Downlink Radio Resource Allocation for Multi-cell OFDMA System

Guoqing Li and Hui Liu
Department of Electrical Engineering, Box 352500
University of Washington, Seattle, WA 98195
{lgq, hliu}@ee.washington.edu

Abstract

Orthogonal frequency division multiple access (OFDMA) provides a feasible broadband platform where different forms of multiuser diversities (e.g., mutual interference, selective fading channels and bursty traffics) can be efficiently exploited. While completely centralized resource allocation can take full advantages of these diversities in a multicell environment, the signalling overhead and computational complexity are often prohibitive expensive. This article presents a radio resource control (RRC) scheme for OFDMA systems where dynamic resource allocation is realized at both a radio network controller (RNC) and base stations (BTSs). The scheme is semi-distributed in the sense that the RRC decision is splitted between RNC and BTSs. RNC makes decision on which channel is used by which BTS at super-frame level and BTSs then make decision on which user is assigned to which channel at frame-leve. Two optimizaton problems for RNC and BTSs are formulated and computationally efficient algorithms that perform the function of interference avoidance and traffic/channel adaptation are developed. Numerical analysis is performed under several cell configurations to show tradeoffs between sector interference suppression and dynamic interference avoidance. The results indicate that with reasonable signalling overhead, the protocol and the associated algorithms yield excellent performance for both real-time and non real-time services, even under fast fading.

Index Terms–OFDM, OFDMA, dynamic channel allocation.

I. INTRODUCTION

Orthogonal frequency division multiplexing(OFDM) has emerged as one of the prime modem schemes for broadband wireless networks, e.g., digital video broadcasting (DVB-T), wireless LAN (IEEE 802.11a), and fixed broadband access system (IEEE 802.16a). Its form of modulation that transmits high-speed data via multiple parallel low-rate streams presents an excellent platform for fast implementation [1]. For multi-user applications, one way of applying OFDM is through OFDM-TDMA or OFDM-CDMA [2], where different users are allocated with different time slots or different spreading codes. However, the fact that each user has to transmit its signal over the entire spectrum leads to an averaged-down effect in the presence of deep fading and narrowband interference. Alternatively, one can divide the total bandwidth into traffic channels (one or a cluster of OFDM subcarriers) so that multiple access can be accommodated in an orthogonal frequency division multiple access (OFDMA) fashion. An OFDMA system is defined as one in which each user occupies a subset of traffic channels and each traffic channel is assigned exclusively to one user at any time [3]. An advantage of OFDMA over OFDM-TDMA and OFDM-CDMA is the elimination of intra-cell interference, avoiding CDMA type of multiuser detection (MUD) [4]. A more important feature of OFDMA is its capability of exploiting the so-called network/multiuser diversity by avoiding “null” traffic channels (due to deep fading and narrowband interference) [5]. Since different users perceive different channel qualities, a null traffic channel for one user may still be favorable to other users. Thus through judicious subcarrier allocation, the system can potentially outperform interference-averaging techniques by a factor of 2 to 3 in spectrum efficiency [6]. In addition, it is pointed out in [7] that under different system imperfections, OFDMA is the multiple access scheme that is least sensitive to estimation errors.

Clearly, radio resource allocation (also known as channel loading) plays an important roll in optimizing the performance of OFDMA systems. Despite of the absence of inter-cell interference, optimum channel loading is still an NP-hard problem that is fundamentally difficult to tackle. In practice, additional constraints, e.g. individual users’ rate requirements, further complicate the problem. A Lagrangian relaxation algorithm is introduced in [8] to minimize the total power consumption subject to users’ transmission rate constraints. The algorithm yields significant gain over fixed assignment strategies but requires a large number of iterations to converge. [9] and [10] developed channel loading schemes based on [8] with less computational complexity. However [9] still needs a lot of iterations to converge and the linearization in [10] can not be generalized to all types of modulations. Recently, non-iterative algorithms are proposed in [11] to reduce computation complexity. However up to date, most algorithms on OFDMA resource allocation have been limited to single-cell scenarios.

Dynamic channel allocation (DCA) has been widely studied for multi-cell cellular networks [12]-[14]. However, applying existing DCA algorithms to broadband OFDMA is nontrivial for several reasons. First of all, unlike traditional DCA that assumes a predetermined SINR threshold (for homogeneous applications such as voice), modern data networks utilize adaptive modulation which makes channel assignment decision non-binary from SINR standpoint [15]. People employ different modulation and coding schemes with different SINR [16], thus different throughputs (or achievable rates) are obtained at different SINR levels. Secondly, users' channels are frequency selective and their data rate requirements are also different. The third challenge associated with OFDMA DCA is the measurement and signalling overhead since OFDMA channels are broadband, fully centralized schemes are often too heavy for implementations as all the interference information on all channels has to be gathered at a central controller. These added degree-of-freedoms severely complicate the DCA problem for OFDMA systems. Recently, a distributed multicell approach called PCA is developed in [17] for OFDM systems. However the algorithm hinges upon instantaneous traffic channel establishment which is problematic in practice. In addition, completely distributed schemes may have difficulties dealing with unevenly loaded cells.

In this paper, we investigate an OFDMA RRC scheme where RRC control is realized at both a radio network controller (RNC) and base stations (BTSs). Our focus is on OFDMA downlink with no intra-cell interference, although the results presented can be extended to uplink applications with minor modifications. The new protocol attempts to capture three types of multiuser diversities, namely, mutual interference diversity, traffic diversity, and selective fading channels diversity. In particular, RNC controls a set of BTSs and makes radio resource allocation decisions on a super-frame level to maximize the system throughput and captures mutual interference diversity gain (or interference avoidance gain). This is feasible since the interference-floor effect is semi-stationary and within full control of RNC. The bursty traffic diversity, along with the fading channel diversity (depending on the Doppler frequency), is exploited at the BTS level within shorter frames based on users' changing channel conditions and buffer occupancies. As a result, the channel is always assigned to the user with the highest utility value, which is a function of both channel and traffic conditions.

Another contribution of this paper is a set of computationally efficient allocation algorithms for both RNC and BTSs. The algorithms perform the functions of both interference avoidance and traffic adaptation with linear-complexity with respect to the number of users and channels. The protocol and algorithms are evaluated under different cell configurations with both real-time and non-real-time traffics. The results reveal some important insights on how to tradeoff between sector (using sector-antennas) interference suppression and dynamic interference avoidance. In particular, we show that using the proposed DCA scheme, the highest overall system spectrum efficiency can be achieved under four-sector cells with two alternating frequency bands.

The remainder of the presentation is organized as follows. In Section II, the system model and the protocol are described and the constrained optimization problems are formulated. Section III derives the allocation algorithms for RNC and BTSs. In Section IV, the system performance is analyzed by simulations and the paper is concluded in Section V.

II. PROTOCOL DESIGN AND PROBLEM FORMULATION

Fig. 1 depicts a typical OFDMA/TDM system where the radio resource is partitioned in both frequency and time domains. In particular, the frequency resource is divided into N *traffic channels* (each traffic channel is a cluster of OFDM subcarriers) whereas the time resource is divided into time slots. The smallest resource unit through which data is transported is termed as *traffic bearer*. Depending on the application, one or a collection of traffic bearers can be allocated to a user at a time. A super-frame is constructed by a number of frames and a frame is constructed by a number of slots. We invoke the following assumptions for the remainder of this paper:

- Assumption 1: Each traffic bearer can only be assigned to one user within a given cell, i.e., there is no intra-cell interference [18][19].

- Assumption 2: Neighboring cells may reuse the same traffic bearer.
- Assumption 3: The transmission power on each traffic bearer is fixed, whereas the transmission rate is variable (using adaptive coding/modulation) [19].
- Assumption 4: Only the dominant co-channel signal from neighboring cells is regarded as interference. The rest is treated as background noise.
- Assumption 5: Perfect channel state information at both transmitter and receiver.

Assumption 1 is a natural choice based on [18], which proves that the best performance of OFDM is achieved by assigning each channel only to one user in each cell (elimination of intra-cell interference). It also shows that water-filling power allocation only brings marginal performance improvement over fixed power allocation with adaptive modulation. Similar results have also been reported in [19] which shows that the channel capacity with channel state information at both the transmitter and receiver (i.e., water-filling power allocation) is just marginally larger than that with channel state information only at the receiver (i.e., equal power allocation).

A. Protocol description

We consider an OFDMA system where radio resources are allocated to users based on their channel measures and traffic requirements. Basically, a RNC coordinates a cluster of BTSs, each of which communicates with a set of users and possesses users' channel state information (CSI) as well as traffic statuses (e.g., arrival rates, buffer occupancies). In our case specifically, CSI is defined as a pair of achievable rates which characterizes inter-cell interference and fading channels, see Fig. 2 for illustration. CSI is periodically provided by the users who listen to the broadband beacon signals broadcast from each BTS. The beacon signals from BTSs are used by each user to determine the dominant interference. For example, on a particular traffic channel, the SINR received by the user who is communicating with BTS$_i$ can expressed as $\frac{p_i h_i}{\sum\limits_{j\neq i}^{L} h_j p_j + N_0}$, where L is the number of BTSs in the system and $p_i h_i$ represents the signal strength received by the user from BTS$_i$ with p_i and h_i representing the transmitting power of BTS$_i$ and channel gain between BTS$_i$ and the user, respectively. Using beacon signals, the user can determine the strongest interfering base station, e.g., $k = \underset{j\neq i}{argmax}\ p_j h_j$ and calculate its SINR without this dominant interference as $\frac{p_i h_i}{\sum_{j \neq i, j \neq k} p_j h_j + N_0}$. The detailed technique that determines the SINR with and without the dominant interference is beyond the scope of this paper. The actual mapping between the SINR measures and the achievable rates is a function of fading profiles and the available modulation/coding techniques. Examples of the rate function in terms of SINR can be found in [15], [11] and [16]. By defining such a CSI structure, the channel and interference information needs to indicate from the user to BTS$_i$, instead of feeding back the signal strength and interference received from L BTSs, is reduced to a pair of rates. Such information exchange leads to signaling overhead reduction between users and BTSs which is numerically verified in Section IV. Note that the interference floor can be turned on and off by RNC who decides whether each BTS gets the channel and thus becomes dominant interference to some user in other cells.

To perform global optimization, RNC must have knowledge of all users' CSI and traffic status information at all time, from all cells. In reality, the amount of information needed from BTSs to RNC will be prohibitively large. The semi-distributed scheme described here reduces the overhead and computational load by splitting the decisions between RNC and BTSs. Mechanically, RNC updates all users' CSI from all BTSs every super-frame. Decisions made by RNC include the specific set of traffic channels assigned to each BTS for that super-frame and the recommended user assignment for the traffic channel set. Locally the BTS makes the actual pairing between the traffic bearers and the users. In a specific frame, if the recommended user by RNC has traffic to send, the BTS will agree with RNC's recommendation, otherwise the BTS will makes its own decision based on users' traffic conditions (buffer occupancies) and channel fading levels. The decision algorithm of RNC performs interference avoidance and the decision algorithm by BTS performs channel/traffic adaptation, which will be described in the ensuing sections.

Specifically, RNC will be dedicated to coordinate the mutual interference between cells, reducing the information update rate between RNC and BTSs to a super-frame level. BTSs will make real time decisions on channel assignment at user packet level (frame level). As a result, both the mutual interference diversity and the fading channel/bursty traffic diversity can be efficiently exploited. The efficiency of the semi-distributed scheme will be validated in Section

IV which examines the performance gain relative to the duration of the super-frame.

B. Problem formulation for RNC

As stated, the main objective for RNC is to coordinate the mutual interference. To formulate the channel allocation problem for RNC, consider an OFDMA system with N traffic channels and a network of L BTSs (cells). Let $\mathfrak{M}_l$ denote the user set for BTS$_l$. The number of users in the l^{th} BTS is M_l and the entire network has a total of $M_t = \sum_{l=1}^{L} M_l$ users. Let us first define some notations.

- Rate matrices $\boldsymbol{I}_{M_t \times N} = [I_{mn}]$ and $\boldsymbol{S}_{M_t \times N} = [s_{mn}]$: I_{mn} and s_{mn} represent user m's achievable rates (bits/s/traffic channel) on channel n with and without the dominant interference, respectively. In other words $[I_{mn}, s_{mn}]$ defines user m's CSI for channel n.
- Interference index matrix $\boldsymbol{J}_{M_t \times N} = [J_{mn}]$: J_{mn} represents the index of user m's dominant interfering BTS on channel n, i.e., $J_{mn} \in \{1, 2, ..., L\}$.
- Assignment matrices $\boldsymbol{Y}_{M_t \times N} = [y_{mn}]$ and $\boldsymbol{X}_{L \times N} = [x_{ln}]$: $y_{mn} = 1$ indicates that channel n is assigned to user m and 0 otherwise; $x_{ln} = 1$ indicates that channel n is assigned to BTS$_l$ and 0 otherwise.

Define $\delta_{mn} = s_{mn} - I_{mn}$, then $y_{mn} \cdot \left(s_{mn} - x_{J_{mn}} \delta_{mn}\right)$ represents the m^{th} user's transmission rate on channel n. Thus the total throughput on channel n is: $T_n(\boldsymbol{x}_n, \boldsymbol{y}_n) = \sum_{m=1}^{M_t} y_{mn} \cdot \left(s_{mn} - x_{J_{mn}} \delta_{mn}\right)$, where $\boldsymbol{x}_n$ and $\boldsymbol{y}_n$ are the n^{th} column vectors of $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. The total throughput of the system (bits/s) is then given by:

$$T(\boldsymbol{X}, \boldsymbol{Y}) = \sum_{n=1}^{N} T_n(\boldsymbol{x}_n, \boldsymbol{y}_n) = \sum_{n=1}^{N} \sum_{m=1}^{M_t} y_{mn} \cdot \left(s_{mn} - x_{J_{mn}} \delta_{mn}\right).$$

Since for each BTS, the channel is only used by one user at any time, $x_{ln} = \sum_{m \in \mathfrak{M}_l} y_{mn}$. As a result, $\mathbf{X}$ is uniquely determined by $\mathbf{Y}$ and the total throughput can be re-expressed as:

$$T(\boldsymbol{Y}) = \sum_{n=1}^{N} T_n(\boldsymbol{y}_n) = \sum_{n=1}^{N} \sum_{m=1}^{M_t} y_{mn} \cdot \left(s_{mn} - \sum_{i \in \mathfrak{M}_{J_{mn}}} y_{in} \delta_{mn} \right).$$

The RNC seeks to find the assignment matrix $\mathbf{Y}$ so that the total throughput is maximized. Mathematically, the RNC problem is formulated as:

$$\underset{\mathbf{Y}}{Max}\ T(\boldsymbol{Y}) = \sum_{n=1}^{N}\sum_{m=1}^{M_t} y_{mn} \cdot \left(s_{mn} - \sum_{i \in \mathfrak{M}_{J_{mn}}} y_{in}\delta_{mn} \right)$$

$$subject\ to\ 1) \sum_{m \in \mathfrak{M}_l} y_{mn} \in \{0,1\},\ l = 1,2,...,L; n = 1,2,...N, \tag{1}$$

$$2)\ y_{mn} \in \{0,1\},\ m = 1,2,...,M_t; n = 1,2,...N. \tag{2}$$

The resulting $\mathbf{X}$ is the channel assignment decisions for BTSs. In our protocol, $\mathbf{Y}$ will be sent to BTSs together with $\mathbf{X}$ to indicate the recommended user on each assigned channel.

C. Problem formulation for BTSs

After the set of channels are allocated to a cell, the objective of the BTS is to capture the traffic diversity and partial fading channel diversity within the current super-frame. If the recommended user has traffic to send, then the BTS will obey RNC's decision. However for bursty traffic, the recommended user by RNC may not have traffic to send in each frame. In this case, the BTS will make its own decisions and try to maximize the system throughput based on users' fading channel and traffic conditions (buffer occupancies).

Taking BTS$_1$ for example, assume N_1 channels are assigned to its M_1 users. Since the interference level for its users is pre-determined during each super-frame (i.e., X is fixed for current super-frame), the achievable rate of each user is uniquely given by $\boldsymbol{U}_{M_1 \times N_1} = [u_{mn}]$ with $u_{mn} = \left(s_{mn} - x_{J_{mn}}\delta_{mn}\right)$, indicating user m's achievable rate on channel n. Let $\boldsymbol{Z}_{M_1 \times N_1} = [z_{mn}]$ be channel assignment matrix for BTS$_1$, with $z_{mn} = 1$ indicating that channel n is assigned to user m and 0 otherwise. Then the number of bits user m can transmit in one slot is given by

$$r_m = t_s \sum_{n=1}^{N_1} u_{mn} z_{mn},$$

where t_s is the duration of one slot and is regarded as a constant for the BTS. Again, channel n is assigned to only one user at any time, thus $\sum\limits_{m=1}^{M_1} z_{mn} \in \{0,1\}, n = 1,2,...,N_1$.

In each slot, the BTS has the knowledge of each user's buffer occupancy (number of bits in each user's buffer, which characterizes the user's traffic intensity and requirement) which is expressed as:

$$\boldsymbol{q} = (q_1, q_2, ..., q_{M_1}).$$

Based on traffic condition $\mathbf{q}$ and channel condition $\mathbf{U}$, the BTS seeks to find a channel assignment matrix $\mathbf{Z}$ for each slot in current frame so as to maximize the total throughput. Note that the number of bits user m can send in one slot is $\min\{q_m, r_m\}$. Also note that r_m is a function of $\mathbf{Z}$, so the total throughput in each slot $T^{'}$ becomes a function of $\mathbf{Z}$:

$$T^{'}(\boldsymbol{Z}) = \frac{1}{t_s} \sum_{m=1}^{M1} \min\{q_m, r_m\}.$$

As a result, the BTS problem is formulated as:

$$\begin{aligned} &\underset{\boldsymbol{Z}}{Max}\ T^{'}(\boldsymbol{Z}) = \frac{1}{t_s} \sum_{m=1}^{M1} \min\{q_m, r_m\} \\ &subject\ \ to\ 1) \sum_{m=1}^{M_1} z_{mn} \in \{0,1\}, n = 1, 2, ..., N_1, \\ &\qquad\qquad 2) z_{mn} \in \{0,1\}, n = 1, 2, ..., N_1, m = 1, 2, ...M_1. \end{aligned} \tag{3}$$

III. ALGORITHM DEVELOPMENT

The problems formulated for RNC and BTSs are both nonlinear integer optimization problems and they are both NP-complete [20] without computational efficient algorithms to obtain the optimal solution. Cutting plane and branch-and-bound algorithms have been suggested to deal with certain classes of nonlinear integer programming problems. However there is no guarantee that these algorithms yield good performance over large scale problems (e.g., problems with a few dozens of constraints and variables like RNC and BTS problems) [21]. For these reasons, we derive two suboptimal algorithms in this section for the RNC and BTS problems. The algorithms fall into the category of the "local search" method which has been shown to be very effective in a variety of nonlinear integer optimization problems [21].

A. Fast algorithm for RNC

The RNC algorithm uses a greedy approach and it attempts to assign a traffic channel to the user that has the highest *throughput marginal utility* (TMU) value (denoted as Ω_m in the ensuing algorithm) within certain BTS. TMU is defined as the system throughput improvement by assigning the current channel to the user. More specifically, the channel assignment is progressively performed to provide the most improvement to the system throughput. If none of the users in the BTS has a positive TMU value, then the channel is not assigned to this particular BTS. The algorithm performs channel allocation as follows:

```
RNC algorithm
Inputs: $S, I, J$
Outputs: $X$, $Y$
Initialization: $X$=[ ]; $Y$=[ ];
For $n = 1 : N$ do                    % channel loop
    $\boldsymbol{y}_{Mt\times 1} = (0, 0, ..., 0)^T$;   % initialize the n$^{th}$ column vector of $\boldsymbol{Y}$
    $\boldsymbol{x}_{\mathbf{L\times 1}} = (0, 0, ..., 0)^T$;   % initialize the n$^{th}$ column vector of $\boldsymbol{X}$
    $\boldsymbol{\pi} = sort(\ \frac{size(\Phi_l)/n}{R_l})$;  % sort BTSs based on the pre-set throughput ratio and already assigned channels.
    For $j = 1 : L$  do              % BTS loop
        $l = \boldsymbol{\pi}(j)$; starting from the BTS that is most under-assigned based on the pre-set throughput ratios.
        for m$\in M_l$ do              % user loop
            $\Omega_m = T_n(\boldsymbol{y} + e_m^{M_t}) - T_n(\boldsymbol{y})$;     % calculate the TMU value for user m in BTS$_l$
        end for
        $m^* \leftarrow \arg\max_m \Omega_m$;      % find the user with the highest TMU
        if  $\Omega_{m^*} > 0$      % if assigning channel n to BTS$_l$ can improve the system throughput
            $\boldsymbol{x} = \boldsymbol{x} + e_l^L$;      % assign the channel to BTS$_l$
            $\boldsymbol{y} \leftarrow \boldsymbol{y} + e_{m^*}^{M_t}$;      % indicate the recommended user in BTS$_l$
            $\Phi_l = \Phi_l \cup \{n\}$      % add channel n to channel set of BTS$_l$
        end if
    end for
    $X = [X\ \boldsymbol{x}]$;      % assign x to the n$^{th}$ column of $\boldsymbol{X}$
    $Y = [Y\ \boldsymbol{y}]$;      % assign y to the n$^{th}$ column of $\boldsymbol{Y}$
end for
```

e_m^L is a L-length vector with all 0s except 1 at the mth position and Φ_l is the set of channels assigned to BTS$_l$. $\boldsymbol{\pi}$ is the vector containing the order with which BTSs' TMUs are evaluated. $\boldsymbol{\pi}$

is determined based on the pre-set throughput ratios $\{R_l\}$ for BTSs and the number of channels already assigned to BTSs. To prevent handicapped users farther away from BTSs, it is necessary to set pre-set throughput ratios among the cells (e.g., proportional to the number of active users). As a result, the channel assignment always starts with the BTS that is most under-assigned and ends with the BTS that is most over-assigned.

Essentially, the algorithm assigns the channels one by one. For each channel it calculates Ω_m, the TMU value, for each user. Ω_m represents user m's contribution to system throughput improvement due to the higher utilization of the channel minus the effect of the throughput loss due to the increased mutual interference introduced by this user/BTS.

For each channel, the evaluation is executed in two loops, BTS loop and user loop. Clearly, the order with which BTSs are evaluated has a profound impact on the final output. The first BTS usually gets the "clear" channel whereas the rest can only use it if the TMU value is positive. It is possible for Ω_m to be negative for all the remaining users. This means that due to interference from previous assignments, no more BTS can increase the overall throughput of the network by using this particular channel. In this case, the channel will not be assigned to any additional BTS.

To address the fairness issue, the order of BTS loop is adjusted after each channel assignment. More specifically, the BTS that is most under-assigned ($\arg\min\limits_{l} \frac{size(\Phi_l)/n}{R_l}$) will be picked first and the BTS that is most over-assigned ($\arg\max\limits_{l} \frac{size(\Phi_l)/n}{R_l}$) will be examined last in BTS loop.

B. Fast algorithm for BTSs

Once each BTS receives its channel assignment from RNC within a particular super-frame, it will then make instantaneous decisions on pairing the traffic bearers and users. Note that the recommended user by RNC may not always have traffic to send in each frame. In this case, the BTS seeks to solve (3) for each slot in current frame. Note that once RNC has made decisions on which channel is used by which BTS, the interference from BTSs to users are pre-determined for that super-frame. Therefore re-allocating the channel to a user who has traffic to send, instead of maintaining RNC's recommended user on that channel who has an empty queue, will always improve the throughput for the BTS [1].

The algorithm introduced in [22] is modified to solve (3). Note that if channel n is assigned to user m, the number of bits that can be transmitted using channel n in one slot is $F_m = \min(q_m, t_s \cdot u_{mn})$. F_m is defined as the *data traffic utility* (DTU) value and it represents how much throughput can be put on it by user m in one slot. F_m captures the traffic condition (q_m) as well as the fading channel characteristic (u_{mn}). The BTS algorithm attempts to find the user that has the highest DTU value for each traffic bearer, and makes the most use of each traffic channel. The algorithm is carried out for each slot in the frame and it is described as follows (take BTS_1 as an example):

```
BTS Algorithm
Inputs: $\mathbf{U}$ and $\mathbf{q}$
Output: $\mathbf{Z}$
For $n = 1 : N_1$ do
    $F_m = \min(q_m, t_s \cdot u_{mn})$, $m = 1, 2, ..., M_1$;
    $m^* \leftarrow \underset{m}{argmax}\ F_m$    % find the user that can make the most use of channel n
    $z_{m^*n} = 1$;    % assignment channel n to the user
    $q_{m^*} = q_{m^*} - t_s \cdot u_{m^*n}$.    % re-calculate the user's buffer occupancy for next slot
End For
```

As a result, the channel is assigned to the user not only has good channel condition but also has traffic to send.

It can verified that RNC algorithm has complexity of $O(N \times L \times M_t)$ and BTS algorithm has complexity of $O(N_1 \times M_1)$.

IV. PERFORMANCE ANALYSIS

We study the performance of the RRC scheme and allocation algorithms by simulating a multicell, frequency-division duplexing (5MHz + 5MHz) OFDMA system. Four algorithms are

---

[1] In this case, the RNC algorithm does not main optimal in this frame since the BTS does not pairing channels and users according to RNC's decision. However to avoid back-and-forth information and decision exchange between BTSs and RNC in each slot, we still allow BTSs to make their own decisions.

evaluated: 1) RAND, which randomly allocates traffic channels to users and each traffic channel is reused in all BTSs. 2) RANDBTS, which randomly allocates traffic channels to users and each traffic channel is reused in all BTSs; when an assigned user does not have packets to send, the BTS assigns the channel randomly to another user with traffic. 3) RNC, which only performs RNC interference avoidance algorithm. 4) RNCBTS, which performs both RNC algorithm and BTS algorithm.

The performance gains are quantified under two categories: interference avoidance gain (IA) and traffic diversity (TD) gain. The algorithm implemented by RNC captures IA gain and partial fading diversity gain since no traffic feature is accounted. The IA and partial fading diversity gain exploited by RNC can be quantified by the performance comparison between RNC and RAND. The algorithm implemented by BTS exploits both traffic diversity gain and some fading diversity gain. The TD gain and partial fading diversity gain exploited by BTS can be quantified by the performance comparison between RNC and RNCBTS and between RAND and RANDBTS under bursty traffic configurations.

We consider three types of cell configuration: no-sector (omnidirectional), three-sector and four-sector. A total of 7 cells are simulated for no-sector and three-sector cases, while 9 cells are simulated in four-sector cases as shown in Fig. 3. Each color represents a 5MHz band and is coordinated by a RNC. The 5MHz band is further divided into OFDMA traffic channels. Note that in the four-sector configuration, a traffic channel can be reused in different sectors within the same cell.

A. System parameters and signaling overhead

The basic OFDMA setups are given in Table 1. For each traffic channel we use 6 bits to represent its CSI and 3 bits for the identification of the dominant interfering BTS. The amount of information sent by users is thus (6+3) bits/user/channel$\times 32$ channels$\times 100$ users=28.8 Kbits in uplink. Since the information is updated every super-frame, the uplink signaling rate is 28.8 Kbits/600ms=48 Kbps. In downlink, we use 7 bits to represent the user index for every traffic bearer. The signaling rate is thus 7 bits/channel/slot$\times 32$ channels$\times 7$ slots$/10ms = 156.8$ Kbps. If we assume a modest 1 bit/s/Hz bandwidth efficiency, the real traffic throughput for uplink and downlink are (5.12 Mbps-48 Kbps)$=5.07$ Mbps and (5.12 Mbps-156.8 Kbps)$=4.96$ Mbps, respectively. The overall overhead is quite reasonable as summarized in Table 2.

B. Traffic and channel models

Two types of data services: real-time and non-real-time traffics are studied in simulations. The real-time service is modelled as a constant rate data stream with 100% activity [23]. The non-real-time service is simulated using the packet train model [24]. Pareto distribution is used to generate individual traffics: $Prob(ON/OFF\ period > t) = (\frac{t_{\min}}{t})^{\alpha}$, where the mean ON/OFF duration is $\frac{\alpha}{\alpha-1}t_{\min}$. In all simulations, we set $t_{\min}$=0.2 and $\alpha = 1.7$ [17].

Users are uniformly distributed in each sector. Antenna gains are generated according to [23]. Path loss is calculated based on the model recommended by Vinko et al [25]:

$$L = A + 10(a - b \cdot h_b + c/h_b)log_{10}(d/d_0),$$

where a, b, c and h_b are type 2 model parameters (hilly/light tree density or flat/moderate-to-heavy tree density) and A is the reference path loss at d_0. In all simulations, $d_0 = 1Km.$

Shadow fading with a shadowing variance of 8 dB is assumed:

$$\begin{aligned} &Sh\_new = C(d) * Sh\_old + \sqrt{1 - C^2(d)}N(0, \sigma), \\ &C(d) = e^{-\frac{d}{dr}\ln 2}, \end{aligned}$$

where Sh_old represents the shadowing value of the last calculation, Sh_new is the current shadowing value and $N(0, \sigma)$ stands for a Gaussian variable with mean 0 and variance $\sigma^2 = 8^2 = 64$ dB. $C(d)$ determines the correlation of shadowing values between calculations. d_r is a reference distance which is set to 5m and d is the difference of user's positions between adjacent calculations. For simplicity, users' positions are fixed while d is calculated according to the Doppler spread used in simulations (the default Doppler spread is 10Hz). Fast fading is modelled with the filter-noised model and frequency selectivity is generated using COST207-TU model [26]. The adaptive coding/modulation scheme used is shown in Fig. 4. The achievable rates in CSI is obtained by multiplying the corresponding rate with the traffic channel bandwidth.

C. Interference avoidance gain (IA gain)

We first study IA gain with real-time service. With real-time service, every user has packets to transmit at all the time (therefore no traffic diversity gain).

1) Spectrum efficiency: We define the following terms with respect to the spectrum efficiency:

- LF_1 (intra-cell loading factor): defined as the number of sectors in a cell that can use the same traffic channel simultaneously. In no-sector and three-sector cases, LF_1 is 1 and in four-sector case, LF_1 is 2.
- LF_2 (inter-cell loading factor): defined as the percentage of cells that are simultaneously using the same traffic channel.
- E_c (channel spectrum efficiency): calculated as average throughput per channel per sector/channel bandwidth (bits/s/Hz).
- E_s (system spectrum efficiency): calculated as $E_c \times \mathrm{LF}_1 \times \mathrm{LF}_2$.

Table 3 summarizes the performance in difference sector configurations. The results are obtained by averaging over 5000 super-frames and all traffic channels and users. Using RNC algorithm, the improvements due to IA gain (improvement in E_c) are 100%, 37% and 85% for no-sector, three-sector and four-sector configurations respectively. As expected, RNC algorithm are most efficient in scenarios with heavily inter-cell interference (no-sector case). The gain diminishes in the three-sector case where most interference is pre-suppressed by sector antennas.

With or without interference avoidance, sectorization improves the overall system spectrum efficiency E_s. While E_c is not the highest in four-sector case, E_s favors four-sector setup since it has two sectors using the same frequency. Also note that due to the RNC coordination, over 75% of the traffic channels are used by all cells at all time.

2) System utilization: We define the system utilization as follows:

$$\text{system utilization} = \frac{\text{total throughput of all users and channels}}{\text{system bandwidth}}.$$

Clearly the system utilization is a function of traffic load: higher traffic load tends to yield higher throughput. However as traffic load increases, the system utilization converges to the system spectrum efficiency E_s defined in subsection C.1. Fig. 5 shows the system utilization vs.

traffic load with different sectorizations. Among all configurations, the four-sector setup has the highest system utilization by balancing sectorization gain and the interference avoidance gain.

D. Traffic diversity gain (TD gain)

To measure the traffic diversity gain and the partial channel fading diversity gain exploited by BTS, we now consider non-real-time service where traffics arrive at BTSs in bursty fashions.

1) Throughput vs. dropping probability: We define dropping probability as:

$$\text{dropping probability}=\frac{\text{dropped bits due to buffer overflow}}{\text{total number of bits for transmission}}.$$

In simulation, we set buffer size to 2K bits for each user. Fig. 6 gives average throughput per user vs. dropping probability. RNCBTS gives the highest throughput under the same dropping probability while RAND being the worst. The better performance of RNCBTS is due to the interference avoidance scheme used by RNC algorithm as well as traffic diversity gain obtained by BTS algorithm. Note that the difference between RNCBTS and RNC is much greater than the difference between RANDBTS and RAND. This is due to the fact that the BTS in RNCBTS performs BTS algorithm and it re-assigns channels based on users' traffic and channel conditions, however the BTS in RANDBTS only re-assigns channels randomly. In other words, the gap between RAND and RANDBTS quantifies partial TD gain while the gap between RNCBTS and RNC quantifies TD gain as well as part of the fast fading diversity gain.

E. Channel re-assignment impact

One of the key issues in system design is the rate at which RRC should be performed. Fig. 7 studies the system throughput under different channel variation conditions and different RNC resource re-allocation frequencies. The super-frame length varies from 10 ms to 1s and the channel fading is studied under three Doppler spreads: 10Hz, 50Hz and 100Hz.

As expected, slower channel variation yields higher throughput while more frequent RRC update brings higher throughput. Since RNC algorithm captures IA gain and also partial fading diversity gain, slower varying channels (or faster RRC control) allows RNC to capture more accurate information and make better decisions. The drop in throughput due to longer superframes (or faster varying channels) quantifies fading diversity gain relative to the total multiuser diversity gain. It is seen that under very fast fading (e.g., Doppler spread is 100 Hz), the throughput degrades by 15% compared to relatively slower fading (e.g., Doppler spread is 10 Hz). In reality, only a very small portion of the broadband users will be highly mobile. Therefore it is reasonable to expect a even smaller drop in overall system throughput.

V. CONCLUSION

In this paper, we consider OFDMA networks and propose a semi-distributed radio resource control scheme where RRC control is realized at both RNC and BTSs. To capture multiuser diversities in a multicell broadband environment, RNC coordinates intercell interference between BTSs at super-frame level, where each BTS makes its channel assignment decision on frame level based on users' channel as well as traffic conditions. As a result, the radio traffic bearer is always assigned to the user with the highest utility value. Two suboptimal channel allocations algorithms have been derived for RNC and BTSs, respectively. We have evaluated the performance of the protocol and algorithms under a variety of conditions. The results show that the scheme yields excellent performance for both real-time and non real-time services and the algorithm works well even in very fast fading environments. Furthermore, the algorithms are simple to implement in that both have linear complexity with respect to the number of users and channels.

REFERENCES

[1] Gordon L. Stuber, "Principles of mobile communication," *Kluwer Academic Publisher*, 1994.

[2] H. Rohling and R. Grunheid, "Performance comparison of different multiple access schemes for downlink of an OFDM communication system," *IEEE 47th Vehicular Technology Conference*, vol. 3, May, 1997, pp. 1365 -1369.

[3] Didem Kivanc, Guoqing Li and Hui Liu, "Computationally efficient bandwidth allocation and power control for OFDMA," *IEEE Transaction on Wireless Communications*, vol. 2, Nov. 2003, pp.1150 -1158.

[4] Ruxandra Lupas and Sergio Verdu, "Near-far resistance of multiuser detectors in asynchronous channels," *IEEE Transaction on communications*, vol. 38, no. 4, April, 1990.

[5] D. Tse, "Multiuser diversity in wireless networks," *Wireless communication seminar*, Standford University, April. 16, 2001.

[6] G. J. Pottie, "System design choices in personal communications," *IEEE Personal Communication,* vol. 2, no. 5, pp. 50 -67, Oct. 1995.

[7] Marc Moeneclaey, Mark Van Bladel and Hikmet Sari, "Sensitivity of multiple- access techniques to narrow-band interference," *IEEE Transaction on Communications*, March, 2001.

[8] Cheong Yui Wong, Roger S. Cheng, Khaled Ben Letaief and Ross D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE Journal on Selected Areas in Communications*, vol. 17, Oct. 1999, pp. 1747 -1758.

[9] Slawomir Pietrzyk and Gerard J. M. Janssen, “Multiuser subcarrier allocation for QoS provision in OFDMA systems,” *Proc. VTC 2002-Fall*, vol. 2, 2002, pp. 1077 -1081.

[10] Inhyoung Kim, Hae Leem Lee, Beomsup Kim and Lee, Y. H., “On the use of linear programming for dynamic subchannel and bit allocation in multiuser OFDM ,” *IEEE Global Telecommunications Conference*, vol. 6, pp. 3648 -3652, 2001.

[11] Hujun Yin and Hui Liu, “An efficient multiuser loading algorithm for OFDM-based broadband wireless systems,” *IEEE Global Telecommunications Conference*, vol. 1, 2000, pp. 103 -107.

[12] L. Jorguseki, J. B. Punt and D. Sparreboom, “A comparison of dynamic channel allocation algorithms,” in *Proc. FRAMES workshop*, 1999, pp. 131-136.

[13] F. Furuya and Y. Akaiwa, “Channel segregation, a distributed adaptive channel allocation scheme for mobile communication systems,” *Trans. on IEICE*, vol. E74, June 1991, pp.1531-1537.

[14] I. Katzela and M. Naghshineh, “Channel assignment schemes for cellular mobile telecommunication systems: a comprehensive survey,” *IEEE Personal Communications*, June 1996, pp. 10-30.

[15] Israel Koffman and Vincentzio Roman, “Broadband wireless access solutions based on OFDM access in IEEE 802.16,” *IEEE Communication Magazine, vol. 40,* April 2002, pp. 96-103.

[16] A. Czywjk, ”Adaptive OFDM for Wideband Radio Channels,” *IEEE Global Telecommunications Conference,* London, U.K, vol. 1, pp. 713 -718, Nov. 1996.

[17] Justin Chuang and Nelson Sollenberger, “Beyond 3G: Wideband wireless data access based on OFDM and dynamic packet assignment,” *IEEE Communication Magazine,* vol. 38, July 2000, pp. 78-87.

[18] Jiho Jang and Kwang Bok Lee, “Transmit power adaptation for multiuser OFDM systems,”, *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 2, Feb. 2003, pp. 171-178.

[19] E. Biglieri, J. Proakis and S. Shamai, “Fading channels: Information-theoreric and communications aspects,” *IEEE Trans. on Information Theory*, vol. 44, Oct. 1998, pp. 2619-2692.

[20] B. Korte, J. Vygen and Jens Vygen, “Combinatorial optimization: theory and algorithms,” Second Edition, Springer-Verlag Telos.

[21] Christos H. Papadimitriou and Kenneth Steiglitz, ”Combinatorial optimization: algorithms and complexity,” Prentice-Hall, Inc, 1982.

[22] Guoqing and Hui Liu, “Dynamic resource allocation with buffer constraints in broadband OFDMA networks,” *WCNC’2003*, March, 2003, New Orleans.

[23] ETSI: TR 101 112 V3.2.0 (UMTS 30.03 v3.2.0).

[24] Walter Willinger, Murad S. Taqqu, Robert Sherman and Deniel V. Wilson, “Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level,” *IEEE Transaction on Networking*, vol. 5, no. 1, Feb. 1997, pp. 71 -86.

[25] Vinko Erceg, Larry J. Greenstein, Sony Y. Jjandra, Seth R. Parkoff, Ajay Gupta, Boris Kulic, Arthur A. Julius and Renee Bianchi, “An empirically based path loss model for wireless channels in suburban environments,” *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 17, July, 1999, pp. 1205 - 1211.

[26] COST207 TD (86)51-REV 3 (WG1), “Proposal on channel transfer functions to be used in GSM tests late 1986,” Sept. 1986.

number of subcarriers	512
subcarrier bandwidth	10 KHz
sys. bandwidth	10K*512=5.12 MHz
number of traffic channels	32
number of active users	100 users/sector
super-frame length	600 ms (default)
frame length	10 ms
slots per frame	7

Table 1. OFDMA system parameters.

information type	traffic throughput	signaling overhead	overhead percentage
uplink	5.07 Mbps	48 Kbps	0.94%
downlink	4.96 Mbps	156.8 Kbps	3.06%

Table 2. Signaling overhead.

	no-sector (w/o IA/w. IA)	three-sector (w/o IA/w. IA)	four-sector (w/o IA/w. IA)
LF_1	1	1	2
LF_2	100% / 85%	100% / 93%	100% / 75%
E_c	2.06 / 4.08 (bits/s/Hz)	3.09 / 4.20 (bits/s/Hz)	1.65 /3.05 (bits/s/Hz)
E_s	2.06 / 3.50 (bits/s/Hz)	3.09 / 3.95 (bits/s/Hz)	3.3 /4.5 (bits/s/Hz)

Table 3. Spectrum efficiency in difference sector cases.

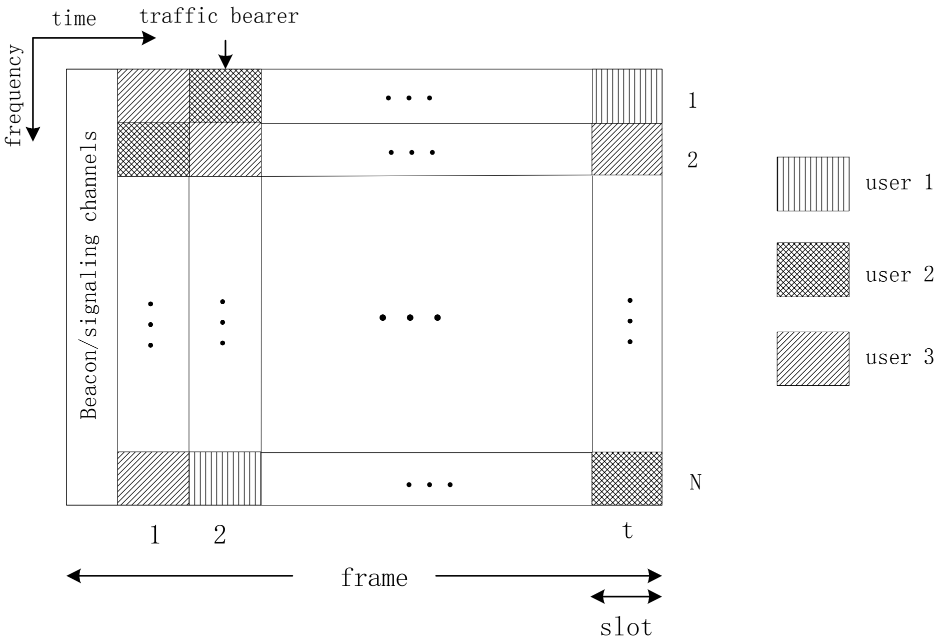


Fig. 1. OFDMA system in frequecy-time axis. In time domain, one frame is devided into t slots. In frequecy domain, the total bandwidth is devided into N traffic channels. A slot within a traffic channel constitutes one traffic bearer.

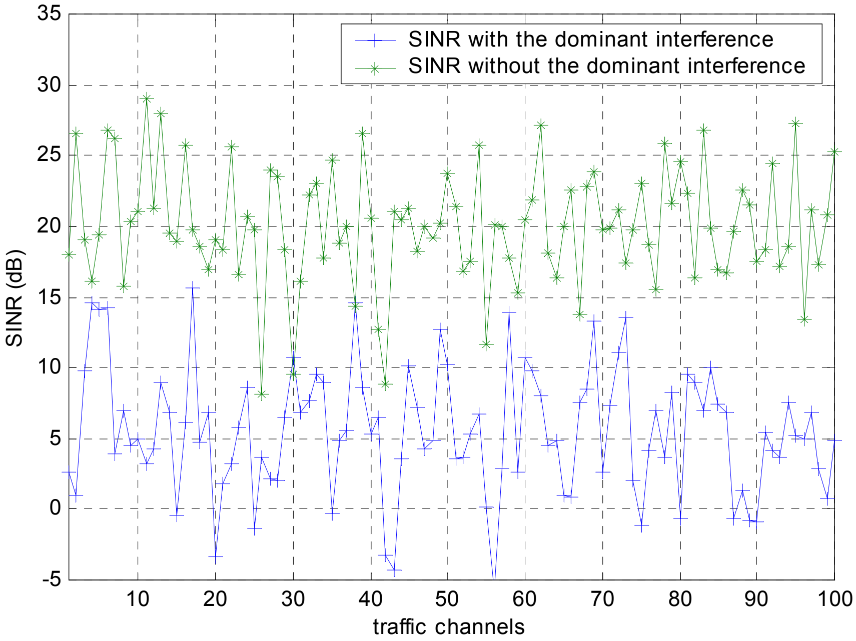
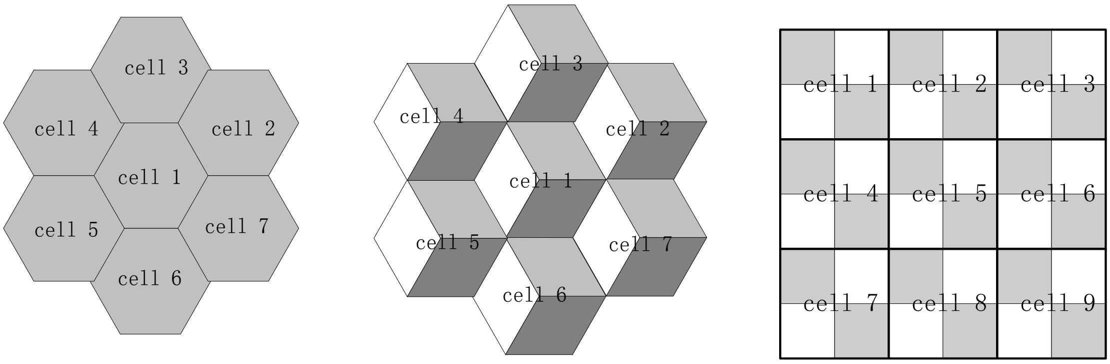


Fig. 2. SINR with and without dominant interference. CSI=[achievable rate with the dominant interference, achievable rate w/o the dominant interference]. Here the rate is measured by bits/s/traffic channel.



(a) no-sector case (b) three-sector case (c) four-sector case

Fig. 3. Sectorizaiton configuration. In three- and four-sector cases, sectors with the same color are using the same frequency band, while in no-sector case, all the cells are sharing the same frequency band.

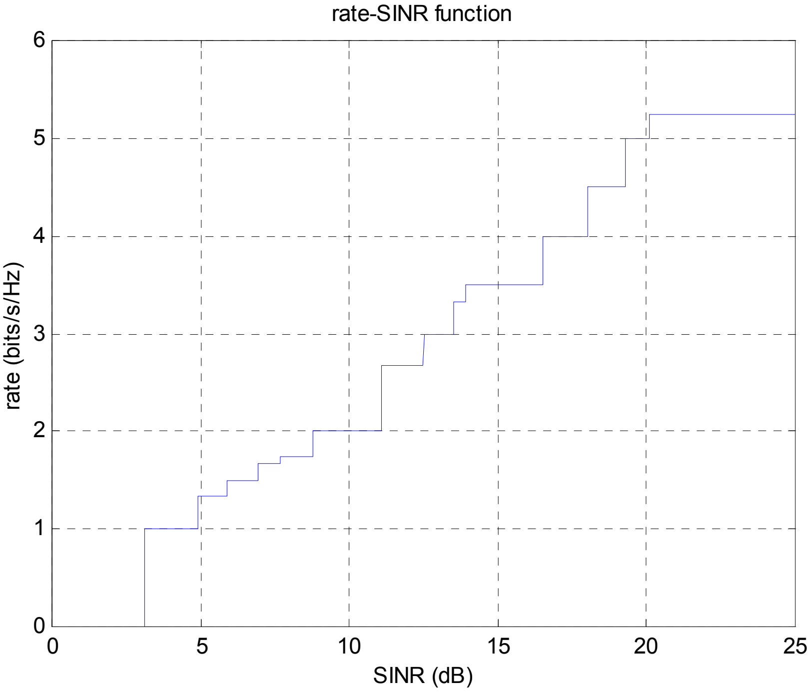


Fig. 4. Adaptive coding/modulation scheme used in simulations. Different rates correspond to different coding/modulation schemes, and they are determined by SINR.

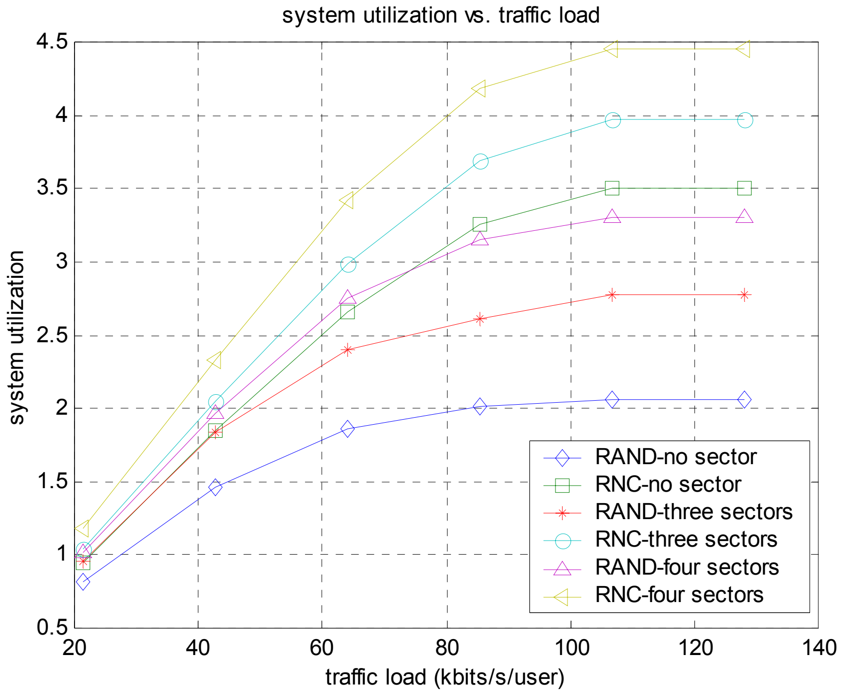


Fig. 5. System utlization vs. traffic load. As traffic load increases, system utlization converges to system spectrum efficiency.

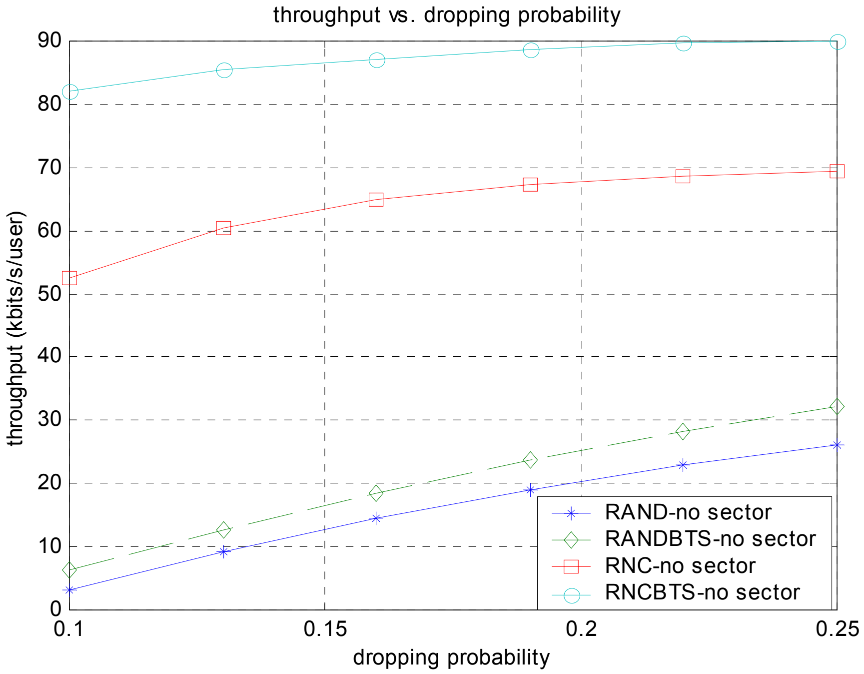


Fig. 6. Study of traffic diversity and fading gain in bursty traffic model. The curves show average throughput per user vs. dropping probability using different algorithms in no-sector case.

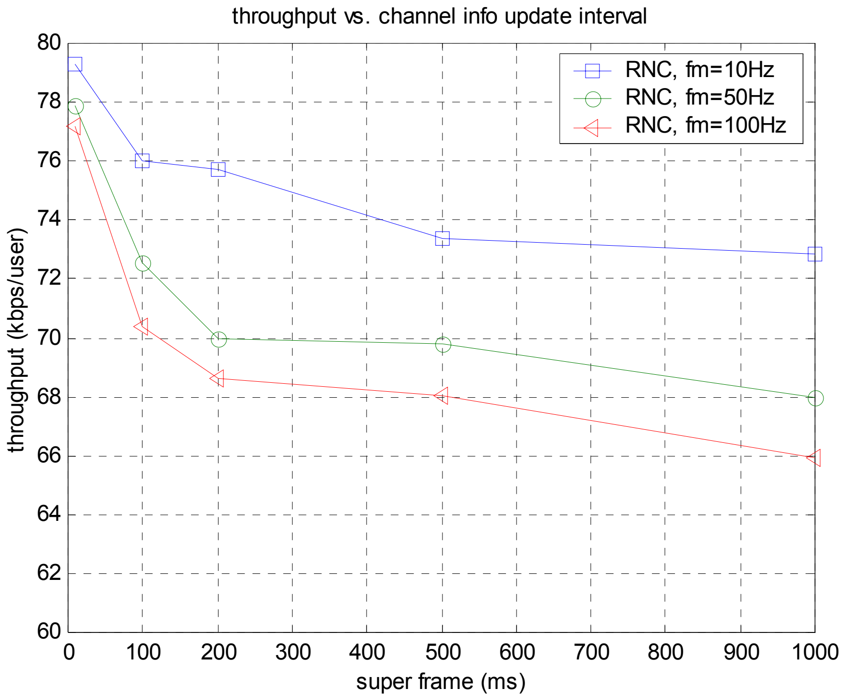


Fig. 7. Average throughput per user vs. channel re-assignment interval under three Doppler spreads: 10Hz, 50Hz and 100Hz. Slower varying channel or faster RRC control allows RNC to capture more accurate information of the system and better performance can be obtained.