

# Fairness-aware Joint Routing and Scheduling in OFDMA-based Cellular Fixed Relay Networks

Mohamed Salem<sup>1</sup>, Abdulkareem Adinoyi<sup>1</sup>, Mahmudur Rahman<sup>1</sup>, Halim Yanikomeroglu<sup>1</sup>, David Falconer<sup>1</sup>, Young-Doo Kim<sup>2</sup>, Wonjae Shin<sup>2</sup>, and Eungsun Kim<sup>2</sup>

<sup>1</sup>Broadband Communications and Wireless Systems (BCWS) Centre  
Dept. of Systems and Computer Engineering,  
Carleton University, Ottawa, Canada  
<sup>2</sup>Samsung Electronics, SAIT, Korea.

**Abstract**— Relaying and orthogonal frequency division multiple access (OFDMA) are the accepted technologies for emerging wireless communications standards. The activities in many wireless standardization bodies and forums, for example IEEE 802.16 j/m and LTE-Advanced, attest to this fact. The availability or lack thereof of efficient radio resource management (RRM) could make or mar the opportunities in these networks. This paper therefore provides a comprehensive RRM algorithm for OFDMA-based multi-cellular fixed relay networks in a way to ensure fairness among users with minimal impact on the network throughput (in contrast, pure opportunistic RRM techniques always favor users with good channel conditions). Unlike the majority of works in the literature, our proposed scheme is queue-aware and jointly performs routing, fair scheduling, and load balancing among cell nodes. The routing strategy has inherent learning ability and it dynamically converges to better routes.

**Index Terms**— RRM, OFDMA, relaying, routing, scheduling, throughput, fairness, load balancing.

## I. INTRODUCTION AND MOTIVATION

Orthogonal frequency division multiple access (OFDMA) is the envisioned air-interface for 4G wireless networks mainly due to its robustness to frequency selective multipath fading, and its flexibility in radio resource allocation [1]. However, in order to truly realize ubiquitous coverage, the high data rate opportunity in OFDMA schemes has to reach to UTs in the most difficult channel conditions, for example, cell edge UTs. Therefore relaying techniques are considered to be the best option to address this problem since relay stations (RS) - with less functionality than a BS - can forward high data rates to remote areas of the cell, and thus overcome the high path losses, while maintaining low infrastructure cost [2]. Hence, the future network roll-out is expected to include various forms of relays.

The combination of relaying and OFDMA techniques has the potential to provide high data rate to UTs everywhere, anytime. In this work we consider wireless fixed RSs deployed in strategic locations by service providers. To alleviate the heavy burden that will be put on radio resources (in particular spectrum) due to the large number of nodes envisaged for these future networks, mechanisms are needed to enable aggressive

resource reuse. Interference problems associated with dense reuse may result in cell edge UTs occupying a highly disadvantageous position. The conventional opportunistic scheduler will rarely serve UTs in such bad channel conditions; this defeats the notion of ubiquitous coverage targeted in future networks, and exposes the importance of fair algorithms, especially when users subscribed to the same service class are charged similarly regardless of their channel conditions.

In the well-established literature of OFDMA-based conventional cellular networks, several queue/traffic-aware fair scheduling algorithms have been proposed, e.g., [3]. However, such algorithms can not be applied to relay-enhanced networks since the problem is not just a scheduling problem. Rather, in principle, it is a joint routing and scheduling problem. The relay-based RRM algorithms developed for the single-cell model, commonly discussed in the literature, are not suitable for multi-cell scenarios since they are not designed to combat inter-cell interference, e.g., [4] - [6]. Also, an observed tendency in the literature is to maximize the total cell capacity, often under no fairness constraints, e.g., [7] and [8]. Whereas capacity does not map directly to throughput due to the burst traffic, unless user buffers are assumed to be full at all potential serving nodes (which is indeed unrealistic). Resorting to simplified channel models, a transmission selection algorithm is proposed in [9] for instance, where selection and resource allocation are solely based on the number of required subcarriers. However, in order to exploit the multiuser and frequency diversities, an RRM scheme has to be designed to cope with channel variations based on practical channel models.

It is imperative therefore to devise intelligent radio resource allocation and management to be able to harness the potentials in such networks. In [10], a centralized joint scheduling and routing algorithm is proposed for a single-carrier relay network, based on CDMA EV-DO (evolution-data only). In that work, several constraints are imposed which are necessary for operability and reduction of computational complexity but preclude implementation in multi-carrier systems. We propose a novel formulation and a novel low-complexity algorithm that achieves a high degree of fairness and enables intra-cell load balancing in OFDMA-based multi-cell fixed relay networks. The proposed scheme utilizes the opportunities provided in OFDM sub-carriers, channel dynamism, and queue and traffic

This work has been patented by Samsung Electronics Co. Ltd., SAIT, Korea. (e-mail: {mrashad, adinoyi, mmrahman, halim, ddf}@sce.carleton.ca, {yd76.kim, wonjae.shin, eungsun.kim}@samsung.com)

diversities. We show that the scheme provides a reasonable trade-off between network throughput and fairness to all UTs, even to those at the cell edge.

## II. SYSTEM DESCRIPTION AND ASSUMPTIONS

In each cell of the multi-cellular network, the BS, serves  $K$  UTs either directly or through  $M$  RSs. All resources are available in each cell resulting in aggressive resource reuse. The scheme operates in the downlink. The total bandwidth is divided into  $N$  subchannels, each composed of a set of adjacent OFDM data subcarriers. The BS and each of the  $M$  relays are equipped with  $K$  user-buffers. User packets arrive at the corresponding BS buffer according to the chosen traffic model. The BS or a RS needs to assess the quality of its links to potential destinations on all subchannels (i.e., signal-to-interference plus noise ratio (SINR)) in every time frame. The channel fading is assumed to be time-invariant within a frame duration. The potential destinations of the BS are the  $K$  UTs plus the  $M$  RSs. We consider a generic scenario that is not restricted to a specific geographical deployment of RSs. Thus, potentially, any UT can be connected to any combination of RSs. Such ‘open routing’ scenario challenges and exposes the ability of our routing strategy to dynamically select the appropriate route(s) for each UT when a specific geographical relay deployment is adopted.

In contrast to [10], a UT can receive from a group of nodes, and any node can transmit to multiple destinations, simultaneously, on orthogonal subchannels. In addition, any RS is assumed to have the ability to receive and transmit concurrently on orthogonal subchannels. A practical concern might arise if the orthogonal transmit and receive subchannels happen to be close in frequency band. However, since relays are fixed, they can be deployed with two antennas (if necessary, at different elevations); a directional antenna to communicate with the serving BS and an omni-directional antenna to communicate with the UTs, thus, alleviating such concern. We also note that the scheme has the flexibility to split the resource allocation (with almost all its features preserved) into two consecutive processes; that yields the two hops of a RS orthogonal in time.

*Load balancing* is a function usually incorporated with the connection admission control mechanisms in conventional cellular networks. In that context, the load balancing function refers to the hand over of some UTs between adjacent cells to distribute the traffic load among BSs network-wide while maintaining users’ quality of service (QoS). Although the load balancing as defined earlier will be an integral part of any prospective RRM scheme, in the literature of OFDMA-based relay networks, researchers often associate the term ‘load balancing’ with a different function which aims at distributing the load evenly among all the nodes, cell-wide. The number of OFDM subcarriers handled by a node is often employed in literature as a good estimate of the traffic load at that node [4].

## III. THE BS ALGORITHM FOR JOINT ROUTING AND FAIR SCHEDULING

The objective of this algorithm is to maximize the total cell throughput while maintaining fairness among users by stabilizing the queues at all nodes. Relays behave like UTs in

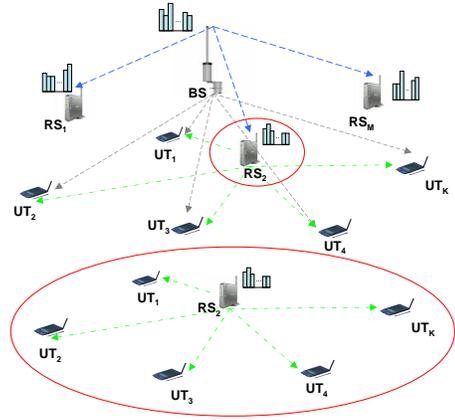


Fig. 1. Example partial network of BS and relays.

uplink and feed back to the BS the per-subchannel channel state information (CSI) of their feeder links as well as the updated queue size information for the previously relayed UTs. Fig. 1 illustrates an example of a cell and captures all the potential links for the BS and RS<sub>2</sub> on one subchannel. Only one link can be active on that subchannel after resource allocation. Let us define the demand metric for any node-UT link on subchannel  $n$  as the product of the achievable rate on that link and the queue length of the user’s buffer at that node, as follows

$$D_{n,m \rightarrow UT_k} = R_{m,k,n} Q_k^m, \quad m = 0, 1, \dots, M, \quad (1)$$

whereas, the demand of any BS-RS link on subchannel  $n$  incorporates the queues at the BS (node 0) and those at the RS and can be expressed as

$$D_{n,BS \rightarrow RS_m} = R_{0,m,n} \max\{(Q_0^0 - Q_k^m)^+\}. \quad (2)$$

The function  $(\cdot)^+$  is a unit slope ramp function, it sets negative arguments to zero.  $Q_k^m$  is the queue length of UT  $k$  at node  $m$  in bits or bytes (shown in blue bars in Fig. 1).  $R_{m,k,n}$  and  $R_{0,m,n}$  are the achievable rates on the links node <sub>$m$</sub> -UT <sub>$k$</sub>  and BS-RS <sub>$m$</sub> , respectively, on subchannel  $n$ . These rates are calculated, without loss of generality, using the following continuous rate formula for adaptive modulation and coding

$$R_{i,j,n} = W \log_2 \left( 1 + \frac{-1.5\beta_{i,j,n}}{\ln(5\text{BER})} \right), \quad (3)$$

where  $\beta_{i,j,n}$  is the received SINR at destination  $j$  from source  $i$  on subchannel  $n$ , considering all dominant interferers. BER and  $W$  are the target bit error rate and the OFDM subchannel bandwidth, respectively.

Although our results show the outstanding performance of the algorithm when employing this metric, it is worth mentioning that designing the structure of the metric is an interesting problem by itself, since different emphasis can be imposed on the arguments. Nevertheless, it can be shown that any monotonically increasing function of the metric defined earlier, in its composite form, will result in the same radio resource allocation (RRA).

### A. The Mathematical Formulation of the RRA at the BS

Through maximizing the sum of the demand of all links in the cell at each allocation instant, the algorithm assigns

the links with the highest capacities at any node to the outstanding queues. Hence, queues are stabilized and no queue is allowed to build up. The resource allocation at the BS can be formulated as a binary integer linear programming (BILP) problem as

$$\max_{\rho, \gamma} \left\{ \sum_{n=1}^N \sum_{m=0}^M \sum_{k=1}^K \rho_{m,k,n} R_{m,k,n} Q_k^m + \sum_{n=1}^N \sum_{m=1}^M \gamma_{0,m,n} R_{0,m,n} \max_k \{ (Q_k^0 - Q_k^m)^+ \} \right\}, \quad (4)$$

$$\text{s.t. } \rho_{m,k,n} \in \{0, 1\} \quad \forall (m, k, n), \quad \gamma_{0,m,n} \in \{0, 1\} \quad \forall (m, n),$$

$$\sum_{m=0}^M \sum_{k=1}^K \rho_{m,k,n} + \sum_{m=1}^M \gamma_{0,m,n} \leq 1 \quad \forall n, \quad (5)$$

$$\sum_{n=1}^N \sum_{k=1}^K \rho_{0,k,n} + \sum_{n=1}^N \sum_{m=1}^M \gamma_{0,m,n} \geq \mu,$$

$$\sum_{n=1}^N \sum_{k=1}^K \rho_{m,k,n} \geq \mu \quad \forall m \neq 0, \quad (6)$$

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

$$T \sum_{n=1}^N \rho_{m,k,n} R_{m,k,n} \leq Q_k^m \quad \forall (m, k), \quad m \neq 0. \quad (7)$$

In the above,  $\rho_{m,k,n}$  is the  $k^{\text{th}}$  UT binary assignment variable to the  $m^{\text{th}}$  node,  $m = 0, 1, 2, \dots, M$ , on the  $n^{\text{th}}$  subchannel ( $m = 0$  corresponds to BS, and the rest correspond to relays). The variable,  $\gamma_{0,m,n}$  is the  $m^{\text{th}}$  relay binary assignment variable to the BS node on the  $n^{\text{th}}$  subchannel. Whereas  $T$  is the transmission time of the downlink frame and  $\mu = \lfloor N/(M+1) \rfloor$  is the minimum number of subchannels to be assigned to any node. The binary indicator  $\kappa_k^m$  takes a value of 1 if user  $k$  has the highest queue difference between the BS and the  $\text{RS}_m$ , and 0 otherwise.

The constraints in (5) ensure that at most one link is active per subchannel (no intra-cell reuse) while in (6) guarantee even distribution of subchannels among all nodes (load balancing). Finally, the constraints (7), unlike the majority of works in the literature, e.g., [4]- [9], ensure efficient bit-loading and prevent scheduling errors which could occur if the total capacity of the links withdrawing from a particular buffer is greater than the queue length at that buffer. Therefore, the novel optimization formulation results in the joint routing and fair scheduling, guarantees efficient use of resources, and balances the load among cell nodes. A discussion on the routing strategy will follow in the next subsection.

**The computational complexity**, however, of such three-dimensional BILP problem is non-polynomial in time and can be approximated to  $O(((M+1)K)^N)$ . As such, the complexity might reach prohibitive limits in a system with high density of UTs and RSs and large number of subchannels. Therefore, we propose a low-complexity iterative algorithm that virtually updates the buffer states between iterations to further improve the fairness while satisfying all of the aforementioned constraints.

## B. The Low-complexity Iterative RRA Algorithm

At the beginning of each allocation time frame, the following steps are executed:

- 1) The demand metric of any  $\text{RS}_m$  on subchannel  $n$  is calculated as the maximum of  $K$  potential links as follows

$$D_{n,m} = \max_k \{ R_{m,k,n} Q_k^m \}, \quad m = 1, 2, \dots, M. \quad (8)$$

Thus,  $D_{n,m}$  is the best proposal of  $\text{RS}_m$  to use subchannel  $n$  while the UT associated with that maximum is marked as the candidate receiver. The demand metric for the BS node is the maximum metric of  $M+K$  potential links and is expressed as

$$D_{n,0} = \max_j \{ D_{n,BS \rightarrow j} \}, \quad (9)$$

where  $D_{n,BS \rightarrow j}$  is calculated using (1) and (2), and  $j$  denotes any of the potential destinations. Thus,  $D_{n,0}$  is the best proposal of the BS to use subchannel  $n$ . The destination associated with that proposal is marked as the candidate receiver. Note that if the destination is an RS, the UT that achieved the highest queue difference on that link is marked as well.

- 2) After calculating the  $(M+1)$  demand metrics on each subchannel, the algorithm solves a one-to-one optimization problem to maximize the total demand by applying the Hungarian algorithm [11] to the  $N \times (M+1)$  demand matrix  $[D_{n,m}]$ .
- 3) The algorithm virtually updates the affected UTs' queues according to the decisions of the previous iteration.

$$Q_k^{m^{(\iota+1)}} = (Q_k^{m^{(\iota)}} - \lfloor R_m^{(\iota)} T \rfloor)^+. \quad (10)$$

In the above,  $Q_k^{m^{(\iota)}}$  is the input queue length to iteration  $\iota$  and  $R_m^{(\iota)}$  is the rate of the link assigned by the Hungarian algorithm to node  $m$  as a result of iteration  $\iota$ . Note that the queues at destination RSs are not incremented between iterations because the transmissions on all subchannels occur simultaneously and the algorithm has to obey the causality law.

- 4) The rows corresponding to the assigned subchannels are eliminated.
- 5) Steps 1-4 are repeated for the unassigned subchannels until all enqueued packets are scheduled or the subchannels are exhausted.

Due to the one-to-one assignment, each iteration will only assign  $M+1$  subchannels to the  $M+1$  nodes. As a result, each node is linked to only one destination per iteration; this prevents, along with step 3, the scheduling errors that could happen when the same queue length is involved in the choice of more than one link as discussed earlier. Furthermore, if  $N \bmod (M+1) = 0$ , each node will be assigned exactly  $N/(M+1)$  subchannels. Hence, load balancing is inherent in the algorithm. The discussion on simulation results will highlight additional advantages of the load balancing feature.

**Routing** of user packets from the BS is dependent on the second-hop link quality as well. Let us assume that  $\text{RS}_M$  in Fig. 1 has a very poor link to  $\text{UT}_3$  (e.g., due to heavy

shadowing) while the BS has forwarded some UT<sub>3</sub> packets to RS<sub>M</sub>. These packets will not be forwarded to UT<sub>3</sub>. However, the algorithm exploits the presence of these trapped packets as they reflect on the quality of the second-hop link. Thus, in the next allocation frame, the BS will identify the user  $k^*$  with the maximum difference  $Q_k^0 - Q_k^M$  as the candidate on that link. Hence, the trapped packets reduce the likelihood of routing UT<sub>3</sub> data again through that link, on any subchannel, while some other user queues at RS<sub>M</sub> are being iteratively discharged. Note that the demand of that BS-RS<sub>M</sub> link has to be compared to the demands of all other BS-RS links as well as to all BS-UT links before it can be considered as the BS's best proposal on a particular subchannel. Thus, the probability of forwarding UT<sub>3</sub> packets to RS<sub>M</sub> converges very rapidly to zero, especially at high density of UTs and RSs. The algorithm therefore, possesses a learning ability. The discussion on simulation results in Section IV will provide a proof of the routing concept.

**The computational complexity** has been significantly reduced using the iterative algorithm since each iteration optimizes a two-dimensional assignment problem using the Hungarian algorithm which has a polynomial complexity of  $O(N^3)$ . This is due to the fact that  $[D_{n,m}]$  is a rectangular matrix with the number of unassigned subchannels usually greater than the number of nodes  $(M + 1)$ . Note that  $M + 1$  rows (subchannels) are eliminated each iteration, whereas the maximum number of iterations required to assign all resources, if necessary, is  $\lceil \frac{N}{M+1} \rceil$ . As such, the complexity of iterations rapidly decreases as the algorithm executes. The complexity of the whole algorithm is therefore  $O(N^4/(M + 1))$ , which is polynomial with a substantial complexity reduction when compared to the BILP problem.

#### IV. SIMULATED NETWORK PERFORMANCE

The simulated network and channel parameters, adopted from the WiMax Forum based on 802.16e, are given in Table I. The pathloss model, RS antenna pattern, and power delay profile of BS-RS links are adopted from the WINNER project [13]. The cellular network consists of 19 hexagonal cells enhanced with 3 or 6 RSs per cell. These RSs are placed at a distance of 0.65 of the cell radius from the BS and with a uniform angular spacing. UTs are uniformly distributed within the cell area. Independent Poisson packet arrival processes are assumed at BS queues. The average arrival rate is 632 packets (188 bytes each) per second per UT. The path-loss model  $PL = 38.4 + A \log_{10}(d)$  dB is used, where  $A = 23.5$  for BS-RS links and  $A = 35.0$  for all other links in the network. Each RS transmits to UTs with an omni-directional antenna, while it receives with a highly directive antenna focused to its BS. The directional antenna gain pattern is given as

$$G(\theta) = -\min[12(\theta/\theta_{3dB})^2, 25] \text{ dB}, \quad -\pi < \theta < \pi, \quad (11)$$

where  $\theta$  is the azimuth arrival angle measured from the antenna bore sight direction. Independent lognormal shadowing is assumed for all links but with different standard deviations. Time-frequency correlated Rician fading is assumed for (LOS) BS-RS links while all other (NLOS) links are assumed to experience Rayleigh fading.

TABLE I  
SIMULATION PARAMETERS

Parameter	Value
BS-BS distance	1 Km
UT min. close-in distance to BS	35 m
BS Tx. antenna gain	15 dB
RS Tx. antenna gain	10 dB
RS Rx. antenna $\theta_{3dB} = 20^\circ$	$\pi/9$
UT Rx. antenna gain	0 dB
Shadowing $\sigma$ for NLOS links	8.9 dB
Shadowing $\sigma$ , for LOS links (BS-RS)	4 dB
Rician K-factor for BS-RS links	10 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	6
Number of channel taps (BS-RS)	8
TDD frame length	2 msec
Downlink : Uplink ratio	2:1
DL Tx. time in OFDM data symbols	11 symbols
OFDM subcarrier bandwidth	10.9375 KHz
OFDM symbol duration	102.86 $\mu$ sec
Subchannel width	18 subcarriers
Total number of subchannels	102
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

Fig. 2 shows a scatter plot of UT average throughput, against UT distance from the BS, over 100 allocation time frames. There are 15 UTs in a cell. Each point in the plot represents the time average throughput for a particular UT within a drop with fixed location and shadowing. The time average is calculated over the downlink frame duration which is 2/3 of the total TDD frame duration. Statistics are collected from 7 cells (the center cell and the surrounding 6 cells) for each of 30 drops. The vertical line represents the RS location. The scatter plots are approximated by a  $3^{rd}$  degree polynomial using least-square curve fitting as a means of averaging over shadowing and path-loss. The uniform average throughput across the cell area is clearly evident from the almost flat performance from the BS to the cell edge. This implies that a fair service and ubiquitous coverage are provided for all users regardless of their location, channel, and interference conditions. It can be observed that the user throughput at the cell edge is comparable to that of users that are close to the BS.

The curve representing the no-relay case (compared with that representing the relay-enhanced case) exposes the significant performance benefits from multihop relaying in terms of coverage, fairness, and throughput. In addition, the no-relay scheme also benefits from the scheduling policy based on the demand metric. This can be realized from the comparison with the Max-SINR scheduler which shows that the scheme without relays tends to sacrifice some throughput gains in the region close to the BS in an effort to provide a fair service to the distant UTs. In contrast, fairness is not a priority for the Max-SINR scheduler which assigns any subchannel to the UT with the highest SINR and non-empty buffer at the BS. Note that since 'non-empty buffer' is a practical condition, the Max-SINR scheduler is forced to serve cell-edge UTs after discharging the best UTs' buffers. As such, the Max-

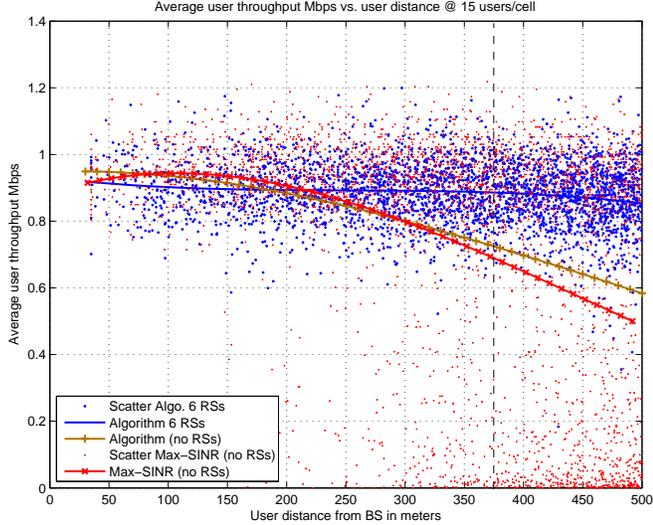


Fig. 2. Average user throughput as function of user location. Scatter points for the algorithm without relays are not shown for figure clarity.

SINR behavior observed at the cell edge outperforms even the benchmark Max-SINR which ignores vacant UT buffers.

Note that, in a conventional cellular network, if a BS-UT link is heavily shadowed then such long-term deep fade will dominate all that UT's subchannels resulting in a very poor average throughput for this user (i.e., a point at the bottom of the scatter plot). Whereas, the scatter points for the algorithm with RSs have high throughput and confined spreading around the mean. This proves the ability of the routing strategy to find the appropriate path(s) for such UTs and to deliver a fair service. Note that the uniform deployment of RSs, given the uniform user distribution, represents a challenge to our 'open routing' strategy since a UT is likely to have poor links to some RSs in the cell.

While fairness can be realized from Fig. 2, we employ Jain's fairness index [12] to further assess the performance of the proposed scheme in terms of short-term (instantaneous) and long-term fairness. Jain's index is defined as  $x(t) = (\sum_{i=1}^K r_i(t))^2 / K \sum_{i=1}^K r_i^2(t)$  which is a positive fraction that takes a value of 1 only if all the  $K$  users in the cell have exactly the same instantaneous rate. The instantaneous values for the index are collected for all frames across all drops to plot a CDF as shown in Fig. 3. Thus, the closer the CDF plot is to a step function at unity, the more fair the network is to the UTs pertaining to the same service class. Note that instantaneous assessment is the most stringent fairness assessment methodology and it is more sensitive to the structure of the index used. However, a substantial fairness advantage is observed for the proposed scheme without relay assistance with respect to the Max-SINR scheduler, while the best fairness behavior is achieved when the proposed scheme employs RSs to provide a fair service for all users. That is evident from the difference between the scenarios with 3 and 6 RSs. As a matter of fact, achieving the stringent instantaneous fairness guarantees long-term fairness. This can be observed from the almost unity index values obtained when Jain's index is applied to the vector of individual user average throughput in

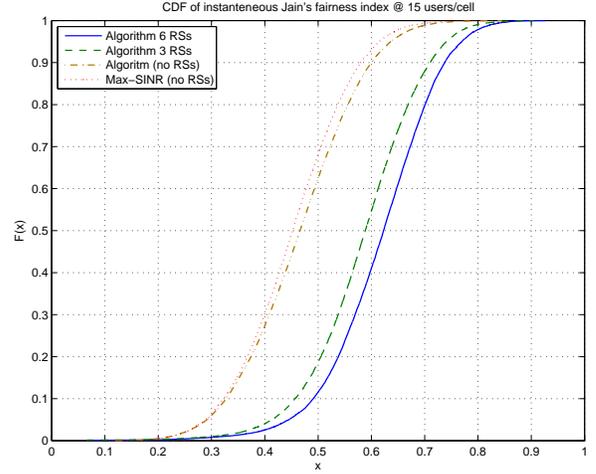


Fig. 3. Jain fairness index for a multi-cell network with 15 users per cell.

a single drop (without averaging over shadowing and pathloss). For instance, the index values; 0.9958, 0.9966, 0.9973, and 0.9968, were realized at 15 users/cell while time averages were computed over a window of 100 frames (0.2 sec).

Fig. 4 shows the average total cell throughput as function of the number of UTs/cell achieved with the proposed scheme with 0, 3 and 6 RSs along with that achieved with the Max-SINR scheduler. The behavior in these curves is in agreement with the multiuser diversity concept and emphasizes the multihop relaying gain compared to the proposed scheme without relays. The comparison between the Max-SINR scheme and our no-relay scheme shows that scheduling based on achievable capacity does not necessarily map to throughput due to the burst traffic, whereas, demand-based scheduling is queue-aware. It is observed that the two schemes interchange roles as the number of users increases. The reason is that the no-relay scheme does not need to sacrifice much of achievable capacity to attain fairness among fewer number of UTs while the Max-SINR scheduler experiences throughput losses in addition to less compensation from the multiuser diversity. It becomes more difficult, at higher number of UTs, to attain the desired fairness what forces the no-relay scheme to sacrifice more capacity and multiuser diversity gains. Whereas the latter provides better compensation to the Max-SINR at that level.

The intra-cell load-balancing behavior of the proposed scheme is shown in Fig. 5 using a normalized histogram. In any iteration, a node is excluded from the assignment, if it has no packets enqueued in all of its buffers. Thus, the very limited perturbations shown in the figure occurred only during the first few initialization frames when RSs were still having no data to send while all resources were assigned to the BS. In addition to distributing the traffic load among cell nodes, the load balancing feature also spatially spreads the co-channel interference across the network. Moreover, given a certain required transmit power per subchannel  $p$ , the load balancing feature guarantees that each RS in the cell will transmit with a total power no greater than  $p \lceil \frac{N}{M+1} \rceil$ , instead of  $p \cdot N$ , which may result in significant cost savings by deploying RSs with less power amplifier rating that is reduced by increasing the number of deployed RSs (power amplifier cost increases

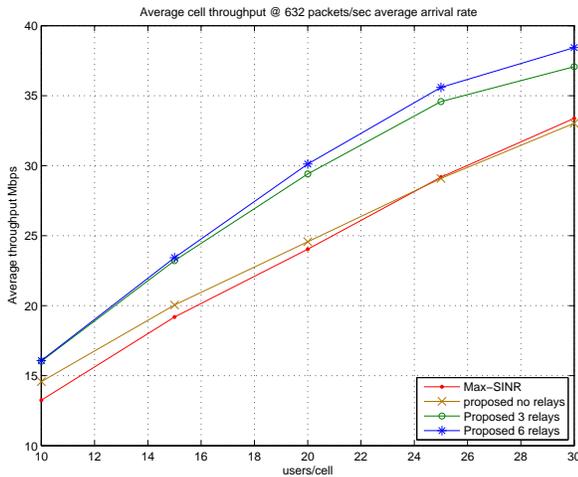


Fig. 4. Average throughput for the proposed scheme.

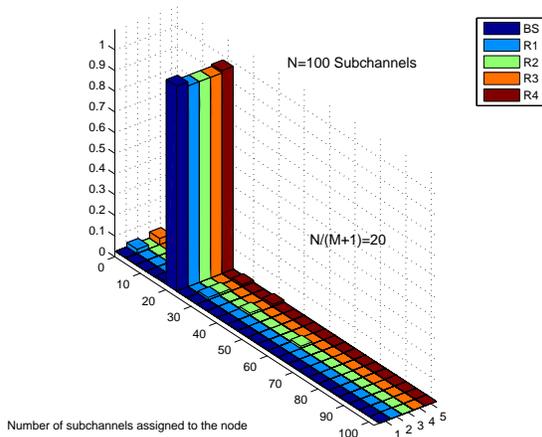


Fig. 5. Normalized histogram of the number of subchannels assigned to transmitting nodes (BS and 4 RSs) during a drop of 100 time frames.

rapidly with its power rating). Therefore, the network operators could employ this feature as a trade-off tool between cost and service requirements in terms of the number of RSs.

#### V. SIGNALLING OVERHEAD AND DELAYS

First we consider the ‘open routing’ strategy investigated earlier in Section IV. In this strategy, we have:

- The required feedback per frame from any UT to the BS is  $N(M+1)$  SINR values or achievable rates. Similarly,  $N$  values are required from an RS to report its feeder link. In addition, RSs report to the BS the status of only the queues that have changed due to the last transmission, at most  $K$  values.
- The minimum delay a relayed packet encounters is  $T_F + 2\tau$  where  $T_F$  and  $\tau$  are the TDD frame duration and the transceiver transmission time, respectively. Although the current simulation platform is quite advanced, individual packet delays are intractable. However, as the results show, the algorithm is designed to maximize throughput while fairness is achieved by stabilizing all queues and avoiding build ups. Thus, the algorithm is expected to minimize the queuing delays as a consequence.

By exploiting the geographical deployment of fixed RSs, the algorithm can circumvent the loss of trapped packets, due to

the generic ‘open routing’, via masking the UT’s buffers at the far RSs located outside its vicinity. Hence, faster convergence is expected and packet loss and/or delays are minimized. Also, substantial savings in overhead can be achieved since no feedback is required for the links from far RSs.

#### VI. CONCLUSION

Efficient RRM algorithms are required to harness the opportunities in the future relay-enhanced OFDMA-based networks in which fairness is crucial. This paper provides a novel fairness-aware joint routing and scheduling algorithm for such networks in multi-cell environments. The algorithm ensures short- as well as long-term fairness among users, including cell-edge users, which is achieved with minimal impact on the network throughput. In contrast, pure opportunistic RRM techniques always favor users with good channel conditions. Our algorithm exploits the opportunities in OFDM subcarriers, channel dynamism, and queue and traffic diversities. Simulation results prove the learning ability and the efficiency of the routing strategy which dynamically converges to better routes, even under the challenging relay deployment examined. The inherent load-balancing feature results in spatial spreading of the co-channel interference and can be used to optimize cost and performance in terms of the number of RSs.

#### REFERENCES

- [1] IEEE P802.16j/D1, “Draft IEEE standard for local and metropolitan area networks Part 16: Air interface for fixed and mobile broadband wireless access systems: Multihop relay specification,” pp. 1002-1007, August 2007.
- [2] R. Pabst, B. Walke, D. Schultz, P. Herhold, H. Yanikomeroglu, S. Mukherjee, H. Viswanathan, M. Lott, W. Zirwas, M. Dohler, H. Aghvami, D.D. Falconer, G.P. Fettweis, “Relay-based deployment concepts for wireless and mobile broadband cellular radio,” *IEEE Communications Magazine*, 42(9), pp. 80-89, September 2004.
- [3] D. Niyato and E. Hossain, “Adaptive fair subcarrier/rate allocation in multirate OFDMA networks: Radio link level queuing performance analysis,” *IEEE Transactions on Vehicular Technology*, 55(6), pp. 1897-1907, November 2006.
- [4] 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.
- [5] R. Kwak and J. M. Cioffi, “Resource-allocation for OFDMA multihop relaying downlink systems,” *IEEE GLOBECOM*, pp. 3225-3229, November 2007.
- [6] M. Kaneko and P. Popovski, “Radio resource allocation algorithm for relay-aided cellular OFDMA system,” *IEEE International Conference on Communications*, pp. 4831-4836, June 2007.
- [7] 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.
- [8] Ö. Oyman, “Opportunistic scheduling and spectrum reuse in relay-based cellular OFDMA networks,” *IEEE GLOBECOM*, pp. 3699-3703, November 2007.
- [9] J. Lee, S. Park, H. Wang, and D. Hong, “QoS-guarantee transmission scheme selection for OFDMA multi-hop cellular networks,” *IEEE International Conference on Communications*, pp. 4587-4591, June 2007.
- [10] H. Viswanathan and S. Mukherjee, “Performance of cellular networks with relays and centralized scheduling,” *IEEE Transactions on Wireless Communications*, 4(5), pp. 2318-2328, September 2005.
- [11] H. W. Kuhn, “The Hungarian method for the assignment problem,” *Naval Research Logistic Quarterly*, 2(1), pp. 83-97, 1955.
- [12] R. Jain, *The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation and Modeling*, New York: Wiley, 1991.
- [13] WINNER, “WINNER II Channel Models,” available online: <https://www.ist-winner.org/WINNER2-Deliverables/D1.1.2.zip>, September 2008.