A New Cross-Layer User Scheduler for Wireless Multimedia Relay Networks

Malcolm Egan, *Student Member, IEEE,* Phee Lep Yeoh, *Member, IEEE,* Maged Elkashlan, *Member, IEEE,* and Iain B. Collings, *Senior Member, IEEE*

***Abstract*—We propose a new scheduler for wireless multimedia relay networks (WMRNs). Our scheduler is designed to account for delay, symbol error probability (SEP), and packet loss probability (PLP) due to buffer overflow. We develop a cross-layer scheduling approach for the downlink to balance these system metrics. Our scheduler is based on a new metric which is referred to as the delay in packet scheduling (DPS). The user with the largest weighted signal-to-noise ratio is scheduled, where the weight is a function of the DPS. We then derive analytical expressions for the probability mass function (PMF) of the DPS, and the SEP of the scheduled user in Rayleigh fading. We derive an analytical approximation for the PMF of the buffer state. An analytical expression is then derived for the PLP due to buffer overflow. Our analysis is verified via simulations. We show the probability that a target DPS is met is 30% higher for our new scheme compared to the standard opportunistic equal weight scheduler, with negligible degradation in the SEP of the scheduled user. This can lead to a 85% improvement in the PLP.**

***Index Terms*—Scheduling, relay, cross-layer.**

I. Introduction

THE explosive demand for real-time audio and video streaming drives the need for new high data-rate transmission strategies in wireless multimedia networks [1, 2]. The defining property of these networks is low delay tolerance, in addition to throughput constraints [1]. Practical implementations of wireless multimedia networks are currently under consideration in current and emerging standards such as the IEEE 802.16 [3–5]. Wireless multimedia relay networks (WMRNs) are one important class of such networks [6–8]. In this paper, we address the important problem of scheduling in WMRNs with multiple data classes and multiple users. Not only does it call for a cross-layer approach to address the multimedia data, but it must also have a cross-user aspect that takes into account the instantaneous channel states of all the users.

Manuscript received March 24, 2012; revised June 28 and October 2, 2012; accepted October 6, 2012. The associate editor coordinating the review of this paper and approving it for publication was Z. Han.

M. Egan is with the School of Electrical and Information Engineering, The University of Sydney, NSW 2006, Australia (e-mail:m.egan@sydney.edu.au).

P. L. Yeoh is with the Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, VIC 3010, Australia (e-mail: phee.yeoh@unimelb.edu.au).

M. Elkashlan is with the School of Electronic Engineering and Computer Science, Queen Mary, University of London, London E1 4NS, UK (e-mail: maged.elkashlan@eecs.qmul.ac.uk).

I. B. Collings is with the Wireless and Networking Technologies Laboratory, CSIRO ICT Centre, Marsfield, NSW 2122, Australia (e-mail: iain.collings@csiro.au).

Digital Object Identifier 10.1109/TWC.2012.120312.120433

Cross-layer scheduling has been successfully applied to other wireless multimedia networks. There are two main approaches: low-complexity heuristic policies [9–12], and network utility maximization (NUM) based algorithms [13, 14]. A common example of the heuristic approach is the weighted sum-rate scheduler, where the weight is dependent on a function of total packet delay [9] or the number of packets in the queue [10]. In contrast, NUM maximizes the average utility subject to a long-term minimum service rate for each user [13, 14]. While NUM achieves long-term utility optimality, it requires the repeated solution of a convex optimization problem using iterative sub-gradient methods. This may not be practical in low-complexity systems.

In this paper, we propose a new cross-layer scheduler for the WMRN that falls within the heuristic approaches. Our new scheduler is specifically designed for the multi-user situation. Unlike the single user case, the overall delay is not only affected by the length of the queue, but can in fact be dominated by the channel state for each user, relative to all the others. To address this, we propose a new scheduler based on a weighted function of the signal-to-noise ratio (SNR) for each user, and what we call the delay in packet scheduling (DPS) for each user. We define the DPS for a user as the number of potential scheduling opportunities that have elapsed for the packet at the head of the user's queue.

To demonstrate the performance of the proposed scheduler, we apply it to the popular WMRN based on cooperative dual-hop amplify-and-forward (AaF) relaying. The majority of works in the literature related to scheduling in these WMRNs consider opportunistic equal weight scheduling, where the user with the highest end-to-end SNR is scheduled (see, for example, [15–17] and the citations therein). Of particular interest is [15] where the outage probability and the symbol error probability (SEP) of opportunistic equal weight scheduling were derived. For the proposed scheduler, we make the important observation that the DPS evolves according to a Markov chain, and use this to calculate both the PMF of the DPS and the SEP. We also calculate the buffer state (number of packets) and the packet loss probability due to overflow. These quantities are required to calculate the average total delay of each packet and the throughput.

We analyze the network under the assumption of independent and identically distributed Rayleigh fading and homogeneous Poisson arrivals. We first derive the probability mass function (PMF) of the DPS by constructing the augmented transition matrix of a truncated multidimensional Markov

chain. We then derive an accurate approximation of the average SEP of the scheduled user with uncoded symbols. Our analysis holds for an arbitrary number of users K and any scheduling policy where the weights are a function of the DPS for each user. We then perform simulations that confirm the analysis. We show that the proposed scheduler can achieve a 30% improvement in the probability the DPS satisfies a given constraint, with negligible degradation in the SEP compared to the opportunistic equal weight scheduler in [15].

To further characterize the performance of the scheduler, we derive a new analytical expression for the PLP due to buffer overflow. Currently, this has not been addressed in the multiuser scheduling scenario we have considered. We show via simulation that our scheme can outperform the equal weight scheduler [15] by 85%. We demonstrate via our analysis and simulations that the PLP does not always improve with increasing buffer size. We show that this problem can be avoided by decreasing the transmission time.

Finally, we simulate the throughput when a convolutional code is used. The throughput is dependent on the SEP of the scheduled user, and the average PLP. We highlight that when the transmission time is sufficiently small, the SEP dominates the throughput. In contrast, when the transmission time is sufficiently large, the PLP dominates. Our simulations show that the throughput can be significantly improved by optimizing the transmission time and the code rate. In particular, the throughput can be increased by over 20% when the transmission time is reduced from 4 ms to 2 ms, at a code rate of $1/4$.

This paper is organized as follows. Section II details the system model and notation. Section III introduces the proposed scheduler. Section IV characterizes the delay performance of our scheduler in terms of the DPS. The SEP of the scheduled user is evaluated in Section V. Section VI presents the analysis of the buffer state and PLP due to buffer overflow. Simulations and numerical results are discussed in Section VII, and Section VIII concludes the paper.

II. System Model

We consider the WMRN where a base station (BS) with K first-in first-out (FIFO) data queues transmits to K corresponding users with the aid of an AaF[1] relay. In our cross-layer scheduling policy (to be described in Section III), a single user with the largest weighted SNR is scheduled for transmission in each scheduling opportunity. We assume independent non-identically distributed (i.n.d) block Rayleigh fading in the two-hop relay links with a coherence time of T_c seconds.

A. MAC Layer Architecture

The BS has K finite queues with buffer size B, each corresponding to a distinct user. A user's packet is lost if the buffer for the queue is full and a new packet arrives. The arrival process of the packets for each queue is assumed to be a homogeneous Poisson process with rate $\lambda_k,\ k = 1, \ldots, K$, where each k corresponds to a different queue. The probability that n packets arrive in an interval of time T for the k-th user is then given by

$$\Pr(N_k(T) = n) = \frac{e^{-\lambda_k T}(\lambda_k T)^n}{n!}. \tag{1}$$

Packets can be re-requested with the caveat that the arrival of the re-requested packet is consistent with the Poisson arrival process. The transmission time T is the same for all users.

Prior to Section VI, we assume that the queues are backlogged such that at least one packet is always available. As a result, the BS is never silent. This assumption is also made in [18]. We relax this restriction in Section VI where we derive the PMF of the buffer state and the PLP due to buffer overflow.

B. Physical Layer Architecture

The BS and the relay each transmit for $T/2$ seconds in half duplex mode such that the total transmission time from the BS to the scheduled user is T seconds, where $T \leq T_c$. The transmission time is chosen such that the BS has knowledge of both the BS-relay and relay-user links for scheduling purposes. In the BS-relay link, the received signal at the relay is given by

$$y_R = \sqrt{E_S} h_{SR} x + z_R, \tag{2}$$

where E_S is the transmit power at the source, h_{SR} is the Rayleigh fading channel coefficient between the source and the relay, x is the transmitted symbol using binary phase-shift keying (BPSK), quadrature phase-shift keying (QPSK) or M-ary pulse amplitude modulation (M-PAM), and z_R is the additive white Gaussian noise (AWGN) with one-sided power spectral density N_0. In the relay-to-user link, the received signal at the scheduled user, denoted by $k^* \in \{1, \ldots, K\}$, is given by

$$y_{k^*} = \sqrt{E_R} h_{Rk^*} \beta y_R + z_{k^*}, \tag{3}$$

where E_R is the transmit power at the relay, h_{Rk^*} is the Rayleigh fading channel between the relay and the scheduled user, z_{k^*} is the AWGN with one-sided power spectral density N_0, and β is the relay amplification factor defined as

$$\beta = \sqrt{\frac{1}{E_{SR}|h_{SR}|^2 + cN_0}}. \tag{4}$$

In (4), we set $c = 1$ for the case where noise power is included in the relay amplification factor and we set $c = 0$ for the case where the noise power is ignored [15].

The end-to-end SNR of the scheduled user is written as

$$\gamma_{eq} = \frac{\gamma_{SR}\gamma_{Rk^*}}{\gamma_{SR} + \gamma_{Rk^*} + c}, \tag{5}$$

where γ_{SR} is the instantaneous SNR in the source-to-relay link and γ_{Rk^*} is the instantaneous SNR in the relay-to-user link. We incorporate the effect of path loss into the instantaneous SNRs such that $\gamma_{SR} = d_S^{-\eta} E_S |h_{SR}|^2 / N_0$ and $\gamma_{Rk^*} = d_R^{-\eta} E_R |h_{Rk^*}|^2 / N_0$, where d_S is the distance between the source and the relay, d_R is the distance between the relay and the scheduled user, and η is the path loss exponent. As both the BS-relay and relay-to-user links experience i.n.d.

[1]AaF is a practical relaying protocol due to its ease of implementation and low power consumption since there is no need for digital decoding at the relay [15]. This makes it preferable over decode-and-forward in low-complexity implementations where digital hardware is not available.

Rayleigh fading, the probability density functions (PDFs) of the instantaneous SNRs are written as

$$f_{\gamma_{SR}}(\gamma) = \frac{1}{\overline{\gamma}_1} e^{-\frac{\gamma}{\overline{\gamma}_1}}, \tag{6}$$

$$f_{\gamma_{Rk}}(\gamma) = \frac{1}{\overline{\gamma}_2} e^{-\frac{\gamma}{\overline{\gamma}_2}}, \ k = 1, 2, \ldots, K. \tag{7}$$

where $\overline{\gamma}_1 = E[\gamma_{SR}]$ is the average SNR in the source-to-relay link and $\overline{\gamma}_2 = E[\gamma_{Rk}], \ \forall k \in \{1, \ldots, K\}$, is the average SNR in the relay-to-user links, with $E[\cdot]$ denoting the expectation.

III. Proposed Cross-Layer Scheduling Policy

In this section we detail our new user scheduling policy. The scheduling policy selects the user with the largest weighted SNR of the second hop. The weight is a function of the DPS. The DPS is formally defined as follows.

Definition 1. *The DPS of user k is the number of potential scheduling opportunities that have elapsed for the packet at the head of user ks queue.*

We note that a packet can only be scheduled at the front of a user's queue. As a result, only delays of the packets at the front of each user's queue are required for our scheduler's computations. The header size of each packet can then be significantly reduced in long queues compared with the scheme in [3], as time stamps with a small number of bits are sufficient. The reduction is due to the impact of the large variation in total packet delay on the scheme in [3], caused by the dependence on the number of packets in the queue when the packet arrives.

While our approach is heuristic, it is practical and efficient. Moreover, our policy does have a theoretical underpinning as we can guarantee the average SEP and PLP due to buffer overflow, as we will show in Sections V and VI. Since the modulation scheme is fixed as described in Section II-B, the instantaneous SEP only varies when the instantaneous SNR varies. As a result, only the instantaneous SNR is required in the scheduling policy to account for the SEP requirements.

The proposed scheduling policy is given by

$$k^* = \arg \max_{k=1,\ldots,K} \gamma_{Rk} W_k, \tag{8}$$

where

(i) k^* is the scheduled user;
(ii) γ_{Rk} are the instantaneous SNRs of the relay-to-user links, where $k = 1, \ldots, K$;
(iii) W_k is the weight satisfying

$$W_k(s_k) \geq 0, \ k = 1, \ldots, K \tag{9}$$

where $W_k(s_k)$ is an arbitrary function of s_k that is the number of potential scheduling opportunities that have elapsed for the k-th user's packet. We do not account for the BS-relay link in the scheduling policy as it is the same for all users.

Users in multimedia applications are often characterized as real time (RT) or best effort (BE). Our scheduling policy can accommodate both types of users by adapting the weight functions W_k. In particular, the weights of the BE users are constant $W_k = 1$, while the RT users weights are functions of the DPS. As a result, the probability that a RT user is scheduled increases when its packet has a high DPS. BE users are then served when the RT users' channels are poor or the RT users' packets packet have a low DPS.

IV. Delay Performance

In this section, we derive the statistics of the scheduling delay for each user when there is at least one packet in each queue. In particular, the statistics of the scheduling delay of each user are given by the stationary distribution of the multidimensional Markov chain.

A. Normalized Service Rate

We first derive the average normalized service rate for the k-th user, i.e., the probability that the k-th user is scheduled. Denote $P_k(\mathbf{s})$ as the normalized service rate when the users' queue states are the elements of the state vector $\mathbf{s} = [s_1, \ldots, s_K]^T$, where each $s_k, \ k = 1, \ldots, K$ denotes the number of scheduling opportunities that the packet for user k has been waiting at the front of the queue. The normalized service rate for user k in state $\mathbf{s}$, $P_k(\mathbf{s})$, is then given by the following lemma.

Lemma 1. *The probability that user k is scheduled queue state $\mathbf{s}$ is given by*

$$P_k(\mathbf{s}) = \sum_{\substack{i=1 \\ i \neq k}}^{K} \frac{W_k(s_k)}{W_k(s_k) + W_i(s_i)} + \sum_{\substack{i=1 \\ i \neq k}}^{K} \sum_{m=1}^{K-2} \sum_{\substack{n_p = n_{p-1}+1 \\ p=1,\ldots,m}}^{K-2-m+p} (-1)^m \times \frac{W_k(s_k)}{W_i(s_i) + W_k(s_k) + W_i(s_i) W_k(s_k) \sum_{j=1}^{m} \tilde{W}_{n_j}^{-1}(s_{n_j})},$$

where

$$(\tilde{W}_q(s_q))_{q=1}^{K-2} = (W_p(s_p))_{p=1, p \neq i,k}^{K}. \tag{10}$$

Proof: See Appendix A. ∎

Remark 1. *In the case where each weight is equal, Lemma 1 reduces to*

$$P_k = \frac{1}{K},$$

which shows that each user is equally likely to be scheduled. As a result, the normalized service rate can be interpreted as the relative priority of each user, and in special cases reduces to an extremely simple form.

B. Delay in Packet Scheduling

Next, we derive the statistics of the DPS. We require the probability that the current state vector is $\mathbf{s}$. Denote $\mathbf{s}(n)$ as the state vector after n transmission slots. The state vectors then form a Markov chain as

$$\Pr(\mathbf{s}(n) | \mathbf{s}(1), \ldots, \mathbf{s}(n-1)) = \Pr(\mathbf{s}(n) | \mathbf{s}(n-1)). \tag{11}$$

We note that the transition probability from state $\mathbf{s}(n-1)$ to state $\mathbf{s}(n)$ when user k is scheduled is given by $P_k(\mathbf{s}(n-1))$. Hence, the scheduler forms a K-dimensional Markov chain with a countably infinite state space. In general, the required eigenvalue equation is intractable and it is not possible to obtain closed form expressions. As a result, we approximate

the steady state characteristics by truncating the Markov chain and forming a 1-dimensional Markov chain with an augmented transition matrix. This technique for approximating the K-dimensional Markov chain is known as generating the augmented Markov chain. It has been well-studied and used in several applications such as [19]. We will show in Section VII that the approximation is accurate. The approximation proceeds as follows:

1) Determine the required maximum DPS for each user to achieve a given accuracy of the approximation. Denote the largest of these as d.
2) Enumerate in lexicographic order all possible state vectors with integer elements greater than or equal to one, with each element less than or equal to d.
3) Let V be the set of states that contain a single element $\mathbf{s}_k^i = 1$, where $\mathbf{s}_k^i$ is the k-th element of the i-th state vector in the lexicographic enumeration. We note that the set V can be written as

$$V = \{\mathbf{s}^i | \exists \text{ a unique } k \in \{1, 2, \ldots, K\} \text{ such that } \mathbf{s}_k^i = 1\}. \tag{12}$$

We then define S as

$$S = \{\mathbf{s}^j \in V | \mathbf{s}_m^j = \mathbf{s}_m^i + 1 \vee \mathbf{s}_m^j = \mathbf{s}_m^i = d, \forall m \neq k\},$$

where i is the index of the enumerated state for the current state vector and j is the index of the enumerated state for the future state vector. We then construct the transition probability matrix $\mathbf{P}$ as

$$p_{ij} = \begin{cases} P_k(\mathbf{s}^i), & \mathbf{s}^j \in S, \\ 0, & \text{otherwise} \end{cases}, \tag{13}$$

where p_{ij} is the (i, j)-th element of $\mathbf{P}$.

4) Adjust p_{i1} such that $\sum_j p_{ij} = 1$ for all i. This ensures that $\mathbf{P}$ is a stochastic matrix. Note that for sufficiently large d, this adjustment is small.

We note that the augmented matrix $\mathbf{P}$ represents a finite 1-dimensional Markov chain, which is irreducible and ergodic. As a result, the stationary distribution π of the Markov chain exists and is unique. We then obtain the steady state characteristics using the stationary distribution given by $\pi = \pi\mathbf{P}$. The stationary distribution obtained from the augmented transition matrix is an approximation for the probability of the scheduler being in any valid state $\mathbf{s}$ when the system is in steady state.

The PMF of the DPS for user k is the probability that user k is in state s_k. This is given by

$$\Pr(s_k = l) = \sum_{\substack{s_n=1 \\ n=1,\ldots,K}}^{\infty} \pi_{\mathbf{s}} \mathbf{1}_{\{s_k=l\}}, \tag{14}$$

where $\mathbf{1}$ is the indicator function and $\pi_{\mathbf{s}}$ is the element of the stationary distribution corresponding to state $\mathbf{s}$. In the case where the truncated Markov chain is used, the PMF of the DPS for user k is well approximated by

$$\Pr(s_k = l) \approx \sum_{\substack{s_n=1 \\ n=1,\ldots,K}}^{d} \pi_{\mathbf{s}} \mathbf{1}_{\{s_k=l\}}, \tag{15}$$

when d is sufficiently large. This result is crucial as it forms the basis of the SEP and PLP analysis in the sequel.

V. Error Performance

In this section, we derive a new expression for the SEP of the scheduled user in backlogged user queues. The expression takes the form of a single-dimensional integral that can efficiently be evaluated using numerical techniques.

A. Channel Statistics

To calculate the SEP of the scheduled user, we require the cumulative distribution function (CDF) of the end-to-end SNR. As the CDF is dependent on the state vector $\mathbf{s}$, we first calculate the conditional CDF. The expression for the conditional CDF of the SNR of the relay-user link of the scheduled user k^* is given in Lemma 2. To simplify the notation, we write $\gamma_2 = \gamma_{Rk^*}$ for the SNR of the scheduled user's relay-to-user link.

Lemma 2. *The conditional CDF of the second hop $\gamma_2 = \gamma_{Rk^*}$ is given by (16)*

Proof: See Appendix B. ∎

We now derive the conditional CDF of the end-to-end SNR of the source-relay-user link. The conditional CDF of the end-to-end SNR, γ_{eq} in (5), is given in terms of the source-relay and relay-user link SNRs. Using a result in [15], the conditional CDF of the end-to-end SNR may be written as

$$F_{\gamma_{eq}|\mathbf{s}}(\gamma|\mathbf{s}) = 1 - \int_0^{\infty} \left(1 - F_{\gamma_2|\mathbf{s}}\left[\gamma + \frac{\gamma^2 + c\gamma}{\omega}|\mathbf{s}\right]\right) \times f_{\gamma_1|\mathbf{s}}(\omega + \gamma|\mathbf{s}) d\omega, \tag{17}$$

where $\gamma_1 = \gamma_{SR}$.

Using (17), we derive the conditional CDF of the end-to-end SNR conditioned on the state vector $\mathbf{s}$ given in Lemma 3.

Lemma 3. *The conditional CDF of the end-to-end SNR conditioned on $\mathbf{s}$ is given by (18) where*

$$G(T_i) = e^{-\gamma T_i} 2\sqrt{\overline{\gamma}_1 T_i(\gamma^2 + c\gamma)} K_1\left(2\sqrt{\frac{T_i(\gamma^2 + c\gamma)}{\overline{\gamma}_1}}\right),$$

with $T_1 = T_4 = \frac{1}{\overline{\gamma}_2}$, $T_2 = \frac{1}{\overline{\gamma}_2}\left(1 + \frac{W_k}{W_i}\right)$, $T_3 = \frac{W_k}{\overline{\gamma}_2}\left(\frac{1}{W_i} + \frac{1}{W_k}\right)$, $T_5 = \left(\frac{1}{\overline{\gamma}_2} + \frac{W_k}{\overline{\gamma}_2}\left(\frac{1}{W_i} + \sum_{j=1}^{m} \frac{1}{\tilde{W}_{n_j}}\right)\right)$, $T_6 = \frac{W_k}{\overline{\gamma}_2}\left(\frac{1}{W_k} + \frac{1}{W_i} + \sum_{j=1}^{m} \tilde{W}_{n_j}^{-1}\right)$, *and K_1 is the first order modified Bessel function of the second kind.*

Proof: See Appendix C. ∎

We now present the average CDF of the end-to-end SNR in Theorem 1, which is obtained by summing over the result in Lemma 3.

Theorem 1. *The average CDF of the end-to-end SNR is given by*

$$F_{\gamma_{eq}}(\gamma) = \sum_{s_1,\ldots,s_K} F_{\gamma_{eq}|\mathbf{s}}(\gamma|\mathbf{s}) \pi_{\mathbf{s}},$$

where $F_{\gamma_{eq}|\mathbf{s}}(\gamma|\mathbf{s})$ is given in (18) and $\pi_{\mathbf{s}}$ is the probability that the DPS state vector is $\mathbf{s}$.

$$F_{\gamma_2|\mathbf{s}}(\gamma|\mathbf{s}) = \sum_{k=1}^{K}\sum_{\substack{i=1\\i\neq k}}^{K}\frac{1}{W_i}\left[-e^{-\frac{\gamma}{\overline{\gamma}_2}}W_i\left(1-e^{-W_k\frac{\gamma}{\overline{\gamma}_2}/W_i}\right)+\frac{1}{\frac{1}{W_i}+\frac{1}{W_k}}\left(1-e^{-W_k\frac{\gamma}{\overline{\gamma}_2}(1/W_i+1/W_k)}\right)\right]$$
$$+\sum_{k=1}^{K}\sum_{\substack{i=1\\i\neq k}}^{K}\sum_{m=1}^{K-2}\sum_{\substack{n_p=n_{p-1}+1\\p=1,\ldots,m}}^{K-2-m+p}\frac{(-1)^m}{W_i}\left[-\frac{e^{-\frac{\gamma}{\overline{\gamma}_2}}}{\frac{1}{W_i}+\sum_{j=1}^{m}\tilde{W}_{n_j}^{-1}}\left(1-e^{-W_k\frac{\gamma}{\overline{\gamma}_2}\left(\frac{1}{W_i}+\sum_{j=1}^{m}\frac{1}{W_{n_j}}\right)}\right)\right.$$
$$\left.+\frac{1}{\sum_{j=1}^{m}\tilde{W}_{n_j}^{-1}+\frac{1}{W_k}+\frac{1}{W_i}}\left(1-e^{-W_k\frac{\gamma}{\overline{\gamma}_2}\left(\frac{1}{W_k}+\frac{1}{W_i}+\sum_{j=1}^{m}\tilde{W}_{n_j}^{-1}\right)}\right)\right] \tag{16}$$

$$F_{\gamma_{eq}|\mathbf{s}}(\gamma|\mathbf{s}) = 1-e^{-\frac{\gamma}{\overline{\gamma}_1}}+\sum_{k=1}^{K}\sum_{\substack{i=1\\i\neq k}}^{K}\frac{1}{W_i}\left(-W_i\frac{1}{\overline{\gamma}_1}e^{-\frac{\gamma}{\overline{\gamma}_1}}G(T_1)+W_i\frac{1}{\overline{\gamma}_1}e^{-\frac{\gamma}{\overline{\gamma}_1}}G(T_2)+\frac{W_iW_k}{W_i+W_k}e^{-\frac{\gamma}{\overline{\gamma}_1}}\right.$$
$$\left.-\frac{W_iW_k}{W_i+W_k}\frac{1}{\overline{\gamma}_1}e^{-\frac{\gamma}{\overline{\gamma}_1}}G(T_3)\right)+\sum_{k=1}^{K}\sum_{\substack{i=1\\i\neq k}}^{K}\sum_{m=1}^{K-2}\sum_{\substack{n_p=n_{p-1}+1\\p=1,\ldots,m}}^{K-2-m+p}\frac{(-1)^m}{W_i}\left(-\frac{1}{\left(\frac{1}{W_i}+\sum_{j=1}^{m}\tilde{W}_{n_j}^{-1}\right)}\frac{1}{\overline{\gamma}_1}e^{-\frac{\gamma}{\overline{\gamma}_1}}G(T_4)\right.$$
$$+\frac{1}{\left(\frac{1}{W_i}+\sum_{j=1}^{m}\tilde{W}_{n_j}^{-1}\right)}\frac{1}{\overline{\gamma}_1}e^{-\frac{\gamma}{\overline{\gamma}_1}}G(T_5)+\frac{1}{\sum_{j=1}^{m}\tilde{W}_{n_j}^{-1}+\frac{1}{W_k}+\frac{1}{W_i}}e^{-\frac{\gamma}{\overline{\gamma}_1}}$$
$$\left.-\frac{1}{\overline{\gamma}_1}e^{-\frac{\gamma}{\overline{\gamma}_1}}\frac{1}{\sum_{j=1}^{m}\tilde{W}_{n_j}^{-1}+\frac{1}{W_i}+\frac{1}{W_k}}G(T_6)\right), \tag{18}$$

B. Symbol Error Probability

Following [15], the SEP of the scheduled user for different modulation formats can be evaluated according to

$$P_S = \frac{a}{2}\sqrt{\frac{b}{\pi}}\int_0^{\infty}\gamma^{-\frac{1}{2}}e^{-b\gamma}F_{\gamma_{eq}}(\gamma)\,d\gamma. \tag{19}$$

The constants a and b are modulation-specific with $a = 1,\ b = 1$ for BPSK, $a = 1,\ b = 0.5$ for QPSK, and $a = 2(M-1)/M,\ b = 3/\left(M^2-1\right)$ for M-PAM.

We note that (19) is absolutely convergent. As such, we can swap the sum in Theorem 1 and the integral in (19) by applying the dominated convergence theorem. This ensures that the infinite sum converges. The integral can then be evaluated efficiently using numerical integration, leading to reduced evaluation time compared with Monte Carlo simulation. In Section VII, we show that the analytical approach also accurately approximates the Monte Carlo simulation.

VI. Packet Loss Performance

In this section, we analyze the PLP of each queue using the proposed scheduling policy. This is achieved by constructing a new Markov chain for the buffer states for each queue with transition probabilities dependent on the scheduling policy, arrival rate, and transmission time.

A. Buffer State

We first obtain the PMF of the buffer state that gives the probability that the buffer has $l,\ 0 \le l \le B$ packets. We note that the buffer state is measured at the beginning of a scheduling slot, after a packet is scheduled in the current slot, and before new arrivals. This is important as the time when the buffer state is measured affects the PMF of the buffer state and subsequently the PLP. We also note that the buffer state is independent of the DPS corresponding to user k. To calculate the PMF of the buffer state, we require the average probability that user k is scheduled, which is given by

$$P_k = \sum_{\substack{s_n=1\\n=1,\ldots,K}}^{\infty}P_k(\mathbf{s})\pi_{\mathbf{s}} \tag{20}$$

where $P_k(\mathbf{s})$ is the probability user k is scheduled in DPS state $\mathbf{s}$ and $\pi_{\mathbf{s}}$ is the probability that the DPS state vector is $\mathbf{s}$. An accurate approximation of (20) can be obtained using the stationary distribution arising from the truncated multidimensional Markov chain given in (15).

To obtain the PMF of the buffer state, we require the stationary distribution of the associated Markov chain. Since the buffer state of each user is only dependent on the individual user's statistics, the Markov chain is one-dimensional and the stationary distribution can be obtained by explicitly constructing the transition matrix. In particular, the construction of the transition matrix for user k, $\mathbf{T}_k$, is shown in Table 1. In the table, t_{ij} is the (i,j)-th element of $\mathbf{T}_k$ where $1 \le i,j \le B+1$. The (i,j)-th element represents the transition from the buffer storing $i-1$ packets to the buffer storing $j-1$ packets. We note that since the buffer is finite, the corresponding irreducible and ergodic Markov chain is also finite. As a result, it has a unique stationary distribution.

In order to analyze the PMF of the buffer state, we assume that it is possible for a user to be scheduled without a packet. This is necessary as the weight in our scheduling policy does not account for the buffer state. We show via simulations in Section VII that the approximation is accurate in spite of this

TABLE I
CONSTRUCTION OF TRANSITION MATRIX $\mathbf{T}_k$

Case 1: $j = i$ i) if $j = 1$, then $t_{ij} = \Pr(N_k(T) = 0) + \Pr(N_k(T) = 1)P_k$ ii) if $j < B$, then $t_{ij} = \Pr(N_k(T) = 0)(1 - P_k)\Pr(N_k(T) = 1)P_k$ iii) if $j = B$, then $t_{ij} = (1-P_k)\Pr(N_k(T) = 0)+(1-\Pr(N_k(T) = 0))P_k$
Case 2: $j = i - 1$ i) if $i < B + 1$, then $t_{ij} = \Pr(N_k(T) = 0)P_k$ ii) if $i = B + 1$, then $t_{ij} = P_k$
Case 3: $j < i - 1$ i) $t_{ij} = 0$
Case 4: $j > i$ i) if $j < B$, then $t_{ij} = \Pr(N_k(T) = j - i)(1 - P_k) + \Pr(N_k(T) = j - i + 1)P_k$ ii) if $j = B$, then $t_{ij} = \Pr(N_k(T) = j - i)(1 - P_k) + (1 - \Pr(N_k(T) \leq B - (i - 1) - 1))P_k$
Case 5: $j = B + 1$ i) if $i < B+1$, then $t_{ij} = (1-\Pr(N_k(T) \leq B-(i-1)-1)(1-P_k)$ ii) if $i = B + 1$, then $t_{ij} = 1 - P_k$

assumption. The PMF of the buffer state approximation for user k is obtained by solving the eigenvalue equation

$$\nu_{B,k} = \nu_{B,k}\mathbf{T}_k. \tag{21}$$

Equation (21) is solved by finding the left eigenvector of $\mathbf{T}_k$ corresponding to an eigenvalue of one. We note that if the scheduling weights are constants independent of the DPS, then (21) provides the exact PMF of the buffer state. This is due to the scheduler not requiring the current DPS of each packet.

B. Packet Loss Probability

The PLP is the probability that a packet is lost due to buffer overflow. Before evaluating the PLP for a given packet, we require the following well known lemma [20, pg. 22].

Lemma 4. *Let the packet arrivals follow a homogeneous Poisson process. Then, the unordered packet arrival times are i.i.d uniformly on $(0, T)$ when conditioned on the number of arrivals in the interval.*

The PLP for each user can now be obtained for a given buffer size by considering the probability that the buffer is full at time $0 < t < T$ after a scheduling opportunity. Here, t is the the time of the new packet arrival. Theorem 2 gives an approximation of the PLP. The approximation arises due to dependence on the stationary distribution (21) and is exact when the scheduling policy weights are fixed constants.

Theorem 2. *The PLP for user k is given by*

$$P_{L,k} \approx \sum_{l=0}^{B} \nu_l \left(1 - \sum_{m=0}^{B-l-1} \frac{\lambda_k^m}{Tm!} \times \left[\frac{m!}{\lambda_k^{m+1}} - e^{-\lambda_k T} \sum_{n=0}^{m} \frac{m!}{n!} \frac{T^n}{\lambda_k^{m-n+1}} \right] \right) \tag{22}$$

where ν_l is the element of the stationary distribution corresponding to a buffer state l for user k. Moreover, $P_{L,k}$ is exact when the weights $W_k(s_k),\ k = 1, \ldots, K$ are constants.

Proof: See Appendix D. ∎

The PLP approximation shows the clear dependence on the the transmission time and arrival rates for the user under consideration. Intuitively, if the arrival rate is high or the transmission time long, the PLP due to buffer overflow is large. We will see in Section VII-C that a consequence of this is that additional redundancy through channel coding does not always improve the throughput.

We note that the expected total packet delay can be obtained via Little's law [21] from the buffer state distribution and the PLP. In particular, we have

$$E[W_k] = L_k/\lambda_{e,k}, \tag{23}$$

where $E[W_k]$ is the expected total packet delay for user k, L_k is the expected number of packets in the queue for user k, which can be obtained using (20), and $\lambda_{e,k}$ is the effective arrival rate for user k given by

$$\lambda_{e,k} = \lambda_k(1 - P_{L,k}), \tag{24}$$

where λ_k is the actual packet arrival rate and $P_{L,k}$ is the PLP of the user given by (21). Surprisingly, this means that we can quantify the expected total packet delay, despite only exploiting the DPS in our scheme.

In the next section, we use numerical and simulation results to evaluate the effect of scheduling policy and examine the accuracy of the approximation when the weights are not fixed.

VII. SIMULATION AND NUMERICAL RESULTS

A. DPS and SEP

We now confirm our analytical results for the PMF of the DPS and the SEP of the scheduled user using Monte Carlo simulations. Simulations are performed with distances $d_S = d_R = 1$ and transmit powers $E_S = E_R = 1$. The WMRN consists of a single RT user and $K - 1$ BE users. To illustrate the behavior of our scheduler, we consider the following exponential-based scheduling policies for the weights of the RT user:

$$W_1 = e^{0.07(j-1)s_1} \tag{25}$$

for $K = 3$, and

$$W_1 = e^{0.06(j-1)s_1} \tag{26}$$

for $K = 4$, where $j \in \{1, 2, \ldots\}$ is the scheduling policy index that is a reference number for each policy. The weights are exponential functions of s_1, which is the number of scheduling opportunities that the packet for the RT user has been waiting at the front of the queue. The weights of the BE users are set to $W_i = 1,\ i = 2, \ldots, K$ that means there are no delay constraints.

In Fig. 1, the probability that the DPS target of $\Pr(s_1 \leq 3)$ is satisfied for the RT user is plotted against the scheduling policy index. We point out the accuracy of the approximation in (15) by noting that the simulations curves are consistent with the analytical curves. The figure shows that the probability of satisfying the DPS target is greater than 90% for $j \geq 5$

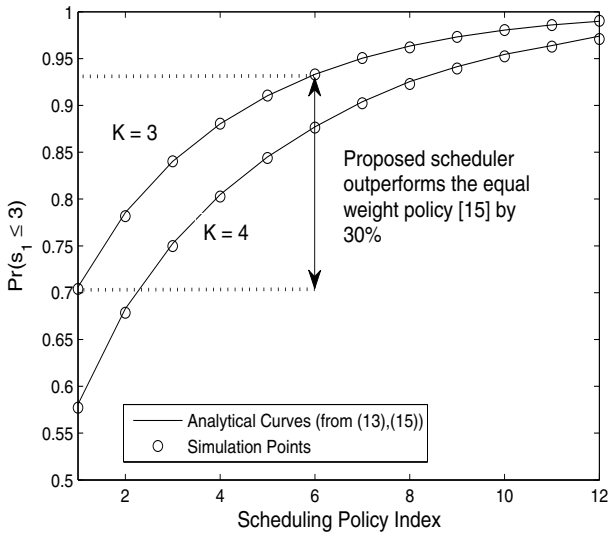

Fig. 1. The probability the DPS target of $s_1 \leq 3$ is satisfied for the RT user versus the scheduling policy index of several policies.

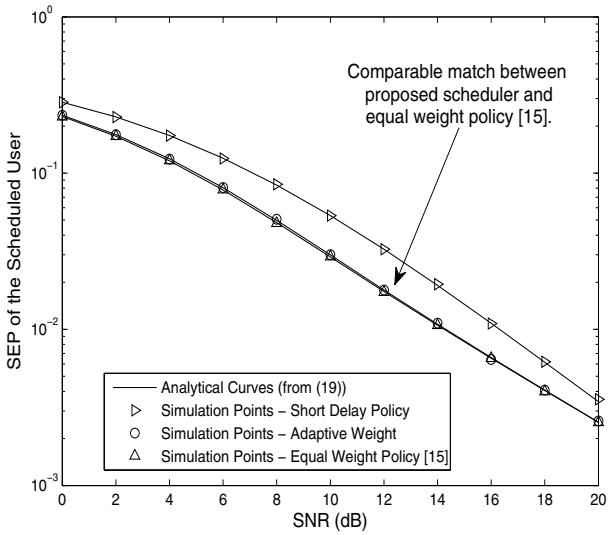

Fig. 2. The SEP of the scheduled user (k^* in (8)) versus the SNR using BPSK with $K = 3$. Three scheduling policies are considered: short delay, adaptive weight and equal weight.

when $K = 3$ and for $j \geq 8$ when $K = 4$. When $K = 3$, we see that the adaptive weight policy of $j = 6$ provides a 30% improvement in satisfying the DPS target relative to the equal weight policy of $j = 1$, which was considered in [15]. This improvement is indicated in the figure by the arrow. We observe that the probability of satisfying the DPS target increases as the scheduling policy index increases. The effect of this on the SEP of the scheduled user is further examined in Figs. 2 and 3.

In Fig. 2, the SEP of the scheduled user (i.e., k^* in (8)) is plotted against the average SNR $\overline{\gamma}_1$. We plot the SEP of the scheduled user for $K = 3$ with the adaptive weight policy of $j = 6$ in (25) and the equal weight policy in of $j = 1$ in (25). For comparison purposes we also plot the SEP of the scheduled user for a short delay policy where $W_1(s_1) = 100$. The short delay policy closely approximates a single user relay network where the probability of satisfying the DPS target is greater than 99% (evaluated using (15)). We point out the accuracy of the approximation in (19) by noting that the simulation and analytical curves are consistent. The figure shows that the SEP of the adaptive weight policy and the equal weight policy outperform the short delay policy by 2 dB at the SEP of 10^{-2}. We also observe that the adaptive weight policy provides a comparable match to the equal weight policy. This means that the 30% improvement in the probability of satisfying the DPS target of the RT user in Fig. 1 is attained with negligible degradation in the SEP of the scheduled user in Fig. 2.

In Fig. 3, the SEP of the scheduled user at $\overline{\gamma}_1 = 10$ dB is compared with the scheduling policy index in (25) for $K = 3$. We observe that an increase in the scheduling policy index causes an increase in the SEP of the scheduled user. Moreover, we observe that the SEP for the adaptive policies of $j \leq 6$ is less than 4% greater than the SEP for the equal weight policy of $j = 1$. This is consistent with the result in Fig. 2 that showed a comparable match between the SEP of the scheduled user using policies $j = 6$ and $j = 1$.

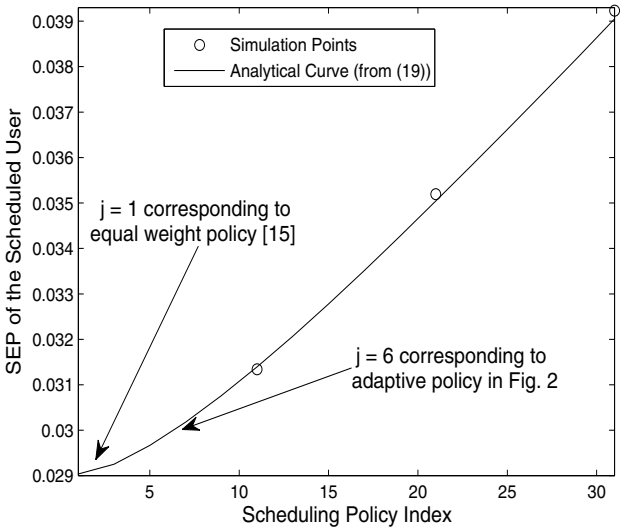

Fig. 3. The SEP of the scheduled user at 10 dB versus the scheduling policy index with $K = 3$ using BPSK.

B. PLP

In Fig. 4, the PLP with $K = 3$ is plotted for varying buffer sizes. The arrival rates are set as $\lambda_k = 0.1$ ms^{-1}, $k = 1, 2, 3$. Two scheduling policies are plotted: an adaptive policy where $W_1(s_1) = e^{0.2s_1}$, $W_2(s_2) = e^{0.1s_2}$ and $W_3 = 1$, and the equal weight policy [15]. We observe that analytical and simulation results are in good agreement for all users and scheduling policies. As expected, the users with higher priorities (i.e., users 1 and 2 of the adaptive policy) have a lower PLP. For $B > 7$, we observe that the PLP is independent of the buffer size for all users with $T = 5$ ms. This is due to the fact that when $T = 5$ ms, the average scheduling rate P_k/T for each user is less than the average arrival rate $\lambda_k = 0.1$ ms^{-1}, $\forall k$. As a result, the queue experiences a positive drift and

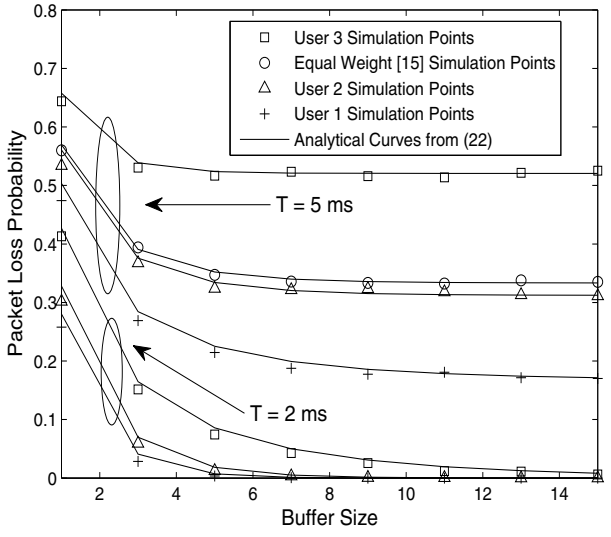

Fig. 4. The PLP for each user versus the buffer size with $K = 3$, transmission times $T = 5$ ms and $T = 2$ ms and arrival rate $\lambda_k = 0.1$ ms^{-1}, $k = 1, 2, 3$. Two scheduling policies are considered: equal weight ($W_i = 1,\ i = 1, 2, 3$) and adaptive ($W_1 = e^{0.2s_1},\ W_2 = e^{0.1s_2},\ W_3 = 1$).

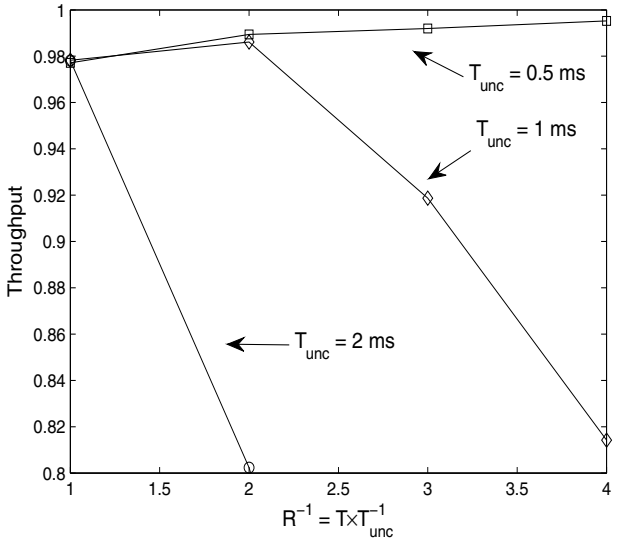

Fig. 5. The throughput of an equivalent single user network versus the inverse of the code rate R^{-1} for varying uncoded transmission times T_{unc}, arrival rate $\lambda_k = 0.1$ ms^{-1}, $k = 1, 2, 3$, and scheduling policy ($W_1 = e^{0.2s_1},\ W_2 = e^{0.1s_2},\ W_3 = 1$)

the buffer overflows regardless of the length. In contrast, the PLP of each user approaches zero with increasing buffer size for $T = 2$ ms. This is consistent with the results for queue stability in the infinite buffer scenario (see for example, Lemma 1 in [22, 23]). This means that the transmission time should be minimized to improve the service rate and allow large buffer sizes to be exploited.

To obtain further insights, we compare the PLP of the RT user with the adaptive weight policy considered in Fig. 2 and the equal weight policy [15]. The PLP of the RT user with the adaptive weight policy of $j = 6$ in (22) is evaluated using (21) as $P_L = 0.04$, for a buffer size $B = 15$, transmission time $T = 5$ ms, and arrival rate $\lambda_k = 0.1$ ms $k = 1, 2, 3$. We observe that this results in a greater than a 85% improvement over the PLP of the equal weight policy plotted in Fig. 4. This shows that our scheduler can achieve a significant improvement in the PLP compared with the equal weight scheduler, with negligible degradation in the SEP of the scheduled user.

C. Transmission Time

We observed in Fig. 4 that increasing the transmission time impacts on the PLP. To determine the optimal transmission time, the effect of channel coding must be accounted for. Of course, when employing coding, a longer transmission time is required to account for the redundancy in the signal. To examine the tradeoff between the coded SEP of the scheduled user and the PLP for each queue we consider the throughput given by

$$\text{Throughput} = (1 - P_{E,coded})(1 - P_{L,ave}), \qquad (27)$$

where $P_{E,coded}$ is the coded SEP of the scheduled user and $P_{L,ave}$ is the average PLP over all queues. The average PLP over all the queues is given by

$$P_{L,ave} = \frac{1}{K}\sum_{i=1}^{K} P_{L,i}. \qquad (28)$$

The throughput expression in (27) approximates the WMRN as a single point-to-point link using a single queue with a PLP given by the average over all queues. As a result, (27) gives a simple characterization of a WMRN as transmission times are varied.

In Fig. 5, the throughput is compared to the inverse of the code rate R^{-1}, for varying uncoded transmission time T_{unc}. Here $T = T_{unc}R^{-1}$, where R is the normalized rate of the coded signal and the rate of the uncoded signal is $R = 1$. The data is coded using punctured convolutional codes. The distances and transmit powers are $d_S = d_R = 1$ and $E_S = E_R = 1$. In the simulation, the buffer size is $B = 20$, the arrival rate is $\lambda_k = 0.1$ ms^{-1}, $k = 1, 2, 3$, the number of transmitted QPSK data symbols is 100, and the scheduling policy is $W_1(s_1) = e^{0.2s_1},\ W_2(s_2) = e^{0.1s_2},\ W_3 = 1$.

We observe in the figure that the throughput increases with increasing transmission time, T, when $T_{unc} = 0.5$ ms. In contrast, the throughput decreases with increasing T when $T_{unc} = 2$ ms. Of particular interest is the scenario where T_{unc} is between 0.5 ms and 2 ms. From the figure we see that when $T_{unc} = 1$ ms, the throughput does not vary monotonically with T. This suggests that an efficient tradeoff between the PLP and the coded SEP of the scheduled user for WMRNs with finite buffers is critically dependent on T_{unc}. This leads to the following key design rules: 1) in the small T_{unc} regime, the code rate should be minimized, 2) in the large T_{unc} regime, the code rate should be maximized and 3) in the intermediate T_{unc} regime, the code rate should lie between $0 < R < 1$.

VIII. Conclusion

We proposed a new user scheduler for WMRNs. We considered a cross-layer scheduling approach that accounted for delay in packet scheduling (DPS), symbol error probability (SEP), and packet loss probability (PLP) due to buffer overflow. The user with the largest weighted SNR was scheduled, where the weight was a function of the DPS. We then derived new analytical expressions for the DPS and the SEP of the scheduled user. We provided analysis and simulation to confirm the performance of the scheduler. We show the probability that a target DPS is met is 30% higher for our new scheme compared to the standard opportunistic equal weight scheduler, with negligible degradation in the SEP of the scheduled user. We show that this can lead to a 85% improvement in the PLP. Finally, we exploited channel coding and the transmission time to optimize the tradeoff between the PLP, and the SEP of the scheduled user. This revealed new design rules for the transmission time when coding is used.

Appendix A
Proof of Lemma 1

Without loss of generality, consider the case where $\overline{\gamma}_{Rm} = 1,\ m = 1, 2, \ldots, K$. The probability that the k-th user is selected in state $\mathbf{s}$ (corresponding to weights $W_k = W_k(s_k)$) is given by

$$P_k(\mathbf{s}) = \int_0^\infty \Pr(W_k\gamma_{Rk} > x | \Gamma_{k^-} = x) f_{\Gamma_{k^-}}(x)dx, \quad (29)$$

where $f_{\Gamma_{k^-}|\mathbf{s}}(x|\mathbf{s})$ is the pdf of $\Gamma_{k^-} = \max\{W_m\gamma_{Rm}\}_{m=1,m\neq k}^K$. The pdf of Γ_{k^-} is derived by differentiating

$$\Pr(\Gamma_{k^-} \leq x|\mathbf{s}) = \prod_{\substack{m=1\\m\neq k}}^K \Pr(W_m\gamma_{Rm} \leq x) = \prod_{\substack{m=1\\m\neq k}}^K (1 - e^{-x/W_m}) \quad (30)$$

that results in

$$f_{\Gamma_{k^-}|\mathbf{s}}(x|\mathbf{s}) = \sum_{\substack{i=1\\i\neq k}}^K \frac{1}{W_i} e^{-x/W_i} \prod_{\substack{m=1\\m\neq i,k}}^K (1 - e^{-x/W_m}), \quad (31)$$

Substituting (31) into (29), noting that γ_{Rk} is exponentially distributed with $\overline{\gamma}_{Rk} = 1$, we obtain

$$P_k(\mathbf{s}) = \int_0^\infty e^{-x/W_k} \sum_{i=1,i\neq k}^K \frac{1}{W_i} e^{-x/W_i} \prod_{\substack{m=1\\m\neq i,k}}^K (1 - e^{-x/W_m}) \times \sum_{\substack{i=1\\i\neq k}}^K \frac{1}{W_i} \int_0^\infty e^{-x/W_k} e^{-x/W_i} \prod_{\substack{m=1\\m\neq i,k}}^K (1 - e^{-x/W_m}). \quad (32)$$

The product term in (32) may be expanded as

$$\prod_{\substack{m=1\\m\neq i,k}}^K (1 - e^{-x/W_m}) = 1 + \sum_{m=1}^{K-2} \sum_{\substack{n_p=n_{p-1}+1\\p=1,\ldots,m}}^{K-2-m+p} (-1)^m \times e^{\left(-\sum_{j=1}^m x\tilde{W}_{n_j}^{-1}\right)}, \quad (33)$$

where we re-expressed the weights as $(\tilde{W}_q)_{q=1}^{K-2} = (W_1, \ldots, W_{i-1}, W_{i+1}, \ldots, W_{k-1}, W_{k+1}, \ldots, W_K)$. Substituting (33) in (32) and evaluating the integrals, we obtain the desired result.

Appendix B
Proof of Lemma 2

We now derive the conditional CDF of the second hop SNR. We condition on the state vector $\mathbf{s}$ that results in weights $W_k = W_k(s_k)$. The CDF of the second hop channel gain $|h_{Rk}^*|^2$ of the scheduled user is evaluated according to

$$\begin{aligned} F_{\Gamma|\mathbf{s}}(x|\mathbf{s}) &= \sum_{k=1}^K \Pr(|h_{Rk}|^2 \leq x, W_k|h_{Rk}|^2 > \Gamma_{k^-}) \\ &= \sum_{k=1}^K \int_0^{W_k x} \Pr(\frac{y}{W_k} < |h_{Rk}|^2 \leq x | \Gamma_{k^-} = y) \\ &\quad \times f_{\Gamma_{k^-}}(y)dy, \end{aligned} \quad (34)$$

where $f_{\Gamma_{k^-}}(y)$ is given in (31). We note that the formulation in terms of $|h_k|^2$ is a special case of γ_{Rk} where $\overline{\gamma}_{Rm} = 1,\ m = 1, 2, \ldots, K$. Applying (31) and (6), we obtain

$$F_{\Gamma|\mathbf{s}}(x|\mathbf{s}) = \sum_{k=1}^K \sum_{\substack{i=1\\i\neq k}}^K \frac{1}{W_i} \int_0^{W_k x} \left[(1 - e^{-x}) - (1 - e^{-y/W_k})\right] \times e^{-y/W_i} \prod_{\substack{m=1\\m\neq i,k}}^K (1 - e^{-y/W_m})dy. \quad (35)$$

Using the same product expansion as in (33) results in

$$F_{\Gamma|\mathbf{s}}(x|\mathbf{s}) = \sum_{k=1}^K \sum_{\substack{i=1\\i\neq k}}^K \frac{1}{W_i} \int_0^{W_k x} dy \left[(1 - e^{-x}) - (1 - e^{-y/W_k})\right] \times e^{-y/W_i} \left(1 + \sum_{m=1}^{K-2} \sum_{\substack{n_p=n_{p-1}+1\\p=1,\ldots,m}}^{K-2-m+p} (-1)^m e^{-\sum_{j=1}^m y\tilde{W}_{n_j}^{-1}}\right). \quad (36)$$

After a number of simple integrations and algebraic simplifications, we obtain (37).

To obtain the CDF of the second hop SNR γ_2, we make the substitution $x = \frac{\gamma}{\overline{\gamma}_2}$.

Appendix C
Proof of Lemma 3

Substituting the conditional CDF of γ_2 in Lemma 2 and the PDF of γ_1 in (6) into (17), we obtain the following expression for the conditional CDF of the end-to-end SNR, given by

$$\begin{aligned} F_{\gamma_{eq}|\mathbf{s}}(\gamma|\mathbf{s}) &= 1 - e^{-\frac{\gamma}{\overline{\gamma}_1}} \\ &\quad + \sum_{k=1}^K \sum_{\substack{i=1\\i\neq k}}^K \frac{1}{W_i} \left(-W_i\Phi_1 + W_i\Phi_2 + \Phi_3 - \Phi_4\right) \\ &\quad + \sum_{k=1}^K \sum_{\substack{i=1\\i\neq k}}^K \sum_{m=1}^{K-2} \sum_{\substack{n_p=n_{p-1}+1\\p=1,\ldots,m}}^{K-2-m+p} \frac{(-1)^m}{W_i} \\ &\quad \times \left(-\Phi_5 + \Phi_6 + \Phi_7 - \Phi_8\right), \end{aligned} \quad (38)$$

$$F_{\Gamma|\mathbf{s}}(x|\mathbf{s}) = \sum_{k=1}^{K}\sum_{\substack{i=1\\ i\neq k}}^{K} \frac{1}{W_i}\left[-e^{-x}W_i\left(1-e^{-W_k x/W_i}\right) + \frac{1}{\frac{1}{W_i}+\frac{1}{W_k}}\left(1-e^{-W_k x(1/W_i+1/W_k)}\right)\right]$$
$$+ \sum_{k=1}^{K}\sum_{\substack{i=1\\ i\neq k}}^{K}\sum_{m=1}^{K-2}\sum_{\substack{n_p=n_{p-1}+1\\ p=1,\ldots,m}}^{K-2-m+p} \frac{(-1)^m}{W_i}\left[-\frac{e^{-x}\left(1-e^{-W_k x\left(\frac{1}{W_i}+\sum_{j=1}^{m}\tilde{W}_{n_j}^{-1}\right)}\right)}{\frac{1}{W_i}+\sum_{j=1}^{m}\tilde{W}_{n_j}^{-1}}\right.$$
$$\left. + \frac{\left(1-e^{-W_k x\left(\frac{1}{W_k}+\frac{1}{W_i}+\sum_{j=1}^{m}\tilde{W}_{n_j}^{-1}\right)}\right)}{\sum_{j=1}^{m}\tilde{W}_{n_j}^{-1}+\frac{1}{W_k}+\frac{1}{W_i}}\right] \tag{37}$$

where $\Phi_1 = \frac{1}{\overline{\gamma}_1}e^{-\frac{\gamma}{\overline{\gamma}_1}}\int_0^\infty e^{-\frac{\omega}{\overline{\gamma}_1}}e^{-\frac{1}{\overline{\gamma}_2}\left(\gamma+\frac{\gamma^2+c\gamma}{\omega}\right)}d\omega$, $\Phi_2 = \frac{1}{\overline{\gamma}_1}e^{-\frac{\gamma}{\overline{\gamma}_1}}\int_0^\infty e^{-\frac{\omega}{\overline{\gamma}_1}}e^{-\left(\frac{1}{\overline{\gamma}_2}+\frac{W_k}{\overline{\gamma}_2 W_i}\right)\left(\gamma+\frac{\gamma^2+c\gamma}{\omega}\right)}d\omega$, $\Phi_3 = \frac{1}{\overline{\gamma}_1}e^{-\frac{\gamma}{\overline{\gamma}_1}}\int_0^\infty e^{-\frac{\omega}{\overline{\gamma}_1}}\frac{W_i W_k}{W_i+W_k}d\omega$, $\Phi_4 = \frac{e^{-\frac{\gamma}{\overline{\gamma}_1}}}{\overline{\gamma}_1}\int_0^\infty \frac{e^{-\frac{\omega}{\overline{\gamma}_1}}W_i W_k}{W_i+W_k}e^{-\frac{W_k}{\overline{\gamma}_2}\left(\frac{1}{W_i}+\frac{1}{W_k}\right)\left(\gamma+\frac{\gamma^2+c\gamma}{\omega}\right)}d\omega$, $\Phi_5 = \frac{1}{\overline{\gamma}_1\left(W_i^{-1}+\sum_{j=1}^{m}\tilde{W}_{n_j}^{-1}\right)}$, $\Phi_6 = \frac{e^{-\frac{\gamma}{\overline{\gamma}_1}}}{\overline{\gamma}_1}\int_0^\infty \frac{e^{-\frac{\omega}{\overline{\gamma}_1}}e^{-\left(\gamma+\frac{\gamma^2+c\gamma}{\omega}\right)\left(\frac{1}{\overline{\gamma}_2}+\frac{W_k}{\overline{\gamma}_2}\left(\frac{1}{W_i}+\sum_{j=1}^{m}\tilde{W}_{n_j}^{-1}\right)\right)}}{\frac{1}{W_i}+\sum_{j=1}^{m}\tilde{W}_{n_j}^{-1}}d\omega$, $\Phi_7 = \frac{e^{-\frac{\gamma}{\overline{\gamma}_1}}}{\sum_{j=1}^{m}\tilde{W}_{n_j}^{-1}+\frac{1}{W_k}+\frac{1}{W_i}}$, $\Phi_8 = \frac{1}{\overline{\gamma}_1}\int_0^\infty \frac{e^{-\frac{\omega}{\overline{\gamma}_1}}e^{-\left(\gamma+\frac{\gamma^2+c\gamma}{\omega}\right)\frac{W_k}{\overline{\gamma}_2}\left(\frac{1}{W_k}+\frac{1}{W_i}+\sum_{j=1}^{m}\tilde{W}_{n_j}^{-1}\right)}}{\sum_{j=1}^{m}\tilde{W}_{n_j}^{-1}+\frac{1}{W_i}+\frac{1}{W_k}}d\omega$.

The integrals in Φ_1, Φ_2, Φ_3, Φ_4, Φ_6, and Φ_8 are solved using the identity $\int_0^\infty e^{-\frac{\omega}{a}-\frac{b}{\omega}}d\omega = 2\sqrt{\frac{a}{b}}K_1\left(2\sqrt{\frac{b}{a}}\right)$ where $K_1(x)$ is the modified Bessel function of the second kind.

APPENDIX D
PROOF OF THEOREM 2

We observe that if there is a new packet arrival at time t, the PLP of a given user can be expressed as $P_L = \Pr(\text{buffer is full for new arrival at time } t)$. Applying Lemma 4 and conditioning on the arrival time t and the number of packets in the buffer l, we approximate the PLP as (which is an exact result for fixed weights)

$$P_L \approx \sum_{l=0}^{B}\int_0^T \sum_{m=B-l}^{\infty} e^{-\lambda t}\frac{(\lambda t)^m}{m!}\nu_l\frac{1}{T}dt \tag{39}$$

The final result is obtained by using (1) and solving the integral using eq. 3.351 in [24]

REFERENCES

[1] I. Akyildiz, T. Melodia, and K. Chowdhury, "A survey on wireless multimedia sensor networks," *Computer Networks*, vol. 51, no. 4, pp. 921–960, 2007.

[2] H. Fattah and C. Leung, "An overview of scheduling algorithms in wireless multimedia networks," *IEEE Wireless Commun.*, vol. 9, pp. 76–83, Oct. 2002.

[3] Q. Liu, X. Wang, and G. Giannakis, "A cross-layer scheduling algorithm with QoS support in wireless networks," *IEEE Trans. Veh. Technol.*, vol. 55, pp. 839–847, May 2006.

[4] C. So-In, R. Jain, and A.-K. Tamimi, "Scheduling in IEEE 802.16e mobile WiMAX networks: key issues and a survey," *IEEE J. Sel. Areas Commun.*, vol. 27, pp. 156–171, Feb. 2009.

[5] C. So-In, R. Jain, and A.-K. Al-Tamimi, "A scheduler for unsolicited grant service (UGS) in IEEE 802.16e mobile WiMAX networks," *IEEE Systems J.*, vol. 4, pp. 487–494, Dec. 2010.

[6] M. Johansson and L. Xiao, "Cross-layer optimization of wireless networks using nonlinear column generation," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 435–445, Feb. 2006.

[7] T. Ng and W. Yu, "Joint optimization of relay strategies and resource allocations in cooperative cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 25, pp. 328–339, Feb. 2007.

[8] J. Tang and X. Zhang, "Cross-layer resource allocation over wireless relay networks for quality of service provisioning," *IEEE J. Sel. Areas Commun.*, vol. 25, pp. 645–656, May 2007.

[9] P. Liu, R. Berry, and M. Honig, "A fluid analysis of a utility-based wireless scheduling policy," *IEEE Trans. Inf. Theory*, vol. 52, pp. 2872–2889, July 2006.

[10] M. Kobayashi and G. Caire, "An iterative water-filling algorithm for maximum weighted sum-rate of Gaussian MIMO-BC," *IEEE J. Sel. Areas Commun.*, vol. 24, pp. 1640–1646, Aug. 2006.

[11] C. Anton-Haro, P. Svedman, M. Bengtsson, A. Alexiou, and A. Gameiro, "Cross-layer scheduling for multi-user MIMO systems," *IEEE Commun. Mag.*, vol. 44, pp. 39–45, Sep. 2006.

[12] A. Eryilmaz and R. Srikant, "Scheduling with QoS constraints over Rayleigh fading channels," in *Proc. 2004 IEEE Conference on Decision and Control*, pp. 3447–3452.

[13] I.-H. Hou and P. R. Kumar, "Utility-optimal scheduling in time-varying wireless networks with delay constraints," in *Proc. 2010 ACM MobiHoc*, p. 31.

[14] D. O'Neill, A. Goldsmith, and S. Boyd, "Wireless network utility maximization," in *Proc. 2008 IEEE MILCOM*, pp. 1–8.

[15] N. Yang, M. Elkashlan, and J. Yuan, "Impact of opportunistic scheduling on cooperative dual-hop relay networks," *IEEE Trans. Commun.*, vol. 59, pp. 689–694, Mar. 2011.

[16] J. Kim, D. S. Michalopoulos, and R. Schober, "Diversity analysis of multi-user multi-relay networks," *IEEE Trans. Wireless Commun.*, vol. 10, pp. 2380–2389, July 2011.

[17] H. Ding, J. Ge, D. B. da Costa, and Z. Jiang, "A new efficient low-complexity scheme for multi-source multi-relay cooperative networks," *IEEE Trans. Veh. Technol.*, vol. 60, pp. 716–722, Feb. 2011.

[18] M. Sharif and B. Hassibi, "Delay considerations for opportunistic scheduling in broadcast fading channels," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 3353–3363, Sep. 2007.

[19] Q. Liu, S. Zhou, and G. Giannakis, "Queuing with adaptive modulation and coding over wireless links: cross-Layer analysis and design," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 1142–1153, 2005.

[20] D. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes: Volume 1: Elementary Theory and Methods*, 2nd edition. Springer, 2003.

[21] A. Zalesky, "To burst or circuit switch?" *IEEE/ACM Trans. Networking*, vol. 17, pp. 305–318, Feb. 2009.

[22] M. Neely, E. Modiano, and C. Rohrs, "Power and server allocation in a multi-beam satellite with time varying channels," in *Proc. 2002 IEEE INFOCOM*, pp. 1451–1460.

[23] E. Yeh and A. Cohen, "A fundamental cross-layer approach to uplink resource allocation," in *Proc. 2003 IEEE Military Communications Conference.*, pp. 699–704.

[24] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series, and Products*, 7th edition, 2007.

Malcolm Egan (S'10) received the B.E. degree in electrical engineering from the University of Queensland, Australia in 2009. He is currently working towards the Ph.D. degree at the University of Sydney, Australia. His research interests include cross-layer design, frame theory, and structured quantization codebook design.

Phee Lep Yeoh (S'08-M'12) received the B.E. degree with University Medal from the University of Sydney, Australia, in 2004, and the Ph.D. degree from the University of Sydney, Australia, in 2012. From 2004 to 2008, he worked in the wireless engineering division at Telstra Australia as radio network design and optimization engineer. From 2008 to 2012, he was with the Telecommunications Laboratory at the University of Sydney and the Wireless and Networking Technologies Laboratory at the Commonwealth Scientific and Industrial Research Organization (CSIRO), Sydney, Australia. In 2012, he joined the Department of Electrical and Electronic Engineering at the University of Melbourne, Australia. He was a recipient of the University of Sydney Postgraduate Award, the Norman I Price Scholarship, and the CSIRO Postgraduate Scholarship. His research interests include cooperative communications, MIMO, cross-layer optimization, and physical layer security.

Maged Elkashlan (M'06) received the Ph.D. degree in Electrical Engineering from the University of British Columbia, Canada, in 2006. From 2006 to 2007, he was with the Laboratory for Advanced Networking at the University of British Columbia. From 2007 to 2011, he was with the Wireless and Networking Technologies Laboratory at the Commonwealth Scientific and Industrial Research Organization (CSIRO), Australia. He has also held adjunct appointment at the University of Technology Sydney, Australia, between 2008 and 2011. In 2011, he joined the School of Electronic Engineering and Computer Science at Queen Mary, University of London, United Kingdom. He also holds visiting faculty appointments at the University of New South Wales, Australia and Beijing University of Posts and Telecommunications, China. His research interests include distributed wireless networks, cooperative wireless networks, MIMO systems, cognitive radio, and physical layer security. He has served as Technical Program Committee member for several IEEE conferences such as IEEE International Conference on Communications (ICC) and IEEE Global Communications Conference (GLOBECOM).

Iain B. Collings (S'92-M'95-SM'02) received the B.E. degree in Electrical and Electronic Engineering from the University of Melbourne in 1992, and the Ph.D. degree in Systems Engineering from the Australian National University in 1995. Currently he is the Research Director of the Wireless and Networking Technologies Laboratory, and one of the CEOs Science Leaders, at the Australian Commonwealth Scientific and Industrial Research Organisation (CSIRO). Prior to this he was an Associate Professor at the University of Sydney (1999-2005); a Lecturer at the University of Melbourne (1996-1999); and a Research Fellow in the Australian Cooperative Research Centre for Sensor Signal and Information Processing (1995). He has published over 200 research papers in the area of mobile digital communications. More specifically, adaptive multi-carrier modulation, multi-user and MIMO systems, and ad hoc and sensor networks. In 2009 he was awarded the Engineers Australia IREE Neville Thiele Award for outstanding achievements in engineering, and in 2011 he was a recipient of the IEEE CommSoc Stephen O. Rice Best Paper Award for IEEE Transactions on Communications.

Dr. Collings served as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2002-2009), and for the *Elsevier Physical Communication Journal* (PHYCOM) (since 2007). He has served as a Co-Chair of the Technical Program Committee for IEEE International Conference on Communications (ICC) 2013 and IEEE Vehicular Technology Conf. (VTC) Spring 2011; and as a Vice Chair of the TPC for IEEE Wireless Communications and Networking Conference (WCNC) 2010 and IEEE VTC Spring 2006; as well as serving on a number of other TPCs and organizing committees of IEEE conferences. He is a founding organizer of the Australian Communication Theory Workshops 2000–11. He also served as the Chair of the Joint Communications & Signal Processing Chapter in the IEEE NSW Section (2008–2010), as well as serving as the Secretary of the IEEE NSW Section (2010).