

## Optimized Opportunistic Multicast Scheduling (OMS) over Wireless Cellular Networks

Tze-Ping Low, Man-On Pun, Y.-W. Peter Hong, C.-C. Jay Kuo

TR2010-008 March 2010

### Abstract

Optimized opportunistic multicast scheduling (OMS) is studied for cellular networks, where the problem of efficiently transmitting a common set of fountain-encoded data from a single base station to multiple users over quasi-static fading channels is examined. The proposed OMS scheme better balances the tradeoff between multiuser diversity and multicast gain by transmitting to a subset of users in each time slot using the maximal data rate that ensures successful decoding by these users. We first analyze the system delay in homogeneous networks by capitalizing on extreme value theory and derive the optimal selection ratio (i.e., the portion of users that are selected in each time slot) that minimizes the delay. Then, we extend results to heterogeneous networks where users are subject to different channel statistics. By partitioning users into multiple approximately homogeneous rings, we turn a heterogeneous network into a composite of smaller homogeneous networks and derive the optimal selection ratio for the heterogeneous network. Computer simulations confirm theoretical results and illustrate that the proposed OMS can achieve significant performance gains in both homogeneous and heterogeneous networks as compared with the conventional unicast and broadcast scheduling.

*IEEE Transactions on Wireless Communications*

This work may not be copied or reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of Mitsubishi Electric Research Laboratories, Inc.; an acknowledgment of the authors and individual contributions to the work; and all applicable portions of the copyright notice. Copying, reproduction, or republishing for any other purpose shall require a license with payment of fee to Mitsubishi Electric Research Laboratories, Inc. All rights reserved.



# Optimized Opportunistic Multicast Scheduling (OMS) over Wireless Cellular Networks

Tze-Ping Low, *Student Member, IEEE*, Man-On Pun, *Member, IEEE*,  
Y.-W. Peter Hong, *Member, IEEE*, and C.-C. Jay Kuo, *Fellow, IEEE*

**Abstract**—Optimized opportunistic multicast scheduling (OMS) is studied for cellular networks, where the problem of efficiently transmitting a common set of fountain-encoded data from a single base station to multiple users over quasi-static fading channels is examined. The proposed OMS scheme better balances the tradeoff between multiuser diversity and multicast gain by transmitting to a subset of users in each time slot using the maximal data rate that ensures successful decoding by these users. We first analyze the system delay in homogeneous networks by capitalizing on extreme value theory and derive the optimal selection ratio (i.e., the portion of users that are selected in each time slot) that minimizes the delay. Then, we extend results to heterogeneous networks where users are subject to different channel statistics. By partitioning users into multiple approximately homogeneous rings, we turn a heterogeneous network into a composite of smaller homogeneous networks and derive the optimal selection ratio for the heterogeneous network. Computer simulations confirm theoretical results and illustrate that the proposed OMS can achieve significant performance gains in both homogeneous and heterogeneous networks as compared with the conventional unicast and broadcast scheduling.

## I. INTRODUCTION

The ever-increasing demand for richer multimedia contents has driven the development of highly efficient content distribution technologies over wireless networks. In particular, opportunistic scheduling has recently emerged as one of the most promising techniques for content delivery. These methods effectively increase throughput with reduced delay by exploiting multiuser diversity inherent in wireless networks [1]–[3]. Most previous research on opportunistic scheduling has focused on applications where the base station (BS) schedules *different* data to each user. Here, we concentrate on broadcast

applications in a single-cell system where the BS sends a common set of data to multiple users undergoing independent channel fading. In fact, with the increased popularity of commercial services such as mobile TV, up-to-date information distribution services (e.g., news, traffic, stock market *etc*), and file distribution (e.g., applications, software upgrades *etc*) [4], [5], the need for such scheduling schemes has generated strong interests in recent years. In this work, our goal is to develop efficient scheduling schemes for such applications by exploiting both multiuser diversity and multicast gain.

We consider the problem of transmitting a common set of data to all users on the downlink of a wireless cellular network. To achieve this task, two scheduling schemes have been studied extensively in the literature. They are the opportunistic unicast scheduling scheme and the broadcasting scheme. The former exploits multiuser diversity by limiting the BS to transmit to the user with the best instantaneous channel in each time slot at the user's highest supportable data rate. In this case, only one user is served at one time and the transmission must be repeated multiple times until all users receive the whole message. In contrast, the latter exploits the broadcast nature of the wireless medium by transmitting at a rate that is supported by all users in the network. In this case, all users are served at once but the rate is constrained by the user with the poorest channel conditions in order to prevent channel outages.

Being motivated by recent advances in erasure codes and fountain codes [6], [7], opportunistic multicast scheduling (OMS) schemes have been proposed recently to better balance the tradeoff between multiuser diversity and the multicast gain [8]–[11]. Assuming that each user experiences independent fading over different time slots, the OMS schemes select a different group of users for service in each time slot at a rate supported by the worst user in the selected group. The adoption of erasure and fountain codes allows users to recover the full original content of a message once a minimum set of encoded symbols is received, regardless of the specific received sequence of encoded symbols. This is in sharp contrast with conventional schemes without erasure or fountain codes where the BS must keep track of the data that each user has received throughout the content delivery process. This is obviously a huge burden for networks with a large number of users. The work in [8], [9] adopted the median OMS scheme where the best half of the users are served in each time slot. Despite its good performance as compared to unicast and broadcast, the optimality of fixing the selection ratio at 50% has not been clearly addressed.

The main contribution of this work is to derive the optimal

Manuscript received March 16, 2009; revised August 27, 2009; accepted October 25, 2009. The associate editor coordinating the review of this paper and approving it for publication was F. A. Cruz-Perez. This research was supported in part by the Croucher Foundation under a post-doctoral fellowship, and in part by the National Science Council, Taiwan, under grant 97-2219-E-007-001. This work was presented in part at the IEEE Globecom 2008, New Orleans, LA, U.S.A, November 2008 and in part at the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2009, Taiwan, April 2009.

Tze-Ping Low and C.-C. Jay Kuo are with the Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089, USA (e-mails: tlow@usc.edu; cckuo@sipi.usc.edu).

Man-On Pun is with the Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA 02139, USA. This work was done when he was a Croucher post-doctoral research fellow at Princeton University, Princeton, NJ (e-mail: mpun@merl.com).

Y.-W. Peter Hong is with the Institute of Communications Engineering National Tsing Hua University, Hsinchu 30013, Taiwan (e-mail: ywhong@ee.nthu.edu.tw).

Digital Object Identifier xxx

user selection ratio that minimizes the total number of time slots required for all users to successfully receive the common set of messages. We first focus on homogeneous networks where all users experience independent and identically distributed (i.i.d.) channel conditions. In this case, the optimal user selection ratio can be derived analytically using central order statistics and extreme value theory [12]. Next, we extend results to heterogeneous networks where users are assumed to be randomly distributed in a cellular network. Clearly, the system performance is constrained by the cell-edge users that have the worst average channel conditions. By exploiting this observation, we devise OMS schemes with the absolute and the normalized instantaneous SNR criteria by optimizing the performance of cell-edge users.

Before proceeding further, some characteristics differentiating our work from other existing studies on OMS [10], [11] and adaptive transmission schemes [13] should be emphasized. Specifically, Ge *et al.* [10] proposed a threshold-based scheme that transmits in the current time slot only if more than  $G$  users (where  $G$  is a pre-defined number) can support a predefined rate  $R$ . In contrast, our scheme investigates the optimal user selection ratio with an adaptive target data rate in each time slot. Another closely related work was conducted by Kozat [11], where the performance of OMS is studied in terms of system throughput by considering infinite encoded messages. The goal of our current research is to minimize the time duration required for transmitting a finite-length message to all users. By capitalizing on extreme value theory, we are able to provide theoretical analysis on the system performance. Finally, despite the fact that both our work and [13] explore the tradeoff between multicast and unicast schemes, our work develops a unified theoretical framework to compute the optimal percentage of users served in each transmission regardless of unicast or multicast design.

The rest of this paper is organized as follows. We first introduce the system model in Section II. Then, under the assumption that all users experience i.i.d. fading, OMS with optimized user selection ratios is proposed to minimize the system delay for homogeneous networks in Section III. By extending results in Section III, we develop OMS for heterogeneous networks in Section IV. Computer simulation results are shown in Section V. Finally, concluding remarks are given in Section VI.

The following convention is adopted in this paper:  $\mathcal{N}(\mu, \sigma^2)$  and  $\mathcal{CN}(\mu, \sigma^2)$  being the real- and complex-valued Gaussian distributions with mean  $\mu$  and variance  $\sigma^2$  while  $\log(\cdot)$  and  $\ln(\cdot)$  denoting the logarithm operators of base 2 and  $e$ , respectively, and  $\lfloor \cdot \rfloor$  being the largest integer no larger than the enclosed quantity. Finally,  $\mathbf{E}[\cdot]$  and  $\mathbf{Var}[\cdot]$  denote the expected value and the variance of the enclosed random variable, respectively.

## II. SYSTEM MODEL

Consider a single-antenna downlink cellular network where the BS is to transmit a common message to  $N$  users in the network. We examine a time-slotted system where the length of each time-slot,  $T$ , is comparable with the channel

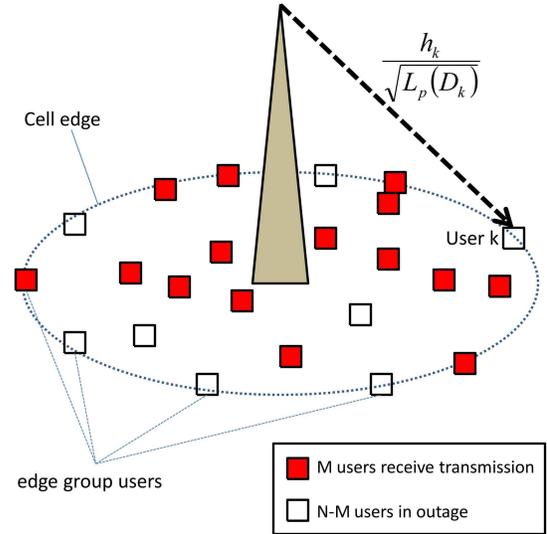


Fig. 1. Illustration of opportunistic multicast scheduling (OMS).

coherence time. As a result, the channel coefficients are assumed to remain constant throughout each time slot but vary independently from one time slot to another. Let  $h_n(k)$  be the instantaneous channel coefficient between the BS and the  $n$ -th user in the  $k$ -th time slot. We assume that  $\{h_n(k), \forall n, k\}$  are independent among users and i.i.d. over different time slots. To distribute the common message to  $N$  users, we adopt a multicast scheduling scheme where a subgroup of users is served simultaneously in each time slot with a rate that depends on the instantaneous channel condition of the users. We assume that the message size  $S$  is large (or that the slot duration  $T$  is small) so that the distribution of the common message to all  $N$  users must be completed over multiple time slots.

By employing a rateless encoding scheme, such as the fountain code [7], the BS can generate a continuous stream of source bits from the intended message so that any collection of  $S$  bits from the data stream can guarantee reliable recovery of the original message. In the proposed transmission scheme, a different portion of the data stream is transmitted in each time slot and a user is assumed to be able to recover the original message whenever it collects a total of  $S$  bits over a series of time slots, regardless of which portions of the bit stream were received.

It is assumed that the BS has knowledge of the instantaneous signal-to-noise ratio (SNR) of each user at the beginning of every time slot. Based on the instantaneous SNR, the BS will select a group of target users, whose index set is denoted by  $\mathcal{I}(k)$  for the  $k$ -th slot, and transmit at a rate constrained by the lowest-SNR user among all *selected* users. The SNR of the  $n$ -th user in the  $k$ -th time slot takes the following form:

$$\gamma_n(k) = \frac{P \cdot |h_n(k)|^2}{N_0 \cdot L_p(D_n)}, \quad (1)$$

where  $P$  is the transmission power,  $N_0$  is the noise variance, and  $L_p(D_n)$  is the path loss over the distance between the

BS and the  $n$ -th user (i.e.,  $D_n$ ). Thus, the maximal rate supportable by the  $n$ -th user in the  $k$ -th time slot is given by

$$r_n(k) = \log(1 + \gamma_n(k)). \quad (2)$$

Subsequently, the maximal data rate that the BS can transmit without incurring channel outage to the *selected* users in  $\mathcal{I}_k$  is given by

$$R(k) = \min_{n \in \mathcal{I}(k)} r_n(k). \quad (3)$$

Therefore, each user in  $\mathcal{I}(k)$  will successfully receive  $R(k) \cdot T$  source bits in the  $k$ -th slot.

Let  $K$  be the total number of time slots required for all  $N$  users to successfully recover the message (i.e., the *system delay*). Clearly, there exists a design tradeoff between the sequence of multicast groups  $\{\mathcal{I}(k), \forall k \geq 1\}$  and the system delay  $K$ . On the one hand, selecting more users will improve the multicast gain at the cost of a lower data rate  $R(k)$ . On the other hand, selecting fewer users enables higher data rates by better exploiting the multiuser diversity gain but limits the number of users that are served simultaneously in each time slot. In the following, optimized OMS is devised by carefully taking into account the tradeoff between multicast gain and multiuser diversity in both homogeneous and heterogeneous networks.

### III. OMS IN HOMOGENEOUS NETWORKS

We begin with homogeneous networks where all users experience i.i.d. fading with the same average SNR,  $\rho_0$ . Mathematically,  $\rho_0$  can be defined as shown in (1) with  $L_p(D_n) = 1$  for all  $n$ :

$$\rho_0 = \mathbf{E}[\gamma_n(k)] = \mathbf{E}\left[\frac{P|h_n(k)|^2}{N_0}\right]. \quad (4)$$

Without loss of generality, let  $F_{\gamma_0}(z)$  be the cumulative distribution function (CDF) of the instantaneous SNR of each user. From (2), the CDF of each user's supportable data rate can be expressed as

$$F_r(y) = \Pr\{r = \log(1 + \gamma) \leq y\} = F_{\gamma_0}(2^y - 1) \quad (5)$$

For example, under the Rayleigh fading scenario, the channel coefficient  $h_n(k)$  is assumed to be circularly symmetric complex Gaussian with zero mean and unit variance, i.e.  $h_n(k) \sim \mathcal{CN}(0, 1)$ , and the CDF of the instantaneous SNR and the supportable data rate are given by

$$F_{\gamma_0}(z) = 1 - \exp\left(-\frac{z}{\rho_0}\right), \quad (6)$$

and

$$F_r(y) = 1 - \exp\left(-\frac{2^y - 1}{\rho_0}\right), \quad (7)$$

respectively. It is worthwhile to remark that no specific fading distributions are assumed in this work and, thus, the results are generally applicable to arbitrary fading scenarios.

To facilitate the following discussion, we rank users according to their instantaneous SNR in the  $k$ -th time slot and define

an order-mapping function as  $\pi_k(i)$  for  $i = 1, 2, \dots, N$  with ordered instantaneous SNR given by

$$\gamma_{\pi_k(1)}(k) \geq \gamma_{\pi_k(2)}(k) \geq \dots \geq \gamma_{\pi_k(N)}(k), \quad (8)$$

or equivalently ordered data rates  $r_{\pi_k(1)}(k) \geq r_{\pi_k(2)}(k) \geq \dots \geq r_{\pi_k(N)}(k)$ . Note that the position of the  $n$ -th indexed user in the  $k$ -th time slot can be easily determined by inverting the order-mapping function, i.e.,  $\pi_k^{-1}(n)$ .

In the proposed OMS scheme, the BS transmits to a fixed number of users, denoted by  $M$ , in each time slot regardless of how many bits they have previously received. Specifically, in the  $k$ -th time slot, the BS chooses the multicast set  $\mathcal{I}(k) = \{\pi_k(1), \pi_k(2), \dots, \pi_k(M)\}$  and transmits at rate  $R(k) = r_{\pi_k(M)}$  so that only the users in the selected set can reliably decode the transmitted data. It is assumed that the BS keeps an updated record of the instantaneous SNRs of all users and that it is able to determine the amount of information each user has decoded in all previous time slots. The process completes when all users are able to accumulate  $S$  encoded bits. Hence, the system delay of the OMS scheme can be computed as a function of  $M$ , i.e.,

$$\mathcal{K}(M) = \min\left\{k : \min_n \left[ \sum_{j=1}^k \mathbf{1}_{\{\pi_j^{-1}(n) \leq M\}} r_{\pi_j(M)}(j) \right] \cdot T \geq S \right\}, \quad (9)$$

where the indicator function  $\mathbf{1}_{\{\pi_j^{-1}(n) \leq M\}}$  is defined as

$$\mathbf{1}_{\{\pi_j^{-1}(n) \leq M\}} = \begin{cases} 1 & \text{if } \pi_j^{-1}(n) \leq M \\ 0 & \text{otherwise,} \end{cases} \quad (10)$$

which indicates whether or not user  $n$  is included in the selected group of the  $j$ -th time slot.

It should be emphasized that (9) provides a generalized form of many existing scheduling schemes, such as those given in the following.

- **Conventional broadcast**

The broadcast scheme that schedules transmission to all users in each time slot is a special case of (9) with  $M = N$ , and the system delay is given by

$$K^{BC} \triangleq \mathcal{K}(N) = \min\left\{k : \sum_{j=1}^k r_{\pi_j(N)}(j) \cdot T \geq S\right\}. \quad (11)$$

- **Opportunistic unicast**

The opportunistic unicast scheduling method is a special case of (9) with  $M = 1$  and the corresponding system delay is

$$K^{UC} \triangleq \mathcal{K}(1) = \min\left\{k : \min_n \left\{ \sum_{j=1}^k \mathbf{1}_{\{\pi_j(1)=n\}} r_n(j) \right\} \cdot T \geq S \right\}. \quad (12)$$

- **Median OMS**

The median user OMS scheme proposed in [8], [9] is the case with  $M = N/2$ , assuming homogeneous users. The

resulting system delay is given by

$$K_{\text{med}}^{\text{OMS}} \triangleq \mathcal{K}\left(\frac{N}{2}\right) = \min\left\{k : \min_n \left\{ \sum_{j=1}^k \mathbf{1}_{\{\pi_j^{-1}(n) \leq \frac{N}{2}\}} r_{\pi_j(\frac{N}{2})}(j) \right\} \cdot T \geq S\right\}. \quad (13)$$

In the following, we analyze, for any given  $N$ , the optimal multicast group size  $M^*$  or, equivalently, the optimal user selection ratio,  $\alpha^* = \frac{M^*}{N}$  that minimizes the average system delay, i.e.,  $\mathbf{E}[\mathcal{K}(M)]$ . To obtain analytically tractable results, we investigate the asymptotic performance in terms of the average transmission rate and system delay for large networks (i.e., large  $N$ ) by resorting to extreme value theory [12].

Lemma 1: Let  $F_r(y)$  and  $f_r(y)$  be the CDF and the probability density function (PDF) of the data rate, which are assumed to be continuous. Given  $\alpha \in (0, 1)$  with  $f_r(F_r^{-1}(1 - \alpha)) \neq 0$ , it follows that, if  $\lim_{N \rightarrow \infty} \frac{M}{N} = \alpha$ , then

$$\sqrt{N} [r_{\pi_k(M)}(k) - F_r^{-1}(1 - \alpha)] \sim \mathcal{N}(0, \sigma^2(\alpha)), \quad (14)$$

with

$$\sigma^2(\alpha) = \frac{\alpha(1 - \alpha)}{[f_r(F_r^{-1}(1 - \alpha))]^2}. \quad (15)$$

The proof follows directly from the asymptotic normality of central order statistics stated in Theorem 7.5 in [12]. More specifically, it follows from (14) that

$$\begin{aligned} \text{Var} [r_{\pi_k(M)}(k) - F_r^{-1}(1 - \alpha)] &= \\ \mathbf{E} \left[ (r_{\pi_k(M)}(k) - F_r^{-1}(1 - \alpha) - 0)^2 \right] &= \frac{1}{N} \sigma^2(\alpha), \end{aligned} \quad (16)$$

which goes to zero as  $N$  approaches infinity. This implies that  $r_{\pi_k(M)}(k)$  converges to  $F_r^{-1}(1 - \alpha)$  in the mean square sense as  $N$  goes to infinity. That is,  $r_{\pi_k(M)}(k)$  can be approximated by  $F_r^{-1}(1 - \alpha)$  for sufficiently large  $N$ . For example, under the Rayleigh fading, the asymptotic transmission rate is given by

$$\lim_{N \rightarrow \infty} r_{\pi_k(M)}(k) = \log \left( 1 - \rho_0 \ln \frac{M}{N} \right),$$

where  $\alpha$  is the user selection ratio to be optimized. Consequently, the number of times that each user must be selected to successfully receive  $S$  source bits is approximated by

$$\kappa = \frac{S}{F_r^{-1}(1 - \alpha) \cdot T}. \quad (17)$$

Next, we analyze the asymptotic characteristic of the system delay. For notational convenience, we use  $X_{n,K}$  to denote the number of time slots that the  $n$ -th user is chosen over total  $K$  time slots. Mathematically, it can be expressed as

$$X_{n,K} = \sum_{k=1}^K \mathbf{1}_{\{\pi_k^{-1}(n) \leq M\}}. \quad (18)$$

By invoking the central limit theorem, we can approximate  $X_{n,K}$  as a Gaussian distributed random variable such that

$$X_{n,K} \sim \mathcal{N}(K\alpha, K\alpha(1 - \alpha)). \quad (19)$$

Clearly, the system delay is bottlenecked by the user that is chosen the least number of times over the  $K$  time slots and can be characterized by

$$X_{\min,K} = \min_n X_{n,K}. \quad (20)$$

To study the asymptotic characteristics of  $X_{\min,K}$ , we need the following result from extreme value theory [12].

Lemma 2: Let  $X_{\min} = \min_n X_n$  where  $X_n$ , for  $n = 1, \dots, N$ , is a set of i.i.d. random variables with CDF  $F_X(x)$ . If the following condition holds,

$$\begin{aligned} \lim_{N \rightarrow \infty} N \cdot F_X \left( F_X^{-1} \left( \frac{1}{N} \right) + x \left[ F_X^{-1} \left( \frac{1}{N \cdot e} \right) - F_X^{-1} \left( \frac{1}{N} \right) \right] \right) \\ = e^{-x}, \end{aligned} \quad (21)$$

there exists normalizing constants  $c_N$  and  $d_N$  such that  $(X_{\min} - c_N)/d_N$  converges in distribution to the Gumbel distribution for minima. The normalizing constants can be determined by

$$c_N = F_X^{-1} \left( \frac{1}{N} \right) \quad (22)$$

$$d_N = c_N - F_X^{-1} \left( \frac{1}{N \cdot e} \right). \quad (23)$$

In this case, we say that  $F_X(x)$  lies in the Domain of Attraction of the Gumbel distribution for minima.

Proof: The proof follows from Theorem 3.4 in [12].

As shown in [12], the Gaussian CDF lies in the Domain of Attraction of the Gumbel distribution for minima. Thus, by the approximation in (19) and by Lemma 2, we know that  $(X_{\min,K} - c_N)/d_N$  converges to the Gumbel distribution with normalizing constants given as

$$c_N = \sigma_x \Phi^{-1} \left( \frac{1}{N} \right) + \mu_x \quad (24)$$

$$d_N = \sigma_x \left[ \Phi^{-1} \left( \frac{1}{N} \right) - \Phi^{-1} \left( \frac{1}{N \cdot e} \right) \right], \quad (25)$$

where  $\Phi(\cdot)$  is the standard Gaussian CDF while  $\sigma_x$  and  $\mu_x$  are the standard deviation and mean of  $X_n$ , respectively. By exploiting the properties of the Gumbel distribution [12], [14], we have

$$\mathbf{E}[X_{\min,K}] \approx c_N + (-\varepsilon)d_N, \quad (26)$$

$$= \sigma_x \left[ (1 - \varepsilon) \Phi^{-1} \left( \frac{1}{N} \right) + \varepsilon \Phi^{-1} \left( \frac{1}{N \cdot e} \right) \right] + \mu_x, \quad (27)$$

and,

$$\text{Var}[X_{\min,K}] = \frac{\pi^2}{6} \cdot d_N^2 \quad (28)$$

where  $\varepsilon = 0.57721$  is the Euler-Mascheroni constant,  $\mu_x = K\alpha$ , and  $\sigma_x = \sqrt{K\alpha(1 - \alpha)}$ . It is straightforward to show from (28) that

$$\text{Var}[X_{\min,K}] = \mathcal{O}(1), \quad (29)$$

for sufficiently large  $N$ , which implies that the system delay variance is insensitive to the number of users in a homogeneous system. This observation will be confirmed by

simulation in Sec.V.

By setting  $\mathbf{E}[X_{\min,K}] = \kappa$ , where  $\kappa$  is defined as in (17), we can compute the number of time slots  $K$  required for all users to be selected at least  $\kappa$  times. This value can then be used to approximate the average system delay as

$$\mathbf{E}[\mathcal{K}(\alpha N)] \approx K = \left[ \frac{-\xi \sqrt{\alpha(1-\alpha)} + \sqrt{\xi^2 \alpha(1-\alpha) + 4\alpha\kappa}}{2\alpha} \right]^2, \quad (30)$$

where  $\xi = (1-\varepsilon) \Phi^{-1}\left(\frac{1}{N}\right) + \varepsilon \Phi^{-1}\left(\frac{1}{N \cdot \varepsilon}\right)$ . The optimal user selection ratio for homogeneous networks can then be obtained by solving the following optimization problem:

$$\alpha_{\text{hom}o}^* = \arg \min_{\alpha} \mathbf{E}[\mathcal{K}(\alpha N)], \quad (31)$$

where  $\mathcal{K}(\cdot)$  is given in (9). It is interesting to note that the optimization in (31) is only over a single parameter,  $\alpha \in (0, 1]$ , and can be solved efficiently using a simple line search. In practice, the user selection ratio  $\alpha_{\text{hom}o}^* = \frac{M}{N}$  can take only  $N$  possible values since the number of users selected in each time slot is an integer, *i.e.*  $M = 1, 2, \dots, N$ . Hence, the line search has a complexity that is linear with  $N$  and, thus, can be done efficiently in practice.

Finally, the operations of the proposed opportunistic scheduling scheme can be summarized as follows. When multicast service is requested by the users, the BS first computes the optimal user selection ratio  $\alpha_{\text{hom}o}^*$  by substituting into (31) the average SNR  $\rho_0$ , the total number of users  $N$ , and the time slot duration  $T$ . The BS transmits the message to all users over multiple time slots during which the rates are chosen adaptively according to the selected group of users in each time slot. Specifically, at the beginning of each time slot, the BS ranks the instantaneous SNR of all users and chooses a data rate that is supported only by the best  $\lfloor \alpha_{\text{hom}o}^* N \rfloor$  users. The instantaneous SNR can be derived by either exploiting channel reciprocity in time-division duplex (TDD) systems or receiving feedback from each user in frequency-division duplex (FDD) systems. By assuming that the source data are encoded by erasure or fountain codes, each user will be able to decode the entire message whenever it receives a total of  $S$  bits over the sequence of time slots, regardless of which specific time slots it was able to receive from. This process continues until all users are able to successfully decode the entire message.

#### IV. OMS IN HETEROGENEOUS NETWORKS

In this section, we extend results of the last section to heterogeneous networks comprising of  $N$  active users uniformly distributed around the BS in a circular cell of radius  $D_{\text{max}}$  as illustrated in Fig. 2. The PDF of the distance  $D_n$  between the  $n$ -th user and the BS is given by [15]

$$f_{D_n}(d) = \frac{2d}{D_{\text{max}}^2}, \quad 0 < d \leq D_{\text{max}}. \quad (32)$$

We model the effect of path loss between the BS and the  $n$ -th user as  $L_p(D_n) = \epsilon D_n^\beta$ , where  $\epsilon$  and  $\beta$  are the path loss

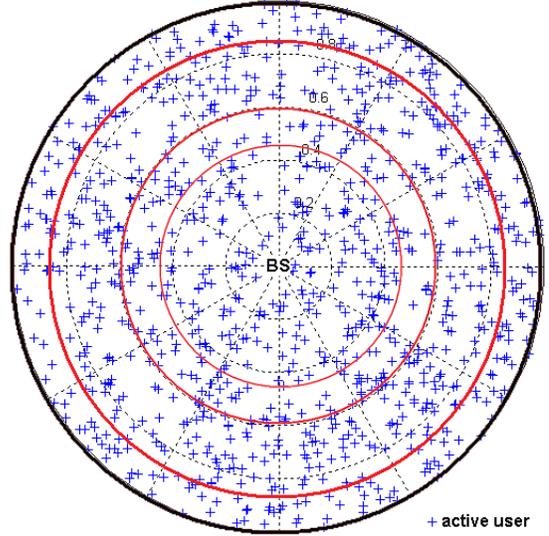


Fig. 2. Homogeneous rings of users in the heterogeneous cell.

constant and exponent, respectively. By taking into account the path loss, the average SNR experienced by the  $n$ -th user can be expressed as

$$\rho(D_n) = \mathbf{E}[\gamma_n(k)] = \frac{\rho_0}{\epsilon D_n^\beta}. \quad (33)$$

Subsequently, for any given  $D_n$ , the CDF of the instantaneous SNR of the  $n$ -th user during the  $k$ -th time slot, denoted by  $\gamma_n(k)$ , can be expressed as

$$F_{\gamma_n|D_n}(z) = \Pr \left\{ \gamma_n(k) = \frac{P \cdot |h_n(k)|^2}{N_0 \cdot \epsilon D_n^\beta} \leq z \mid D_n \right\} = F_{\gamma_0}(z \cdot \epsilon D_n^\beta), \quad (34)$$

where  $F_{\gamma_0}(z)$  was defined in the previous section as the CDF of  $\gamma_n(k)$  with  $L_p(D_n) = 1$ , and the corresponding CDF of the data rate as

$$F_{r_n|D_n}(y) = F_{\gamma_0}((2^y - 1) \cdot \epsilon D_n^\beta). \quad (35)$$

In the following, we investigate two different schemes in applying OMS: the absolute-SNR scheme and the normalized-SNR scheme. In the absolute-SNR scheme, the actual instantaneous SNR is employed to compute the user selection ratio whereas the normalized-SNR scheme first normalizes each user's instantaneous SNR with respect to its average SNR before applying OMS. Similar to the proportional fair scheduling (PFS) scheme in [2] that employs the normalized SNR to ensure fairness among different users, the normalized-SNR scheme is devised to increase the chance of selecting the weaker users whose performance hinders the system delay.

##### A. OMS with Absolute SNR

It is not surprising to observe that the average system delay in heterogeneous networks is typically limited by the users farthest away from the BS, *i.e.*, cell-edge users. For sufficiently dense networks, we can assume that a ring of users at the

cell edge, which is referred to as the *edge group*, form a homogeneous group with average SNR approximately equal to

$$\rho_{\text{edge}} = \frac{\rho_0}{\epsilon D_{\text{max}}^\beta}. \quad (36)$$

Concentrating on this cell-edge group, we use  $\alpha_{\text{edge}}^*$  to denote the optimal selection ratio for the edge group according to (31). Since the delay performance of the heterogeneous network is dominated by this edge group, we argue that users selected by the heterogeneous OMS scheme should include exactly  $\alpha_{\text{edge}}^*$  proportion of users in this group. Based on this observation, we derive in the following optimal user selection ratio for heterogeneous networks,  $\alpha_{\text{abs}}^*$ .

From Lemma 1, we know that, in a homogeneous group of  $N_{\text{edge}}$  users, the rate of the  $U$ -th user, where  $U = \lfloor \alpha_{\text{edge}}^* N_{\text{edge}} \rfloor$ , will converge to a constant

$$r_{\pi_k(U)}(k) \approx F_{r_n|D_{\text{max}}}^{-1}(1 - \alpha_{\text{edge}}^*) \triangleq r_{\text{th}} \quad (37)$$

for large  $N_{\text{edge}}$ . In other words, as long as the OMS scheme chooses a multicast group that results in the transmission rate  $r_{\text{th}}$ , the optimal portion of users in the edge group, *i.e.*  $\alpha_{\text{edge}}^*$ , will be chosen. Suppose that  $M = \lfloor \alpha_{\text{abs}}^* N \rfloor$  users are chosen in each time slot in the heterogeneous OMS scheme. Interestingly, by assuming that the location of users are i.i.d., we can also show by Lemma 1 that the rate transmitted in each time slot converges to constant  $F_{r_n}^{-1}(1 - \alpha_{\text{abs}}^*)$ , where the CDF of the rate  $r_n$  is now averaged over the random location of the user, *i.e.*,

$$F_{r_n}(y) = \int F_{r_n|D_n=u}(y) \cdot f_{D_n}(u) du \quad (38)$$

$$= \int_0^{D_{\text{max}}} F_{r_n|D_n=u}(y) \cdot \frac{2u}{D_{\text{max}}^2} du. \quad (39)$$

The user selection ratio  $\alpha_{\text{abs}}^*$  must be chosen to satisfy  $F_{r_n}^{-1}(1 - \alpha_{\text{abs}}^*) = r_{\text{th}}$ . Thus, we have

$$\alpha_{\text{abs}}^* = 1 - F_{r_n}(r_{\text{th}}). \quad (40)$$

For example, in the case of Rayleigh fading, this is given by

$$\alpha_{\text{abs}}^* = \frac{2}{D_{\text{max}}^2} \int_0^{D_{\text{max}}} u \exp\left(\frac{\ln \alpha_{\text{edge}}^*}{D_{\text{max}}^\beta} u^\beta\right) du, \quad (41)$$

$$= \frac{1}{D_{\text{max}}^2} \int_0^{D_{\text{max}}} \exp\left(\frac{\ln \alpha_{\text{edge}}^*}{D_{\text{max}}^\beta} z^{\beta/2}\right) dz, \quad (42)$$

$$\stackrel{(a)}{=} \frac{1}{D_{\text{max}}^2} \int_0^{D_{\text{max}}} \left(1 + \sum_{m=1}^{\infty} \left[\left(\frac{\ln(\alpha_{\text{edge}}^*)}{D_{\text{max}}^\beta}\right)^m \frac{z^{\frac{m\beta}{2}}}{m!}\right]\right) dz, \quad (43)$$

$$= 1 + \sum_{m=1}^{\infty} \frac{(\ln \alpha_{\text{edge}}^*)^m}{m! \left(\frac{m\beta}{2} + 1\right)}, \quad (44)$$

where (a) follows from the Taylor series expansion of the exponential function about  $z = 0$ . The system delay of the absolute SNR scheme is

$$K_{\text{abs}}^{\text{OMS}} = \mathcal{K}(\alpha_{\text{abs}}^* N), \quad (45)$$

where  $\mathcal{K}$  is given in (9).

## B. OMS with Normalized SNR

In the previous subsection, the proposed OMS scheme selects the best  $M = \lfloor \alpha_{\text{abs}}^* N \rfloor$  users according to their absolute instantaneous SNR values. This method is simple, and it requires no knowledge of the average SNR of each user. However, it suffers from the disadvantage that users far away from the BS are often deprived of their opportunity to transmit even if they have reached their relative maximal SNR. To address this issue, we propose an alternative method where the OMS is derived based on the normalized instantaneous SNR values with respect to the average SNR. Let us define the normalized instantaneous SNR of the  $n$ -th user as

$$\hat{\gamma}_n(k) = \frac{\gamma_n(k)}{\rho_n}, \quad (46)$$

where  $\rho_n = \rho(D_n)$  is the average SNR of the  $n$ -th user.

1) *Two-User Example:* We first use a two-user network to provide some insights. Consider a network consisting of two users: User 2 is closer to the BS than User 1, *i.e.*  $\rho_2 > \rho_1$ . For such a simplified network, the unicast scheme selects either User 1 or User 2, while the broadcast scheme selects both users at the same time. In contrast, the OMS scheme selects either one or both users depending on the achievable delay performance. Interestingly, it is shown in the Appendix that when the OMS scheme selects one user, the delay experienced by User 1 under the normalized-SNR scheme is always smaller than that under the absolute-SNR scheme, *i.e.*,  $K_1^{\text{abs}} \geq K_1^{\text{norm}}$ . Recall that delay is limited by User 1 and that, when the OMS scheme selects two users, the delay is the same for both absolute-SNR and normalized-SNR versions. As a result, the normalized-SNR OMS scheme will perform better or equally well as compared to the absolute-SNR OMS scheme for the two-user case. Simulation results will be provided in Sec. V-B1 to confirm the analysis reported in the Appendix.

2) *Multiple-User Case:* Thus inspired, we establish the normalized-SNR scheme for general networks with  $N$  active users. Following the preceding analysis, let us define the ordered-mapping function  $\phi_k(i)$  for  $i = 1, 2, \dots, N$  according to the normalized SNR such that

$$\hat{\gamma}_{\phi_k(1)}(k) \geq \hat{\gamma}_{\phi_k(2)}(k) \geq \dots \geq \hat{\gamma}_{\phi_k(N)}(k). \quad (47)$$

The position of the  $n$ -th indexed user in the order is given by the inverse function  $\phi_k^{-1}(n)$ . In the normalized-SNR OMS scheme, the scheduler selects a group of  $M = \lfloor \alpha N \rfloor$  users indexed by  $\hat{\mathcal{I}}(k) = \{n : \phi_k^{-1}(n) \leq M\}$  for transmission, where  $\alpha$  is the user selection ratio. Since the normalized SNR's are i.i.d., all users are selected with equal probability  $\alpha$ . Furthermore, because delay is limited by the edge group users, the OMS scheme should set the *normalized* selection ratio equal to the optimal  $\alpha_{\text{edge}}^*$  according to (31), *i.e.*,  $\alpha_{\text{norm}}^* = \alpha_{\text{edge}}^*$ .

To ensure that all users in  $\hat{\mathcal{I}}(k)$  receive the transmission, the BS needs to transmit at a rate constrained by the worst user in  $\hat{\mathcal{I}}(k)$ ; namely,

$$R_{\text{norm}}(k) = \min_{n \in \hat{\mathcal{I}}(k)} r_n(k) = \min_{n \in \hat{\mathcal{I}}(k)} \log(1 + \gamma_n(k)). \quad (48)$$

It is important to note that this rate is typically smaller than the  $M$ -th best user according to the absolute instantaneous

SNR. As a result, a total of  $\hat{M}(k)$  users are able to receive the transmission in the  $k$ -th time slot with

$$\hat{M}(k) = \max_{n=1, \dots, M} \pi_k^{-1}(\phi_k(n)). \quad (49)$$

Thus, the system delay for the normalized SNR scheme is given by

$$K_{norm}^{OM} = \min \left\{ k : \min_n \left[ \sum_{j=1}^k \mathbf{1}_{\{\pi_j^{-1}(n) \leq \hat{M}(j)\}} R_{norm}(j) \right] \cdot T \geq S \right\}. \quad (50)$$

For networks with a large number of users, we expect that the number of users actually selected in the normalized-SNR scheme will also converge to a constant and the difference between the two schemes will be negligible.

## V. SIMULATION RESULTS

In this section, we study the performance of the proposed OMS schemes via computer simulation. Unless otherwise specified, we set  $N = 100$  so that a selection ratio of  $\alpha = M\%$  effectively selects  $M$  users in the following experiments. The Rayleigh fading model is assumed throughout the experiments.

### A. Homogeneous networks

We first verify the accuracy of the optimal selection ratios of a homogeneous network predicted by (31). Fig. 3 shows the simulated optimal selection ratios (i.e., the user selection ratios obtained by minimizing the simulated average delay) and the analytically optimized ratios using (31) with respect to the average SNR. We can see that the median OMS that fixes the selection ratio at 50% may not be optimal for all SNR values. It is interesting to note that, in practice as well as in the simulations, the number of time slots needed for each user to receive the entire message must be an integer value, causing a quantization effect that is not well predicted by our asymptotic analysis. This quantization effect is less critical at low SNR where the number of time slots needed for each user to receive the entire message is large, but it is more evident at moderate to high SNR where the entire message may be received within only a few time slots. Interestingly, this effect may also yield more than one locally optimal values of  $\alpha_{homo}^*$ . In fact, we observed in our experiments two locally optimal solutions of  $\alpha_{homo}^*$  for certain SNR values. The two locally optimal user selection ratios yield small difference in delay and can be used equally effective in the proposed OMS scheme. Both of these solutions are plotted in Fig. 3. The analytical approach presented in (31) is able to predict one of these solutions with good accuracy.

Next, we compare the delay performance of the proposed OMS scheme against three other existing scheduling schemes mentioned in Section III (i.e., the conventional broadcast, unicast, and the median OMS schemes) in Fig. 4. Inspection of Fig. 4 confirms the conventional wisdom: the unicast scheme outperforms the broadcast scheme at the low SNR region by exploiting the multiuser diversity gain while the broadcast scheme has more advantages when the multicast gain becomes

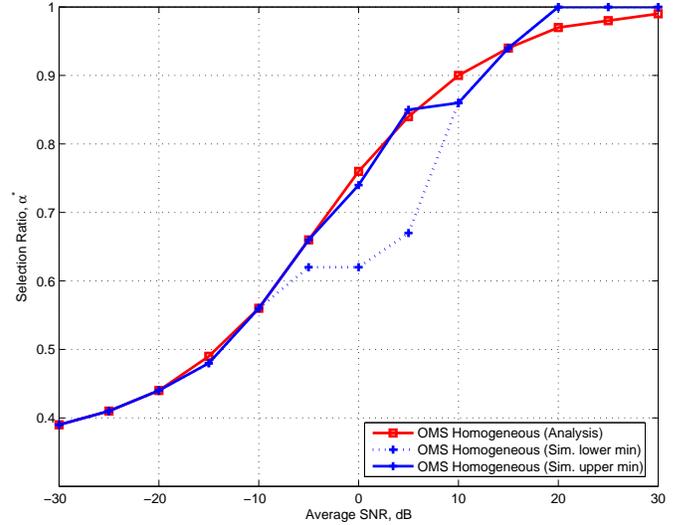


Fig. 3. Comparison of theoretical and simulated optimal selection ratios in homogeneous networks.

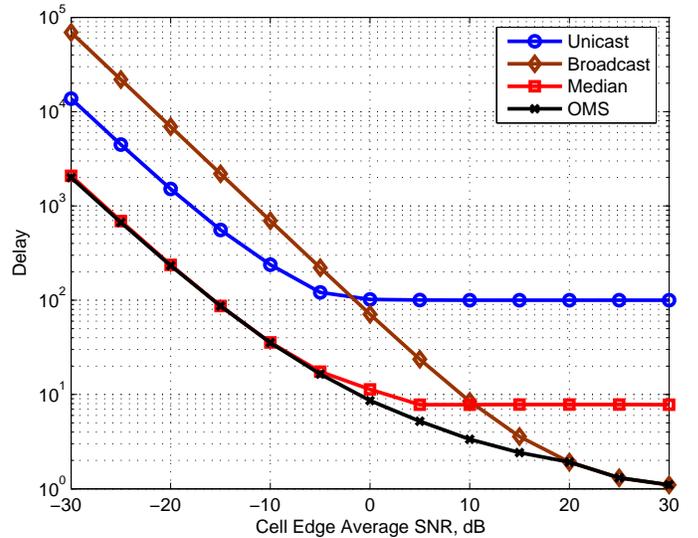


Fig. 4. Performance comparison of four multicast schemes in homogeneous networks.

more dominant with the increase in SNR. The median OMS scheme outperforms both unicast and broadcast schemes for SNR below 20 dB. For very large SNR values, the OMS scheme chooses to serve all users in every time slot, which results in the delay performance converging to that of the broadcast scheme.

Finally, in Fig. 5, we plot the average OMS delay and its variance with respect to the number of users at SNR of 10 dB. It is easy to show from (30) that the average delay is  $\mathcal{O}(\xi^2) = \mathcal{O}(\ln N)$ , which is confirmed in Fig. 5. Furthermore, inspection of Fig. 5 also reveals that the delay variance is insensitive to the number of users, which agrees

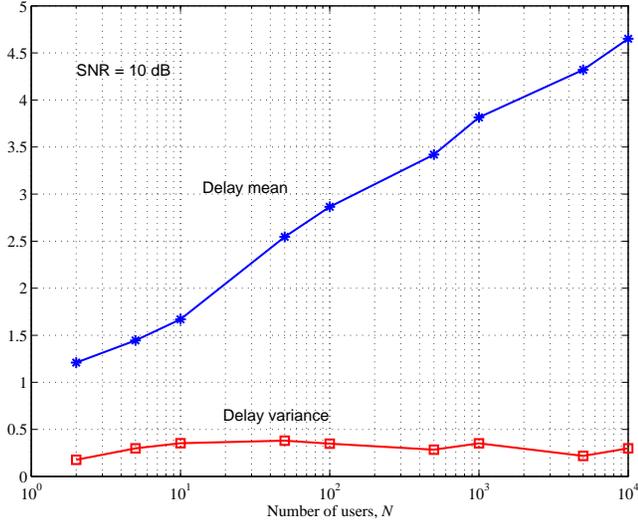


Fig. 5. Delay performance as a function of the number of users with SNR of 10dB.

well with the discussion in Sec. III.

### B. Heterogeneous networks

Here, we simulate the heterogeneous scenario and provide further insight into the performance of the proposed OMS scheme when users' channel statistics are not i.i.d..

1) *Two-User Example*: We first perform simulation to confirm the analysis on two-user networks reported in the Appendix. In Fig. 6, we simulate the delay performance of the two-user example where User 1 is at a fixed distance from the BS and has an average SNR of  $-20$  dB. The average system delay is depicted for a range of SNR separation between the two users. Fig. 6 indicates that the performance gain by normalizing the SNR increases as the separation increases and attains its maximum at an SNR separation of about 2 dB. For an SNR separation larger than 4.5 dB, the broadcast scheme is optimal among all schemes under consideration. Similar results can also be observed in Fig. 7, which is plotted by increasing the edge user's average SNR to  $-10$  dB (e.g. a smaller cell). The system delay is reduced for all schemes due to the higher average SNR values. However, the improvement achieved by the normalized-SNR scheme remains non-negligible for an SNR separation under 2 dB.

2) *Networks with more than two users*: Next, we simulate general heterogeneous networks with more than two users. The users are distributed uniformly in a circular cell area of radius  $D_{max} = \left(\frac{\rho_o}{\epsilon \rho_{edge}}\right)^{\frac{1}{\beta}}$ , with path-loss constant  $\epsilon = 10^{3.15}$  and path-loss exponent  $\beta = 3.5$  [16]. We first examine the impact on the whole system delay due to the worst users in the network. In Fig. 8, we fix the cell area and the average cell-edge SNR at  $-20$  dB and  $-10$  dB. Then, we increase the number of users in the cell from 10 to 1000 and uniformly distribute them over the cell. Note that the system delay performance of the network is limited by the two outermost

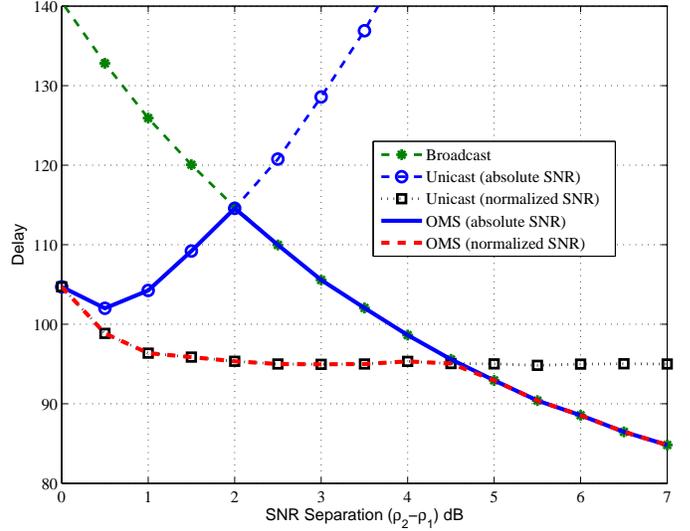


Fig. 6. Delay performance as a function of SNR separation for the two-user example with  $\rho_1 = -20$  dB.

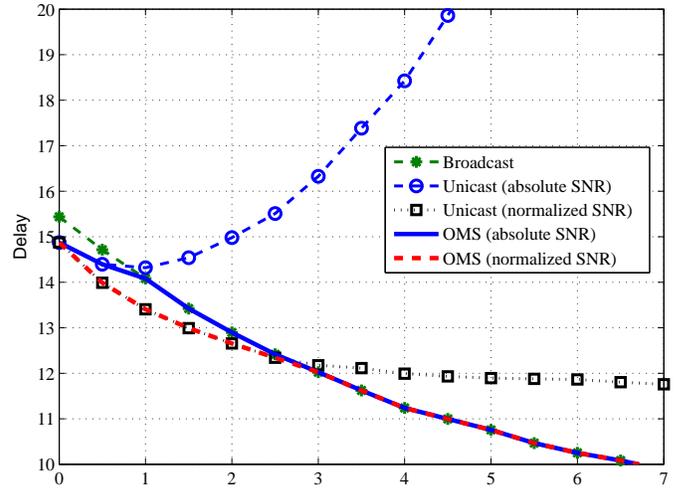


Fig. 7. Delay performance as a function of SNR separation for the two-user example with  $\rho_1 = -10$  dB.

users having the worst average SNR. Hence, we investigate the delay performance as the average SNR separation between the two outermost users changes. Clearly, as the number of users increases, the user density in the cell increases and the average SNR separation between the two outermost users decreases. Inspection of Fig. 8 shows that, as the average SNR separation becomes large (i.e. the network becomes less dense), the advantage of the normalized-SNR scheme is more apparent. Furthermore, Fig. 8 suggests that normalizing the SNR in OMS results in less improvement for the same SNR separation for a smaller cell (i.e. the average cell-edge SNR of  $-10$  dB).

Next, we use Fig. 9 to confirm the optimal selection ratio for the OMS scheme using the absolute-SNR and the normalized-SNR selection schemes as described in Sec. IV. For the

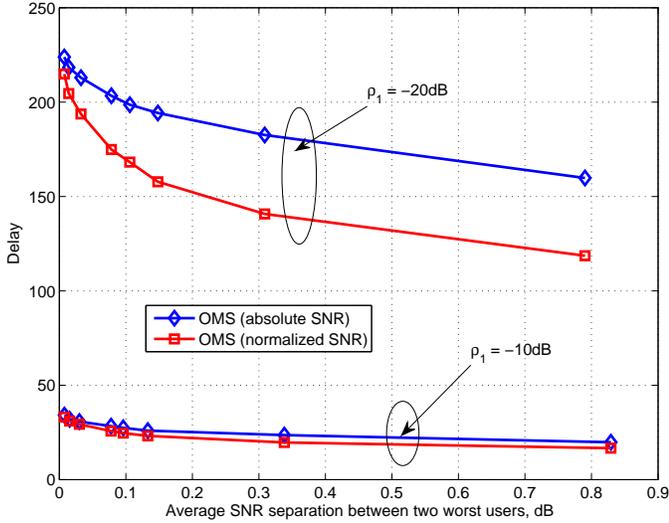


Fig. 8. Comparison of OMS with the normalized-SNR scheme and the absolute-SNR scheme in heterogeneous networks.

absolute-SNR selection, we compare the optimized selection ratio analytically derived from (44) with that obtained by exhaustive search over the simulated average delay. As with the homogeneous case, we also observe, for certain SNR values, two locally optimal user selection ratios that yield similar delay performances. Inspection of Fig. 9 reveals that the analytical result closely matches the optimal selection ratio obtained by simulations when the minimum is unique and it matches one of the solutions when more than one exists. Furthermore, we examine the normalized-SNR selection by comparing the analytical and simulated ratios. Fig. 9 suggests that the analytical and simulated results agree well with each other in a similar manner. Comparison of selection ratios obtained from the absolute-SNR and the normalized-SNR selection schemes shows that the normalized-SNR scheme tends to select a smaller selection ratio. However, it should be borne in mind that the normalized-SNR scheme chooses the user selection ratio based on the normalized SNRs while the actual ratio of users whose supportable data rates are higher than the selected rate is larger than the selection ratio. Finally, the advantage of the normalized-SNR scheme is exemplified in Fig. 10, where we compare the delay performance of the proposed OMS schemes against other existing scheduling schemes in the heterogeneous scenario. For the unicast and median OMS schemes, *normalized* SNRs are employed as the user selection criteria to guarantee each user a fair chance to be served regardless of their average SNRs. Fig. 10 shows that the unicast scheme suffers from the worst delay performance in the SNR range under consideration. Furthermore, the broadcast scheme is more advantageous when the cell-edge average SNR is sufficiently large whereas the median scheme cannot fully exploit the multicast gain at large cell-edge average SNRs. In contrast, the proposed OMS schemes with the normalized-SNR scheme achieve robust performance throughout the SNR range examined.

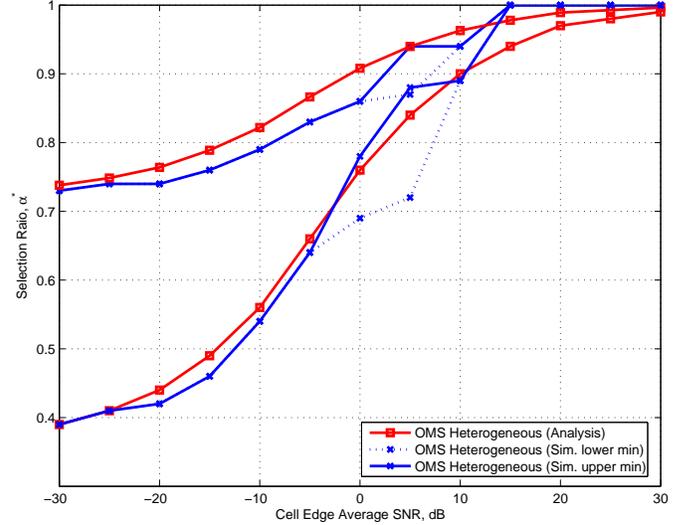


Fig. 9. The optimal selection ratio as a function of the cell-edge average SNR in heterogeneous networks.

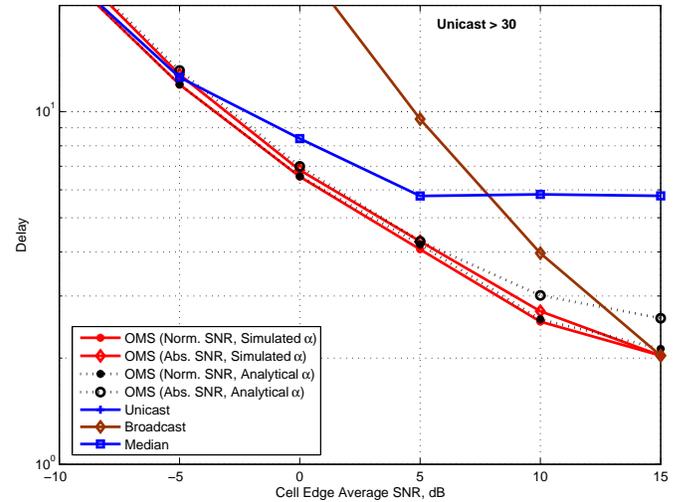


Fig. 10. Performance comparison of different scheduling schemes in heterogeneous networks.

## VI. CONCLUSION

In this work, we studied the opportunistic multicast scheduling (OMS) scheme with optimal user selection for both homogeneous and heterogeneous cellular networks. Capitalizing on extreme value theory, we provided theoretical analysis on the optimal user selection ratio that minimizes the system delay. Moreover, we demonstrated that the proposed OMS scheme with the normalized-SNR scheme can obtain further performance gain in heterogeneous networks via analysis on a two-user network example. Computer simulation confirmed that the proposed optimized OMS scheme achieves robust delay performance in both homogeneous and heterogeneous

networks by exploiting the optimal tradeoff between multiuser diversity and multicast gain.

Throughout this work, it has been assumed that the instantaneous SNR of each user is available to the BS. This assumption has significantly simplified the theoretical analysis while providing insights into the performance upper bound achieved by the proposed scheme. However, such an assumption may be restrictive in practice, particularly for networks with a large number of users. Some modifications of the proposed scheme are necessary for its practical deployment. For instance, limited-feedback techniques similar to that proposed in [17] merit future investigation.

#### APPENDIX A

##### TWO-USER OMS WITH NORMALIZED SNR

Suppose that the network consists of two users. One is farther from the BS and labeled as User 1, and the other is closer to the BS and labeled as User 2. Let  $\gamma_n(k)$  be the instantaneous SNR of the  $n$ -th user in the  $k$ -th time slot. Denote by  $\rho_n$  the average SNR of the  $n$ -th user such that  $\rho_1 < \rho_2$ . With a slight abuse of notation, let us define  $K_n$  as the delay required by the  $n$ -th user to successfully receive  $S$  encoded bits. The average system delay under a particular scheduling scheme can be expressed as

$$\mathbf{E}[K] = \mathbf{E}[\max\{K_1, K_2\}] \geq \max\{\mathbf{E}[K_1], \mathbf{E}[K_2]\} = \mathbf{E}[K_1]. \quad (51)$$

Notice that the expectation can be closely approximated by its lower bound when the difference of the average SNR of the two users are sufficiently large. In this case, the system delay is dominated by User 1, *i.e.*, the user with the worst average SNR.

To compare the performance between the absolute-SNR and the normalized-SNR schemes, we consider the case with  $M = 1$  (*i.e.* only one user is selected in each time slot) since the case with  $M = 2$  is identical for both schemes in the two-user example. In this case, User 1 is selected as the target user (and, thus, able to successfully receive the packet) under the absolute-SNR scheme if and only if  $\gamma_1 \geq \gamma_2$ . However, it also holds that

$$\begin{aligned} \{\gamma_1 \geq \gamma_2\} &= \left\{ \gamma_1 \geq \gamma_2, \frac{\gamma_1}{\rho_1} \geq \frac{\gamma_2}{\rho_2} \right\} \cup \left\{ \gamma_1 \geq \gamma_2, \frac{\gamma_1}{\rho_1} < \frac{\gamma_2}{\rho_2} \right\} \\ &\stackrel{(a)}{=} \left\{ \gamma_1 \geq \gamma_2, \frac{\gamma_1}{\rho_1} \geq \frac{\gamma_2}{\rho_2} \right\} \subset \left\{ \frac{\gamma_1}{\rho_1} \geq \frac{\gamma_2}{\rho_2} \right\}, \end{aligned}$$

where (a) follows from the fact that  $\gamma_1 < \frac{\gamma_2 \rho_1}{\rho_2} < \gamma_2$  since  $\rho_1 < \rho_2$ .

This shows that whenever User 1 is able to successfully receive a packet under the absolute-SNR scheme for a particular channel realization, it will also be able to successfully receive under the normalized-SNR scheme. Hence, the delay of User 1 can only be smaller under the normalized-SNR scheme, *i.e.*  $K_1^{\text{abs}} \geq K_1^{\text{norm}}$  and, hence, the normalized-SNR scheme will not perform worse than the absolute-SNR scheme for the two-user case.

#### REFERENCES

- [1] R. Knopp and P. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. IEEE International Conference on Communications (ICC)*, Seattle, WA, June 2005.
- [2] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Transactions on Information Theory*, vol. 48, pp. 1277–1294, June 2002.
- [3] M. Sharif and B. Hassibi, "Delay considerations for opportunistic scheduling in broadcast fading channels," *IEEE Transactions on Wireless Communications*, vol. 6, pp. 3353–3363, September 2007.
- [4] 3GPP, "Multimedia Broadcast/Multicast Service (MBMS) user service guidelines," 3rd Generation Partnership Project (3GPP), TR 26.946, Jun. 2007. [Online]. Available: <http://www.3gpp.org/ftp/Specs/htmlinfo/26946.htm>
- [5] M. Luby, T. Gasiba, T. Stockhammer, and M. Watson, "Reliable multimedia download delivery in cellular broadcast networks," *IEEE Transactions on Broadcasting*, vol. 53, pp. 235–246, March 2007.
- [6] M. Luby, M. Watson, T. Gasiba, T. Stockhammer, and W. Xu, "Raptor codes for reliable download delivery in wireless broadcast systems," in *Proc. the 3rd IEEE Consumer Communications and Networking Conference*, Las Vegas, NV, January 2006, pp. 192–197.
- [7] D. J. C. MacKay, "Fountain codes," *IEE Proceedings-Communications*, vol. 152, pp. 1062–1068, December 2005.
- [8] P. K. Gopala and H. E. Gamal, "Opportunistic multicasting," in *Proc. IEEE Thirty-Eighth Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, November 2004, pp. 845–849.
- [9] —, "On the throughput-delay tradeoff in cellular multicast," in *Proc. International Conference on Wireless Networks, Communications and Mobile Computing*, Maui, HI, June 2005, pp. 1401–1406.
- [10] W. Ge, J. Zhang, and S. Shen, "A cross-layer design approach to multicast in wireless networks," *IEEE Transactions on Wireless Communications*, vol. 6, pp. 1063–1071, March 2007.
- [11] U. C. Kozat, "On the throughput capacity of opportunistic multicasting with erasure codes," in *Proc. the 27th IEEE International Conference on Computer Communications (INFOCOM08)*, Phoenix, AZ, May 2008, pp. 520–528.
- [12] E. Castillo, *Extreme Value Theory in Engineering*. Boston: Academic Press, 1988.
- [13] S. Y. Baek, Y.-J. Hwang, and D. K. Sung, "Adaptive transmission scheme for mixed multicast and unicast traffic in cellular systems," *IEEE Transactions on Vehicular Tech.*, vol. 58, pp. 2899–2908, July 2009.
- [14] R. Reiss and M. Thomas, *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*, 2nd ed. Boston, MA : Birkhauser Verlag, 2001.
- [15] S. Y. Baek, H. Y. Hwang, and D. K. Sung, "Performance analysis of scheduling-based systems in Rayleigh fading channels," in *Proc. IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications*, Helsinki, Finland, September 2006.
- [16] IEEE, "802.20 Channel Models Document for IEEE 802.20 MBWA System Simulations," IEEE 802, IEEE 802.20-PD-08, Sep. 2005. [Online]. Available: <http://www.ieee802.org/20/PDdocs/IEEE802.20-PD-08.doc>
- [17] N. Ravindran and N. Jindal, "Multi-user diversity vs. accurate channel state information in MIMO downlink channels," *Submitted to IEEE Trans. Communications*, 2009.



**Tze-Ping Low** (S'08) received his B.Eng. and M.S. degrees in Electrical Engineering from National University of Singapore in 1999 and 2002, respectively. From 2000 to 2008, he worked as a radio network engineer in the telecommunications industry, designing and implementing cdma2000/EVDO and WCDMA/HSDPA radio networks. He is currently pursuing his Ph.D. at the University of Southern California, Los Angeles. His current research interests are in signal processing for communications and wireless multicast scheduling.



**Man-On Pun** (M'06) received the BEng. (Hon.) degree in electronic engineering from the Chinese University of Hong Kong in 1996, the MEng. degree in computer sciences from University of Tsukuba, Japan in 1999 and the Ph.D. degree in electrical engineering from University of Southern California, Los Angeles, in 2006, respectively. He is a research scientist at Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA. He held research positions at Princeton University, Princeton, NJ from 2006 to 2008 and Sony Corporation in Tokyo, Japan,

from 1999 to 2001.

Dr. Pun received the best paper award - runner-up from the IEEE Conference on Computer Communications (Infocom), Rio de Janeiro, Brazil in 2009, the best paper awards from the IEEE International Conference on Communications, Beijing, China in 2008 and the IEEE Vehicular Technology Fall Conference, Montreal, Canada in 2006. He is a recipient of several scholarships including the Japanese Government (Monbusho) Scholarship, the Sir Edward Youde Memorial fellowship for Overseas Studies and the Croucher postdoctoral fellowship.



**Y.-W. Peter Hong** (M'05) received his B.S. degree in Electrical Engineering from National Taiwan University, Taipei, Taiwan, in 1999, and his Ph.D. degree in Electrical Engineering from Cornell University, Ithaca, NY, in 2005. He joined the Institute of Communications Engineering and the Department of Electrical Engineering at National Tsing Hua University, Hsinchu, Taiwan, in Fall 2005, where he is now an Associate Professor. He was also a visiting scholar at the University of Southern California during June-August 2008. His research

interests include cooperative communications, distributed signal processing for sensor networks, and PHY-MAC cross-layer designs for next generation wireless networks.

Dr. Hong received the best paper award for young authors from the IEEE IT/COM Society Taipei/Tainan chapter in 2005, the best paper award among unclassified papers in MILCOM 2005, and also the Junior Faculty Research Award from the College of EECS at National Tsing Hua University in 2009. He is a co-editor (along with A. Swami, Q. Zhao, and L. Tong) of the book entitled "Wireless Sensor Networks: Signal Processing and Communications Perspectives" published by John-Wiley & Sons in 2007.



**C.-C. Jay Kuo** (S'83, M'86, SM'92, F'99) received the B.S. degree from the National Taiwan University, Taipei, in 1980 and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology, Cambridge, in 1985 and 1987, respectively, all in Electrical Engineering. Dr. Kuo was Computational and Applied Mathematics (CAM) Research Assistant Professor in the Department of Mathematics at the University of California, Los Angeles, from October 1987 to December 1988. Since January 1989, he has been with the University of Southern

California, where he is currently Professor of Electrical Engineering and Computer Science and Director of the Signal and Image Processing Institute. His research interests are in the areas of digital signal and image processing, multimedia compression, communication and networking technologies. He is a co-author of about 160 journal papers, 790 conference papers and 9 books.

Dr. Kuo is a Fellow of IEEE and SPIE. He is co-Editor-in-Chief for the Journal of Visual Communication and Image Representation, and Editor for the Journal of Information Science and Engineering, LNCS Transactions on Data Hiding and Multimedia Security and the EURASIP Journal of Applied Signal Processing. Dr. Kuo received the National Science Foundation Young Investigator Award (NYI) and Presidential Faculty Fellow (PFF) Award in 1992 and 1993, respectively.