

A Fair Radio Resource Allocation Scheme for Ubiquitous High-data-rate Coverage in OFDMA-based Cellular Relay Networks

Mohamed Salem¹, Abdulkareem Adinoyi¹, Halim Yanikomeroglu¹, David Falconer¹, and Young-Doo Kim²

¹Department of Systems and Computer Engineering,
Carleton University, Ottawa, Canada

²Samsung Electronics, SAIT, Korea

Abstract—The provision of very high data rates in a ubiquitous and fair manner throughout the service area is a great challenge for 4G and beyond-4G wireless networks. Towards that end, the deployment of fixed relays by the operators has become an accepted radio access network concept in various standardization activities including LTE-A and 802.16m. In this paper, we present a fair resource allocation scheme that can significantly reduce the co-channel interference and improve spectrum utilization in OFDMA-based multi-cellular networks enhanced by fixed relays. The major goal is to obtain an effective throughput-fairness trade-off through the combination of dynamic resource allocation, in-cell routing, traffic diversity, and history of users' past service. The proposed scheme achieves substantial fairness, throughput, and latency improvements in comparison to the reference schemes as well as to another scheme proposed, earlier by the authors. Numerical results indicate that the proposed scheme provides an almost ubiquitous high-data-rate service to users, irrespective of their locations (including cell-edge users), with a minimal sacrifice in the overall network throughput.

Index Terms— OFDMA, RRM, relaying, throughput, fairness.

I. INTRODUCTION AND MOTIVATION

Fourth-Generation (4G) and beyond-4G wireless networks are expected to provide ubiquitous high-data-rate coverage. A number of standardizations and consortiums are currently exerting efforts towards achieving this objective. These initiatives, as in Long Time Evolution-Advanced (LTE-A), IEEE 802.16m, and WINNER+ [1] have considered Orthogonal Frequency Division Multiple Access (OFDMA) and multihop relaying among the key technologies to deliver the promise of the reliable, high-data-rate, everywhere coverage.

Orthogonal Frequency Division Multiplexing (OFDM) technology is the favored air-interface technology for the next-generation broadband wireless networks. There is a great potential in relay networks for enhancing the multiuser and frequency diversity benefits inherent in OFDMA networks. For instance, routing (relay selection) can exploit the increased spatial diversity to combat shadowing and large pathloss.

Relay deployment provides a cost-efficient way of expanding network size without incurring the backhaul cost associated with deployment of additional full-fledged BSs. However, the additional radio resources (or channels) required for multihop relaying poses a serious cost implication on system bandwidth, especially when a large number of relays will be part of network deployments. Hence, efficient radio resource

management (RRM) schemes with aggressive spectrum reuse capability cannot be overemphasized.

In prospective cellular networks, the applicability of the schemes designed for maximizing the total achievable capacity, or even allocating fair shares of this capacity to users, is doubtful. That is due to the burst traffic that precludes a one-to-one mapping between channel achievable capacity and user's throughput. Hence, such schemes neither deliver a fair service nor can they exploit the traffic diversity. Clearly, the inability to maintain fairness defeats service reliability and ubiquity as it becomes channel and location dependent. On the other hand, users pertaining to the same service class will be charged similarly while the service is not evenly distributed.

Traffic-aware or queue-aware RRM schemes are therefore of great research interest. One way of incorporating queue-awareness into RRM schemes is involving the buffer states in the formulation so that the optimization problem can be worked out as sum-utility or sum-demand maximization. The demand metric can be proportional to both the queue length at the source node and the quality of the link to the destination. In [2] and [3], modulated versions of this metric are used in conventional (non-relaying) cellular SDMA/TDMA and OFDMA networks, respectively. In [4], this demand metric is used in a relay-enhanced single carrier CDMA network.

The centralized joint routing and fair scheduling algorithm proposed in [5] represents a different optimization problem in an OFDMA-based relay-enhanced network. Despite the significant performance returns this joint algorithm has achieved, we observe that it suffers from a performance limiting bottleneck as the traffic load increases. In addition, relays in that scheme transmit and receive different data concurrently on orthogonal OFDM subchannels. This quasi full-duplex relaying (quasi-FDR) raises a practical concern due to the limitations in hardware technology. In this paper, different variants of an efficient half-duplex relaying (HDR) scheme are developed. The contributions and features of the proposed scheme are summarized as follow:

- We propose a radio resource allocation (RRA) scheme with two variants (called Variant-A and Variant-B) that offer an outstanding throughput-fairness trade-off for OFDMA-based relay networks.
- A novel formulation of the RRA is developed for HDR operation. Half-duplex relaying is considered realistic for practical implementations [6].
- We provide low complexity iterative algorithms to solve

Patent filings from this work have been made by Samsung Electronics Co. Ltd., SAIT, Korea. E-mail: {mrashad, adinoyi, halim, ddf}@sce.carleton.ca, yd76.kim@samsung.com

the formulated optimization problem.

- The scheme performs joint in-cell routing and scheduling in contrast to most works in the literature, e.g., [7] [8].
- The network capacity (users/traffic load) has been increased significantly and substantial latency improvement has been achieved as compared to reference [5].
- Traffic diversity and queue-awareness are better exploited as compared to [5].
- Relay fairness is achieved via load balancing among relay stations (RSs) [9] [10].

II. SYSTEM MODEL AND ASSUMPTIONS

In the multi-cellular network, we consider the downlink scenario. Fig. 1 illustrates a representative cell and shows the employed joint scheduling and routing concept. The BS in each cell communicates with its K user terminals (UTs) through the assistance of M RSs. Based on the routing strategy, any UT may communicate simultaneously with multiple nodes (including the BS in a single hop). Therefore, the BS and each of the M relays has K user buffers. All users are subscribed to the same service and thus have the same traffic arrival statistics. Aggressive frequency reuse is adopted which implies that the same spectrum is available in each cell. The available bandwidth is divided into N subchannels. Each subchannel is a set of adjacent OFDM data subcarriers. The schemes operate in time division duplexing (TDD) mode. The downlink frame is partitioned into two consecutive sub-frames, the BS sub-frame followed by the RS sub-frame as shown in Fig. 2. The coherence time of the multipath fading channel is assumed to be greater than the downlink frame duration.

In line with the frame structure, the proposed scheme can be realized in two modes, which we referred to as Variant-A and Variant-B. In the BS sub-frame, only the BS transmits to the selected UTs and RSs. In Variant-A, only the RSs transmit to selected UTs in the RS sub-frame. In Variant-B, however, while RSs transmit to their connected UTs, the BS continues to share the resources with the RSs to transmit to some directly connected UTs. The sub-frame times may not necessarily be of equal length; the sub-frame division is an important optimization parameter. This optimization will not be discussed further in order not to shift the focus of this paper. The equal sub-frame division is adopted to expose the excellent features of the proposed scheme. Adaptive modulation and coding (AMC) is employed. Therefore, on each subchannel the achievable transmit rate for a target bit error rate is a function of the received signal-to-interference-plus-noise ratio (SINR) at the receiving node, and the subchannel bandwidth. Fixed power allocation is considered for both BSs and RSs. While the required CSI at the transmitter can easily be obtained by invoking reciprocity in TDD, feedback is often deemed necessary due to practical concerns. Therefore, to alleviate the amount of feedback overhead, indices of the discrete AMC levels can be reported rather than exact SINR values.

III. MATHEMATICAL FORMULATION OF THE RRA

The scheme aims at maximizing the total cell throughput over two consecutive sub-frames while maintaining fairness among users. This is achieved by iteratively stabilizing the user

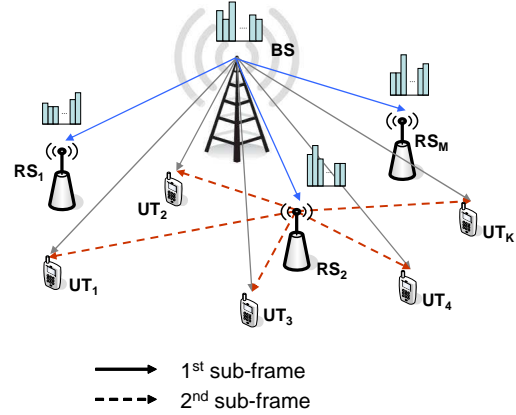


Fig. 1. A representative cell in the multicellular network.

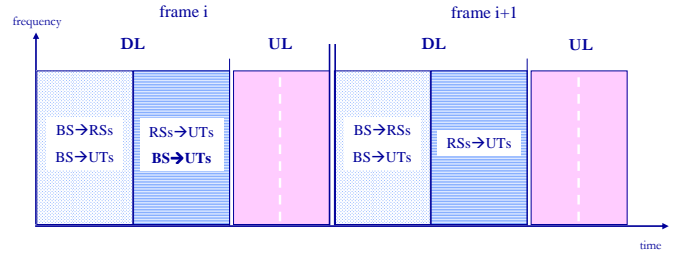


Fig. 2. Frame structure for Variant-B. Note that in the case of Variant-A, only RSs transmit in the second portion of the DL frame.

queues at all nodes via a throughput-optimal policy [4] [11], given the same traffic mean arrival rates. In each allocation instant, two separate optimization procedures are performed (for the two sub-frames) before the BS starts transmitting in the first sub-frame. The BS sends the allocation results for the second sub-frame to the associated RSs on separate control channels. During the uplink portion of the frame, RSs feedback the status of the updated queues due to the previous DL transmission.

A. The Joint Routing and Scheduling for the BS Sub-frame

The joint routing and scheduling optimization problem at the BS can be formulated, for both variants, as a binary integer linear programming (BILP). We note that the demand of any BS-RS link on subchannel n incorporates the maximum difference between the queues at the BS and those at the RS. Therefore, the sum-demand maximization problem is formulated as,

$$\begin{aligned} \max_{\rho^{(1)}, \gamma^{(1)}} & \sum_{n=1}^N \sum_{k=1}^K \rho_{0,k,n}^{(1)} R_{0,k,n} Q_k^0 \\ & + \sum_{n=1}^N \sum_{m=1}^M \gamma_{0,m,n}^{(1)} R_{0,m,n} \max_k \{Q_k^0 - Q_k^m\}^+, \end{aligned} \quad (1)$$

subject to the constraints

$$\begin{aligned} \rho_{0,k,n}^{(1)} & \in \{0, 1\} \forall m, k, n, \\ \gamma_{0,m,n}^{(1)} & \in \{0, 1\} \forall m, n, \end{aligned}$$

$$\sum_{k=1}^K \rho_{0,k,n}^{(1)} + \sum_{m=1}^M \gamma_{0,m,n}^{(1)} \leq 1 \quad \forall n,$$

$$\sum_{n=1}^N \rho_{0,k,n}^{(1)} R_{0,k,n} T + \sum_{n=1}^N \sum_{m=1}^M \gamma_{0,m,n}^{(1)} R_{0,m,n} T \lambda_k^m \leq Q_k^0 \quad \forall k, \quad (2)$$

where

$$\lambda_k^m = \begin{cases} 1, & k = \arg \max_j \{Q_j^0 - Q_j^m\}^+ \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

In (1) - (3), $\rho_{0,k,n}^{(1)}$ is the k^{th} user binary assignment variable to the BS (node 0) on the n^{th} subchannel, during the BS sub-frame. The variable $\gamma_{0,m,n}^{(1)}$ is the m^{th} relay binary assignment variable to the BS node on the n^{th} subchannel ($m = 1, 2, \dots, M$, representing the relays). $R_{m,k,n}$ is the achievable rate on the link between relay m and user k on subchannel n . Q_k^m is the queue length, expressed in bits or bytes, of user k at node m . This queue length might change based on the allocation decisions of the BS sub-frame and will therefore be referred to as q_k^m in the RS sub-frame's allocation. The function $(z)^+$ is defined as $(z)^+ = \max\{0, z\}$. The link achievable rate is calculated as [12]

$$R_{i,j,n} = W \log_2 \left(1 + \frac{-1.5 \beta_{i,j,n}}{\ln(5 \text{ BER})} \right), \quad (4)$$

where $\beta_{i,j,n}$ is the SINR at a receiving node j due to source i on subchannel n . BER and W are the target bit error rate and the bandwidth of the OFDM subchannel, respectively. Whereas T is the transmission time of the sub-frame. The binary indicator λ_k^m is unity if user k has the highest queue difference between the BS and RS_m , and 0 otherwise. The constraints prevent the allocation of a subchannel to more than one link and ensure efficient bit-loading, i.e., resource loading with data bits, to avoid wasting resources.

B. Formulation of Variant-A for the RS Sub-frame

The optimization formulation of Variant-A scheme during the RS sub-frame can be stated as

$$\max_{\rho^{(2)}} \sum_{n=1}^N \sum_{m=1}^M \sum_{k=1}^K \rho_{m,k,n}^{(2)} R_{m,k,n} q_k^m, \quad (5)$$

subject to the constraints

$$\begin{aligned} \rho_{m,k,n}^{(2)} &\in \{0, 1\} \quad \forall m, k, n, \\ \sum_{m=1}^M \sum_{k=1}^K \rho_{m,k,n}^{(2)} &\leq 1 \quad \forall n, \end{aligned} \quad (6)$$

$$\sum_{n=1}^N \sum_{k=1}^K \rho_{m,k,n}^{(2)} \geq \mu_A \quad \forall m \neq 0, \quad (7)$$

$$T \sum_{n=1}^N \rho_{m,k,n}^{(2)} R_{m,k,n} \leq q_k^m \quad \forall m, k, m \neq 0. \quad (8)$$

The binary variable $\rho_{m,k,n}^{(2)}$ assigns subchannel n to UT_k at RS_m during the RS sub-frame where $m = 1, 2, \dots, M$. The parameter $\mu_A = \lfloor N/M \rfloor$ is the minimum number of subchannels that should be assigned to any RS to achieve relay fairness through load balancing.

C. Formulation of Variant-B for the RS Sub-frame

Here, the BS shares the system resource, i.e., subchannels, with the RSs to communicate directly with some selected UTs. Thus, the optimization formulation of Variant-B scheme during the RS sub-frame can be given as

$$\max_{\rho^{(2)}} \sum_{n=1}^N \sum_{m=0}^M \sum_{k=1}^K \rho_{m,k,n}^{(2)} R_{m,k,n} q_k^m, \quad (9)$$

subject to the constraints

$$\begin{aligned} \rho_{m,k,n}^{(2)} &\in \{0, 1\} \quad \forall m, k, n, \\ \sum_{m=0}^M \sum_{k=1}^K \rho_{m,k,n}^{(2)} &\leq 1 \quad \forall n, \end{aligned} \quad (10)$$

$$\sum_{n=1}^N \sum_{k=1}^K \rho_{m,k,n}^{(2)} \geq \mu_B \quad \forall m, \quad (11)$$

$$T \sum_{n=1}^N \rho_{m,k,n}^{(2)} R_{m,k,n} \leq q_k^m \quad \forall m, k. \quad (12)$$

Similarly, $\rho_{m,k,n}^{(2)}$ is the binary assignment variable where $m = 0, 1, 2, \dots, M$ accounts for the BS and the RSs. Whereas $\mu_B = \lfloor N/(M+1) \rfloor$ is the minimum number of subchannels that should be assigned to any node for balancing the load.

IV. REALIZATION OF THE PROPOSED RRM SCHEME THROUGH LOW COMPLEXITY ITERATIVE ALGORITHMS

The BS sub-frame allocation procedure is the same for both Variant-A and Variant-B. The BS has full access to all of the N subchannels but the transmission occupies only a portion of the DL frame duration (see Fig. 2). The demand metric of any BS-UT link on subchannel n is given as

$$D_{n,BS-k} = R_{BS,k,n} Q_k^{BS}. \quad (13)$$

Whereas the demand metric of any BS-RS link on subchannel n is expressed as

$$D_{n,BS-RS_m} = R_{BS,RS_m,n} \max_k \{ (Q_k^{BS} - Q_k^{RS_m})^+ \}. \quad (14)$$

We denote the destination of the best BS link, out of $K+M$ potential links on subchannel n , as \hat{j}_n . The algorithm finds the highest demand across subchannels and the associated BS link \hat{j} is then selected. The algorithm runs another iteration after eliminating the assigned subchannel and updating the associated queues. The iterative process stops when subchannels are exhausted or the queues at the BS are evacuated.

To facilitate understanding of the mechanics of the proposed scheme, we present the pseudo codes of the iterative algorithms for the two variants (Variant-A and Variant-B) where \mathcal{U} , \mathcal{N} , \mathcal{K} , and \mathcal{M} denote the sets; unassigned subchannels, all available subchannels, users, and RSs, respectively. We note that node index is used in the notation of the RS sub-frame code for simplicity.

In Variant-A (as explained in Fig. 2), the BS does not transmit at all and only RSs share the resources to transmit to the selected users during the RS sub-frame. Similar to BS-UT links, the algorithm finds the best link from any relay RS_m out of the K links to UTs on subchannel n .

Optimization code for the BS sub-frame for both variants

```

Initialization:  $\mathcal{U} = \mathcal{N}$ , update  $\mathbf{Q}^{\text{BS}} = [Q_1^{\text{BS}} \dots Q_K^{\text{BS}}]$  by
new arrivals, update affected queues in  $\mathbf{Q}^{\text{RS}m}$  using
feedback.
while  $|\mathcal{U}| \neq 0$  and  $\mathbf{Q}^{\text{BS}} \neq \mathbf{0}$  do
  for each  $n \in \mathcal{U}$ 
    for  $m = 1$  to  $M$ 
       $D_{n,BS-RS_m} = R_{BS,RS_m,n} \max_k \{(Q_k^{\text{BS}} - Q_k^{\text{RS}m})^+\}$ 
       $\kappa^m = \arg \max_k \{Q_k^{\text{BS}} - Q_k^{\text{RS}m}\}$ 
    end for
    for  $k = 1$  to  $K$ 
       $D_{n,BS-k} = R_{BS,k,n} Q_k^{\text{BS}}$ 
    end for
     $D_{n,BS} = \max_j \{D_{n,BS \rightarrow j}\}$ ,  $j \in \mathcal{K} \cup \mathcal{M}$ 
     $\hat{j}_n = \arg \max_j \{D_{n,BS \rightarrow j}\}$ 
  end for
   $\hat{n} = \arg \max_n \{D_{n,BS}\}$ ,  $\mathcal{U} = \mathcal{U} - \{\hat{n}\}$ ,  $\hat{j} = \hat{j}_{\hat{n}}$ 
  if  $\hat{j} \in \mathcal{M}$  then
     $\hat{k} = \kappa^{\hat{j}}$ ,  $b = \min\{Q_{\hat{k}}^{\text{BS}}, \lfloor R_{BS,\hat{j},\hat{n}} T \rfloor\}$ 
     $Q_{\hat{k}}^{\text{BS}} = Q_{\hat{k}}^{\text{BS}} - b$ ,  $Q_{\hat{k}}^{\hat{j}} = Q_{\hat{k}}^{\hat{j}} + b$ 
  else
     $\hat{k} = \hat{j}$ ,  $Q_{\hat{k}}^{\text{BS}} = (Q_{\hat{k}}^{\text{BS}} - \lfloor R_{BS,\hat{k},\hat{n}} T \rfloor)^+$ 
  end if
end while

```

Optimization code for RS sub-frame for Variant-A (B)

```

Initialization:  $\mathcal{U} = \mathcal{N}$ ,  $\mathbf{q}^m = \mathbf{Q}^{\text{RS}m} \forall m \neq 0$ ,  $\mathbf{q}^0 = \mathbf{Q}^{\text{BS}}$ .
while  $|\mathcal{U}| \neq 0$  and  $\sum \mathbf{q}^m \neq \mathbf{0}$  do  $\rightarrow (\text{B} : m = 0, \dots, M)$ 
  for each  $n \in \mathcal{U}$ 
    for  $m = 1$  to  $M$   $\rightarrow (\text{B} : m = 0 \text{ to } M)$ 
       $D_{n,m} = \max_k \{R_{m,k,n} q_k^m\}$ 
       $\kappa_{n,m} = \arg \max_k \{R_{m,k,n} q_k^m\}$ 
    end for
  end for

  %  $\mathbf{D} = [D_{n,m}]$  is the demand matrix.
   $(\hat{\mathbf{n}}, \hat{\mathbf{m}}) \leftarrow \text{Hungarian}(\mathbf{D})$  % Vectors of indices
   $\mathcal{U} = \mathcal{U} - \{\hat{\mathbf{n}}\}$ ,  $N_{\text{assigned}} = \|\hat{\mathbf{n}}\| = \|\hat{\mathbf{m}}\|$ 
  %  $N_{\text{assigned}} \leq \min\{M, |\mathcal{U}|\} \rightarrow (\text{B} : M \rightarrow M+1)$ 

  for  $i = 1$  to  $N_{\text{assigned}}$ 
     $\hat{n} = \hat{\mathbf{n}}(i)$ ,  $\hat{m} = \hat{\mathbf{m}}(i)$ ,  $\hat{k} = \kappa_{\hat{n},\hat{m}}$ 
     $q_{\hat{k}}^{\hat{m}} = (q_{\hat{k}}^{\hat{m}} - \lfloor R_{\hat{m},\hat{k},\hat{n}} T \rfloor)^+$ 
  end for
end while

```

Since only one link will be active per subchannel, we need to find the best demand across all RSs for each subchannel. The algorithm therefore solves a one-to-one optimization problem to maximize the total demand by applying the Hungarian algorithm [13] iteratively to the N -subchannels-by- M -nodes demand matrix $[D_{n,m}]$. After each iteration, queue lengths are updated (not the actual buffers), based on the assignments. The iteration is performed until all the packets in the RSs' queues are scheduled or the subchannels are exhausted.

For Variant-B, the BS shares the resources with the RSs to transmit to its selected users in the RS sub-frame and its demand values are calculated as in (13) using its updated queues \mathbf{q}^0 . Hence, the demand matrix is now N -by- $(M+1)$.

Table I shows the **computational complexity** of Variant-A and Variant-B algorithms proposed in this paper. It can be observed that the complexity upper bounds are polynomial

TABLE I
THE COMPLEXITY OF PROPOSED RRM ALGORITHMS.

Frame	Complexity	
	Variant-A	Variant-B
Sub-frame 1	$\mathcal{O}\left(\frac{N^2}{2}(K+M)\right)$	$\mathcal{O}\left(\frac{N^2}{2}(K+M)\right)$
Sub-frame 2	$\mathcal{O}\left(\frac{N^2(N+M)^2}{4M}\right)$	$\mathcal{O}\left(\frac{N^2(N+M+1)^2}{4(M+1)}\right)$

TABLE II
SIMULATION PARAMETERS

Parameter	Value
BS-BS distance	1 Km
RS distance from BS	$0.65 \times \text{cell radius}$
UT min. close-in distance to BS	35 m
BS Tx. antenna gain	15 dB
RS Tx. antenna gain	10 dB
UT Rx. antenna gain	0 dB
Carrier frequency	2.5 GHz
Total bandwidth	20 MHz
UT mobility	20 Km/hr
BS-RS links max. Doppler spread	4 Hz
Number of channel taps (BS-RS)	8
Number of channel taps otherwise	6
OFDM subcarrier bandwidth	10.9375 KHz
OFDM symbol duration	102.86 μsec
Subchannel width	18 subcarriers
CR-QAM target BER	10^{-3}
Noise power density at Rx. nodes	-174 dBm/Hz
BS total Tx. power	46 dBm
RS total Tx. power	37 dBm

in time. The algorithm incurs less complexity if it stops before allocating all the subchannels. The brute-force solution of the BILP formulation is non-polynomial in time with a complexity of $\mathcal{O}\left((MK)^N\right)$. However, due to the first sub-frame allocation, the proposed algorithms incur a slight increase in complexity as compared to the iterative algorithm of the quasi-FDR reference scheme in [5].

V. NUMERICAL RESULTS

Table II provides the simulation parameters that have been adopted in the study. We have used 19 hexagonal cells with 3 or 6 relays, with equal angular spacing, in each cell. The UTs in each cell are uniformly distributed. The TDD frame length is 2 ms with a 2:1 downlink-to-uplink ratio. Independent Poisson packet arrival process is assumed at BS queues. The average arrival rate is 632 packets (188 bytes each) per second per UT. On top of the 4-dB lognormal shadowing, the BS-RS links experience time-frequency correlated Rician fading with a Rician factor of 10 dB. All the other links are NLOS and experience 8.9 dB independent lognormal shadowing with time-frequency correlated Rayleigh fading. The path-loss model is

$$PL = 38.4 + 10\beta \log_{10}(d) \text{ dB}, \quad (15)$$

where $\beta = 2.35$ for BS-RS links and $\beta = 3.50$ for all other links in the network. The RSs employ an omni-directional antenna to communicate with UTs, while highly directive antennas are used with a gain pattern given as in [5].

Figure 3 shows a scatter plot of user time-average throughput against user distance from the BS for $K = 30$ and $M = 6$. The figure compares the performance of the Variant-A algorithm with the reference quasi-FDR [5] in which BSs and RSs transmit concurrently for the whole DL duration. Each point in the figure represents the time average throughput for a particular user within a drop with fixed location and shadowing. The time average is calculated over the DL frame duration. The scatter plots are approximated using a 3rd degree polynomial curve fitting as a means of averaging out shadowing. The same amount of resources are provided for both schemes, i.e., same DL frame length, bandwidth, and transmit powers. The figure indicates that uniform average throughput across the cell area is achieved by the proposed scheme and thus ubiquity is attained. This is deduced from the flatness and confined spreading of the scatter points of the Variant-A scheme as compared to that of the quasi-FDR which suffers at this loading point ($K = 30$) from a performance limiting bottleneck.

This bottleneck is due to the fact that all nodes in the reference scheme, including the BS, are assigned around the same number of subchannels ($\lfloor N/(M+1) \rfloor$) for the whole DL time. The BS uses this share of resources to transmit to its directly connected UTs as well as forwarding the traffic of relayed UTs to the associated feeder links. This does not limit the performance for light to moderate loadings, whereas at higher loading, this share of resources becomes insufficient to handle the traffic of both direct and relayed users. On the other hand, the proposed scheme allows full access for the BS to the whole bandwidth for the sub-frame duration. As a result, the cell-edge user throughput has been improved by 34%. It is worth mentioning that in Variant-B, more resources are devoted to the BS than in Variant-A due to allocation in the RS sub-frame. However, under the employed routing and scheduling policy, that impedes the distribution of the relayed traffic as the load increases and results in a slightly inferior performance to Variant-A, using the same number of RSs.

Two other conventional (non-relaying) schemes are examined to emphasize the gains achieved by different schemes through exploiting multihop relaying, spatial diversity, queue-awareness, and traffic diversity. The first scheme is the queue-aware ‘RQ’ scheduler, which exploits the traffic diversity by employing the same demand metric at the BS, but this scheduler cannot effectively combat large pathlosses. That applies also to the second scheme, the Max-SINR scheduler, which assigns any subchannel to the best user among only the users with buffered data. With such a (practical) constraint on user buffers, this scheduler enjoys gains from traffic diversity and thus outperforms the well-known conventional Max-SINR.

Figure 4 is provided to complement Fig. 3 and to show the significant improvement of the proposed scheme over the reference schemes in terms of the number of unsatisfied users or the outage behavior. At 30 users/cell, it can be observed that while less than 1.5% of the users have throughput less than 1.0 Mbps in the proposed Variant-A scheme with 6 RSs, an outage probability of 36% is experienced with the Max-SINR and the RQ schedulers whereas the quasi-FDR scheme results in 17.8% outage. A comparison with the Variant-A with 3

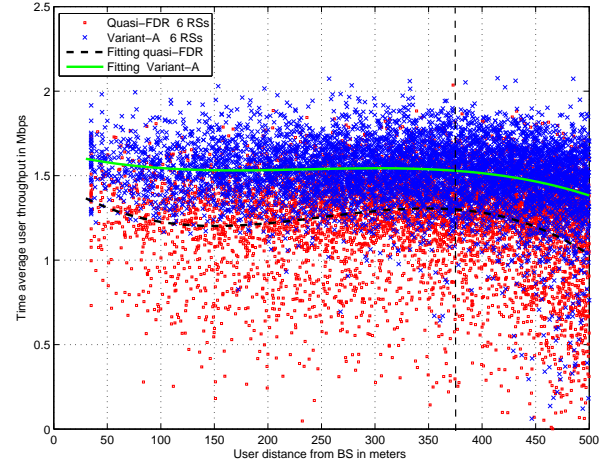


Fig. 3. The throughput comparison of Variant-A with the reference scheme at $K = 30$. Other scatters are not shown for figure clarity.

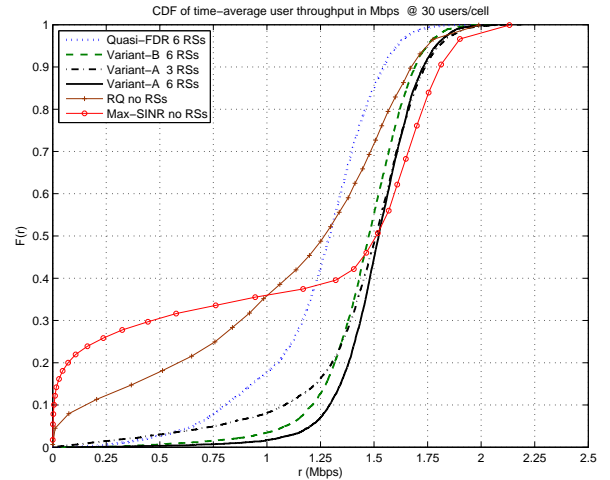


Fig. 4. The CDF of time average user throughput for the proposed scheme and the reference schemes.

RSs, which results in 8.2% outage, shows the impact of the number of RSs on the lower tail behavior and exposes the ability of the proposed scheme to utilize the RSs and exploit the increased spatial diversity. The performance of Variant-B with 6 RSs remains to be bounded between Variant-A with 6 RSs and with 3 RSs and attains an outage of 3.5% which is though 5 times less than the quasi-FDR scheme. Another informative way of reading these results, according to the LTE methodology, is comparing the cell-edge user throughput attained by the schemes at the 5th percentile.

Figure 5 shows the average total cell throughput as a function of the number of users/cell. The figure shows that the performance gap increases significantly as the load increases and becomes insignificant at low to moderate loading levels. The impact of the bottleneck in the quasi-FDR scheme can be realized by comparing the slope of these curves at the high loading end. The proposed scheme outperforms the reference one even with fewer RSs.

Finally, the time-average fairness performance of the proposed scheme is presented in Fig. 6 using the IEEE 802.16m

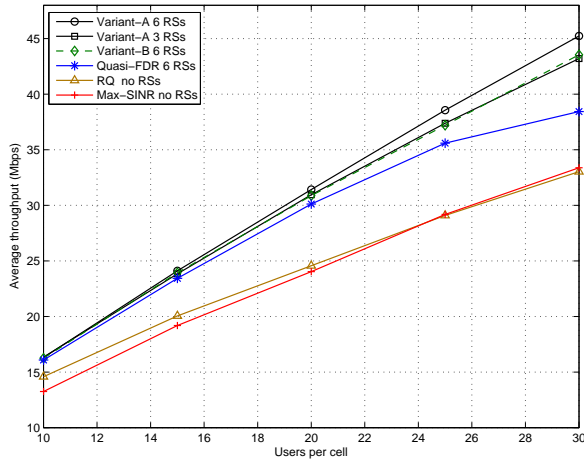


Fig. 5. The total cell average throughput vs. the number of users per cell.

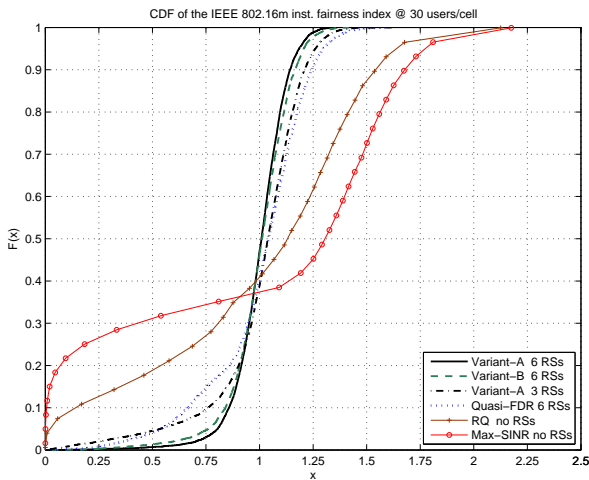


Fig. 6. The IEEE 802.16m fairness index at 30 users per cell.

fairness criteria. A step function at unity in the CDF plot of Fig. 6 indicates absolute fairness. Therefore, the closer the curve is to a step function at unity the more fair the scheme is. It is observed therefore that the proposed scheme achieves the best fair performance as compared to the reference schemes. Further results show that such superior fairness performance is achieved in the more stringent short-term sense. The substantial improvement in both throughput and fairness implies that an efficient throughput-fairness trade-off has been achieved.

Due to the frame structure of the quasi-FDR scheme in [5], the minimum delay of a relayed packet is $T_F + 2\tau$ where T_F and τ are the TDD frame duration and the OFDM transceiver transmission time, respectively. The *latency* is reduced significantly when employing the proposed scheme as the minimum delay for a relayed packet is down to $T + 2\tau$, where T is the BS sub-frame duration. Hence, the lower bound of delay is reduced by 66.6% using equal sub-frames and 2:1 DL-to-UL ratio.

It is worth mentioning that the proposed scheme enjoys better queue-awareness since the feeder link traffic is accounted for when allocating resources to the RSs for the second sub-frame transmission. It is infeasible in the quasi-FDR scheme

to account for feeder link traffic during the same frame due to the concurrent transmission on feeder links and RS-UT links. Furthermore, the proposed algorithms, despite the slight increase in complexity, exploit the finer resource granularity of the proposed frame structure. However, at low to moderate loading levels, the quasi-FDR scheme achieves the same throughput and fairness performance of the proposed scheme with the same *feedback overhead* yet with less computational complexity. Therefore, at such loading levels, the quasi-FDR is more adequate for delay-insensitive applications.

VI. CONCLUSION

This paper presents fairness-aware radio resource management methods that efficiently utilize spectrum in relay-enhanced OFDMA-based multicellular networks. The proposed scheme seeks an effective throughput-fairness trade-off through the combination of dynamic resource allocation, in-cell routing, traffic diversity and queue-awareness to ensure high throughput for all users in fair manner including the cell-edge users. Numerical results show that the proposed scheme offers substantial fairness, throughput and latency improvements to the reference schemes, especially at high loading levels. The proposed scheme also achieves relay fairness in terms of load balancing.

REFERENCES

- [1] WINNER+, <http://projects.celtic-initiative.org/winner+>.
- [2] M. Kobayashi and G. Caire, "Joint beamforming and scheduling for a multi-antenna downlink with imperfect transmitter channel knowledge," *IEEE Journal on Selected Areas in Communications*, 25(7), pp. 1468 - 1477, September 2007.
- [3] P. Parag, S. Bhashyam, and R. Aravind, "A subcarrier allocation algorithm for OFDMA using buffer and channel state information," *IEEE Vehicular Technology Conference (VTC)*, pp. 622 - 625, September 2005.
- [4] H. Viswanathan and S. Mukherjee, "Performance of cellular networks with relays and centralized scheduling," *IEEE Transactions on Wireless Communications*, 4(5), September 2005.
- [5] M. Salem, A. Adinoyi, M. Rahman, H. Yanikomeroglu, D. Falconer, Y.-D. Kim, W. Shin, and E. Kim, "Fairness-aware joint routing and scheduling in OFDMA-based multi-cellular fixed relay networks," *IEEE International Conference on Communications (ICC)*, June 2009.
- [6] W. Nam, W. Chang, S.-Y. Chung, and Y. Lee, "Transmit optimization for relay-based cellular OFDMA systems," *IEEE International Conference on Communications*, pp. 5714 - 5719, June 2007.
- [7] M. Kaneko, P. Popovski, "Radio resource allocation algorithm for relay-aided cellular OFDMA system," *IEEE International Conference on Communications (ICC)*, pp. 4831 - 4836, June 2007.
- [8] Ö. Oyman, "Opportunistic scheduling and spectrum reuse in relay-based cellular OFDMA networks," *IEEE GLOBECOM*, November 2007.
- [9] C. Bae and D.-H. Cho, "Fairness-aware adaptive resource allocation scheme in multihop OFDMA Systems," *IEEE Communications Letters*, 11(2), pp. 134 - 136, February 2007.
- [10] G. Li and H. Liu, "Resource allocation for OFDMA relay networks with fairness constraints," *IEEE Journal on Selected Areas in Communications*, 24(11), pp. 2061 - 2069, November 2006.
- [11] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *IEEE Transactions on Automatic Control*, pp. 1936 - 1948, December 1992.
- [12] X. Qiu and K. Chawla, "On the performance of adaptive modulation in cellular systems," *IEEE Transactions on Communications*, 47(6), pp. 884 - 895, June 1999.
- [13] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval Research Logistic Quarterly*, 2(1), pp. 83 - 97, 1955.