 585-594.

[23] J. Wieselthier, G. Nguyen, and A. Ephremides, “Algorithm for bandwidth limited energy-efficient wireless broadcasting and multicasting,” in *Proc. MILCOM 2000*, vol. 1, pp. 485-490.

Weiyan Ge received his B.S. in the Department of Electrical Engineering from Peking University, Beijing, China, in 2001, and his M.Phil degree in the Department of Information Engineering from the Chinese University of Hong Kong, Hong Kong, China, in 2003. Currently, he is a Ph.D. student in the Department of Electrical Engineering, Arizona State University, Tempe, AZ. His research interests include wireless multicast and cooperative communications in wireless sensor networks.

Junshan Zhang was born in September 1972. He received his Ph.D. degree from the School of Electrical and Computer Engineering at Purdue University in 2000. He joined the Department of Electrical Engineering at Arizona State University in August 2000, where he is currently an Associate Professor. His research interests fall in the general area of wireless networks, spanning from the networking layer to the physical layer. His current research focuses on fundamental problems in wireless ad-hoc networks and sensor networks, including cross-layer optimization and design, network management, network information theory, stochastic analysis.

He is a recipient of the ONR Young Investigator Award in 2005 and the NSF CAREER award in 2003. He has also received the Outstanding Research Award from the IEEE Phoenix Section in 2003. He was chair of the IEEE Communications and Signal Processing Phoenix Chapter from Jan. 2001 to Dec. 2003. He has served as a TPC co-chair for IPCCC 2006 and TPC vice chair for ICCCN 2006, and will be the general chair for IEEE Communication Theory Workshop 2007. He has been on the technical program committees for many conferences, including INFOCOM, SECON, GLOBECOM, ICC, MOBIHOC, BROADNETS, and SPIE ITCOM. He has been an Associate Editor for *IEEE Transactions on Wireless Communications* since 2004.



Xuemin (Sherman) Shen received the B.Sc. (1982) degree from Dalian Maritime University (China) and the M.Sc. (1987) and Ph.D. (1990) degrees from Rutgers University, New Jersey (USA), all in electrical engineering. From September 1990 to September 1993, he was first with the Howard University, Washington DC, and then the University of Alberta, Edmonton (Canada). Since October 1993, he has been with the Department of Electrical and Computer Engineering, University of Waterloo, Canada, where he is a Professor and the Associate Chair for Graduate Studies.

Dr. Shen’s research focuses on mobility and resource management in interconnected wireless/Internet interworking, UWB, WiFi/WiMAX, sensor and ad hoc wireless networks. He is a coauthor of three books, and has over 200 publications in wireless communications and networks, control and filtering. Dr. Shen received the Outstanding Performance Award in 2004 from the University of Waterloo, the Premier’s Research Excellence Award (PREA) in 2003 from the Province of Ontario for demonstrated excellence of scientific and academic contributions, and the Distinguished Performance Award from the Faculty of Engineering, University of Waterloo, for outstanding contributions in teaching, scholarship and service. He serves as one of the General/Technical Chairs, for QShine’06, IWCMC’06, IEEE ISSPIT’06, IEEE Broadnet’05, QShine’05, IEEE WirelessCom’05, IFIP Networking 2005, ISPAN’04, and IEEE Globecom’03 Symposium on Next Generation Networks and Internet. He also serves as the Associate Editor for *IEEE Transactions Wireless Communications*, *IEEE Transactions on Vehicular Technology*, *ACM Wireless Networks*, *Computer Networks*, and *WCMC* (Wiley), among others. He also serves as Guest Editor for *IEEE Wireless Communications*, *IEEE Journal on Selected Areas in Communications*, and *IEEE Communications Magazine*. Dr. Shen is a senior member of the IEEE, and a registered Professional Engineer of Ontario, Canada.