Ph.D. DISSERTATION

OPTIMIZATION OF CELLULAR SYSTEM UTILIZATION UNDER REAL PROPAGATION ENVIRONMENTS

실제 전파 환경을 반영한 이동통신 시스템의 최적화 연구

BY

JEONGSIK CHOI

AUGUST 2016

DEPARTMENT OF ELECTRICAL ENGINEERING AND
COMPUTER SCIENCE
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

Abstract

The 4th generation cellular systems, such as LTE (Long-Term Evolution) or LTE-Advanced, significantly improve the speed and the quality of data service as compared to the previous generation systems. In this situation, many applications generating a huge amount of mobile traffic (e.g., high definition (HD) video streaming or cloud-based storage services) have been widely spread. For this reason, the amount of mobile data traffic keeps increasing and sometimes even exceeds the capacity of the system. In order to accommodate explosively increasing mobile data traffic, service providers try to enhance the spatial reuse of wireless resources by deploying more base stations (BSs). Furthermore, small-sized BSs, such as pico and femto BSs, draw much attention as an economical and easy to deploy solution for relieving the load of macro BSs.

In this dissertation, I investigate several strategies for optimizing the utilization of cellular systems. Especially, load balancing algorithms, which forcibly redirect users associated with a congested BS thereby experiencing low service quality to nearby BSs, are proposed. As a first step, I propose methods for predicting the service quality (or equivalently the long-term average throughput) of each individual user when multiple users share the same BS. During developing these algorithms, the time-varying characteristic of wireless channel due to multi-path propagation environment is considered to reflect real propagation environments. To this end, the fluctuation phenomenon of the received signal strength is expressed by a random variable, and then, two types of user throughput estimation schemes are developed. The proposed algorithms can be easily implemented in a practical system, and prediction errors are less than 10% for almost every case.

Based on the proposed throughput estimation methods, I deal with a user association problem in multi-cell environments. At first, a centralized user association algorithm is developed, where a central node collects all the channel information between

every BS and every user and then assigns an optimal base station to each individual user. However, transferring a lot of information to the central node requires excessive uplink feedback and backhaul usage. In addition, such overheads are increased with the density of BSs. For this reason, I propose a decentralized version of user association algorithm, where users themselves choose an optimal BS by considering not only their service quality but also network-wide utilization. The proposed decentralized algorithm especially can be compatible with heterogeneous cellular networks, where there are abundant BSs in the vicinity of each user.

Finally, I study an inter-tier interference management problem between macro and small cell BSs in heterogeneous cellular networks. As the name indicates, small cell BSs are designed to consume much less power as compared to conventional macro BSs. For this reason, users associated with small cell BSs experience severe interference from macro BSs. To mitigate inter-tier interference, the eICIC (enhanced Inter Cell Interference Coordination) method was proposed. In this scheme, macro BSs periodically mute data transmission in order to guarantee the signal quality of users at the small cell BSs. In this dissertation, I try to optimize both user association and inter-tier interference management problems. As a result, users change their association and the system alters data transmission strategies in order to optimize network-wide utilization.

Keywords: Cellular system, heterogeneous network, wireless resource management, user association, interference management

Student number: 2012-30947

Contents

List of Tables

List of Figures

Chapter 1

INTRODUCTION

In practical wireless propagation environments, the signal strength is exponentially decreasing with the distance between the transmitter and the receiver [1–3]. Moreover, an additional attenuation of signal strength occurs when obstacles (e.g., buildings, cars, human bodies) exist in the propagation paths. In order to provide an acceptable signal strength over a wide area and to minimize coverage holes, cellular systems divide the service region into numerous cells and deploy a base station (BS) per cell. According to the Ministry of Science, ICT and Future Planning, the number of LTE (Long-Term Evolution) BSs in South Korea is reported by 120,000-170,000 for each service provider to cover the whole territory [4]. Because of such a huge scale of the system, there are a number of issues, for instance, wireless channel measurement and modeling, physical layer transmission scheme, resource allocation strategy, interference management, etc. Among these topics, I specifically study interference management, user association, and scheduling algorithms in this dissertation.

If two or more neighboring BSs share the same wireless resources, the transmitted signal from a BS degrades the signal quality of users associated with other BSs. If the distance between these BSs becomes too far, the effect of inter-cell interference will be disappear, but users at the boundary region may not have an enough signal quality for communication. For this reason, determining an optimal spacing between BSs is

important when deploying the system. Once the location of BSs are determined, an intelligent interference management method can enhance the utilization of the pre-installed infrastructures. One of traditional ways to mitigate inter-cell interference is the frequency reuse scheme [2]. This scheme divides the whole frequency band into several sub-bands (e.g., three, seven, etc) and assigns each sub-band to cells according to a frequency reuse pattern. Even though this scheme increases the signal strength of users, especially those located in cell boundary regions so that experience severe inter-cell interference, the enhancement of link capacity is questionable. Because, according to Shannon's formula, the capacity is linearly increased with the frequency bandwidth, and therefore, exploiting only a part of frequency resource inevitably decreases the capacity.

In order to increase both cell-edge and overall performances of the system, a variant of the traditional frequency reuse scheme is introduced [5–8]. This scheme is called fractional frequency reuse (FFR) scheme, because it divides the whole frequency band into two parts, where the first part is allocated to every BS whereas the second part is allocated according to the traditional frequency reuse manner. By doing so, users closely located to the BS are generally served by the first part of frequency band, and those located at cell-edge regions are served by the second part of sub-band. Furthermore, few works even try to enhance the utilization of frequency resource, by dynamically changing the FFR pattern according to the traffic load at each BS [9–11].

Another important issue in cellular system is user scheduling. Because the coverage of a BS changes from few hundred meters to several kilometers, each BS should serve many users simultaneously. There have been plenty of strategies for user scheduling [12–16]. For instance, the round-robin scheduling algorithm allocates equal time to users in order to guarantee the fairness among users. In contrast, the max-rate scheduler try to enhance the overall system capacity by serving a user having the best channel quality. In order to balance throughput and fairness among users, Kelly proposed the proportional fair (PF) criterion [13]. The PF scheduling algorithm serves each user

based on the ratio between the average amount of received-data in the past and the instantaneous data rate in the present. Due to its low implementation complexity and remarkable performance gain, various properties have been explored in the literature, including convergence [17], optimality [18, 19], throughput analysis [20–22], etc.

As next step, user association algorithms in multi-cell environments are investigated. As I mentioned earlier, cellular systems are composed of numerous infrastructures. Therefore, selecting an optimal BS is an important issue. One of the easiest ways to choosing a BS is considering the received signal strength, because the capacity of a communication link is determined by signal to interference plus noise ratio (SINR) [23, 24]. However, if too many users share the same BS, each of them may struggle in low service quality. For this reason, cell-site selection based only on a signal strength does not always guarantee the best result, and avoiding a congested BS becomes a desirable option for choosing BS.

There have been plenty of literatures to introduce load-aware cell-site selection rules [25–35]. In an early stage, a cell breathing technique is proposed [25, 26]. This scheme controls the transmission power of each BS so that the coverage of the BSs changes accordingly, and as a consequence, well-balanced loads among neighboring BSs can be achieved. Another approach is considering the number of active users at each BS to complement the traditional signal strength based cell-site selection rule [27–29]. Especially, the work in [29] considers both user association and interference management problems. A distributed association algorithm, where users autonomously choose an optimal BS, is proposed in [31].

Finally, I investigate a joint optimization problem of user association and inter-tier interference management in heterogeneous cellular networks. Also, several remarkable methods that focus on this environment have been proposed. A maximum resource utilization scheme, which solves the optimization problem for channel selection, user association, and power control, is introduced in [36, 37]. Another important approach is the cell range expansion method, which compensates the received signal strength

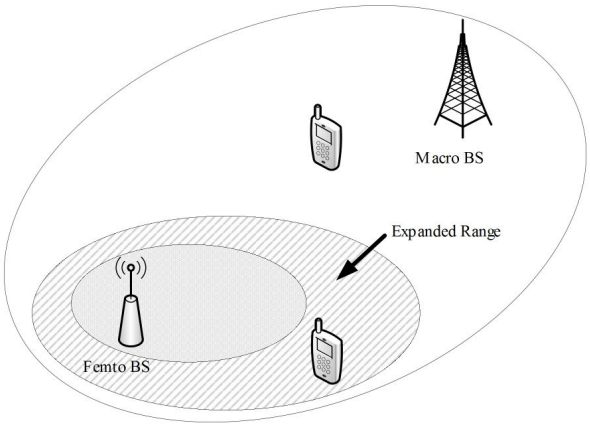


Figure 1.1: An example of cell range expansion scheme in heterogeneous cellular networks.

from a low power BS by virtually adding an offset [38–40]. Using this scheme, users try to be associated with a BS according to the biased signal strength. Figure 1.1 illustrates an example of cell range expansion operation. The study in [38] proposes a range expansion scheme for achieving the maximum sum rate. The bias level for cell selection can be adjusted adequately to guarantee fairness among users [40].

The dissertation is organized as follows. In chapter 2, throughput estimation methods when many users share the same BS is studied. In detail, the proposed methods calculate the receivable throughput of each user for the PF scheduling algorithm under two propagation environments: Rayleigh fading and general conditions. By using these scheme, a centralized user association algorithm for multi-cell cellular environments is introduced in Chapter 3. This algorithm predicts the network-wide utility difference when a user changes association and then redirects the user if it increases the overall system performance. In chapter 4, a decentralized user association algorithm in heterogeneous cellular networks is introduced. A joint optimization of user association and inter-tier interference among different types of cells is analyzed in Chapter 5. Finally, the concussion of the dissertation is represented in Chapter 6.

Chapter 2

USER THROUGHPUT ESTIMATION FOR THE PF SCHEDULING ALGORITHM

2.1 Motivation

The motivation of this chapter comes from the consideration of recent cellular environments, where each user tries to be associated with a BS providing the best received signal strength. However, the association based only on the signal strength results in the loads between neighboring BSs being unbalanced. For this reason, estimating the long-term achievable throughput from each BS and selecting a BS providing the highest long-term throughput are the key factors for improving system performance. For this reason, an accurate user throughput estimation method is required.

In this chapter, I consider two kinds of throughput estimation methods for the PF scheduling algorithm. At first, I develop a throughput estimation scheme for the special case, where a single antenna system is deployed under Rayleigh fading environments. According to [17], the long-term throughput of each user can be obtained by solving the set of ordinary differential equations. Furthermore, the authors also claim that if the data rate function is linear, the PF scheduler yields multi-user diversity gain over

the round-robin scheduler. Such a gain is defined by

$$G(K) = \sum_{q=1}^{K} \frac{1}{q}, \tag{2.1}$$

where K is the number of users sharing the same BS. By expanding this result, I propose a modified throughput estimation equation which is valid and provides an accurate result for non-linear data rate functions.

The first proposed throughput estimation method is designed for the special case, and thus, it may yield a wrong result for other scenarios. For example, in MIMO channels, the statistics of user data rates are widely varying with transmission schemes, receiver types, number of antennas, correlation between antennas, etc. For this reason, I propose another throughput estimation method which covers more general cases. Instead of providing analytical throughput estimation equations for various system configurations, the proposed method monitors the variation of user data rates during a short time period (i.e. a few hundreds of time slots), and then estimates the achievable throughput from the monitored samples.

2.2 System Model

I consider a single cell downlink scenario, where a BS serves the set of users denoted by $\mathcal{K} = \{1, ..., K\}$. The number of antennas is given by n_T for the BS and n_R for each user. Time is slotted, and the BS uses the time multiplexing scheme to serve multiple users. Unlike time resource, the frequency resource is not divided into several parts. For this reason, at each time slot, only a user is selected to use the whole system bandwidth exclusively. For the purpose of determining the modulation and coding scheme (MCS) level and making resource allocation decision at the BS, users periodically measure the channel quality and then report it to the BS via reverse link. The $n_R \times 1$ received signal vector for user k can be expressed by

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x} + \mathbf{n}_k, \tag{2.2}$$

where $\mathbf{H}_k$ is the $n_R \times n_T$ channel matrix from the BS to the user, including pathloss, shadowing and small-scale fading. The $n_T \times 1$ transmitted signal vector $\mathbf{x}$ is restricted by the total signal power P (i.e., $E[\mathbf{x}^H\mathbf{x}] \leq P$), and the $n_R \times 1$ white Gaussian noise vector for the user is characterized by the covariance matrix of $E[\mathbf{n}_k\mathbf{n}_k^H] = N_0\mathbf{I}_{n_R}$. In here, $E[\cdot]$ represents the expectation and $(\cdot)$ implies the hermitian operation. Also, $\mathbf{I}_{n_R}$ denotes the $n_R \times n_R$ identity matrix. Note that every matrix and vector represented in the equation (2.2) becomes a complex scalar variable under the single antenna scenario (i.e., $n_T = n_R = 1$).

According to the Shanon's formula, the capacity between the BS and the user is determined by the SINR of the corresponding link. For instance, in case of single antenna scenarios, the capacity is given by

$$r = W \log_2(1 + \gamma), \tag{2.3}$$

where W is the system bandwidth and γ is the SINR of the link. In case of multiple antenna scenarios, the BS can transmit multiple data streams to the user. Assume that $n_T \leq n_R$ and the BS simultaneously transmits n_T data streams with equal power. Even under the same environment, the capacity can be differed by the type of receiver. In case of the optimal receiver, the instantaneous data rate of user k is theoretically given by

$$r_k = W \log_2 \det\left(\mathbf{I}_{n_R} + \frac{P}{n_T N_0}\mathbf{H}_k\mathbf{H}_k^H\right), \tag{2.4}$$

where $\det(\cdot)$ represents the determinant operator of the matrix.

The linear receivers, such as the zero-forcing (ZF) and the minimum mean squared error (MMSE) receivers, decode each data streams independently. Therefore, the instantaneous data rate of user k is represented by the sum of every data stream's capacity as follows:

$$r_k = W \sum_{m=1}^{n_T} \log_2(1 + \gamma_{k,m}), \tag{2.5}$$

where $\gamma_{k,m}$ means the SINR of the m-th data stream decoded by user k, and it is

expressed by

$$\gamma_{k,m}^{ZF} = \frac{P}{n_T N_0 [\mathbf{H}_k^H \mathbf{H}_k]_{m,m}^{-1}} \tag{2.6}$$

for the ZF receiver, and

$$\gamma_{k,m}^{MMSE} = \frac{1}{[(\mathbf{I}_{n_T} + \frac{P}{n_T N_0} \mathbf{H}_k^H \mathbf{H}_k)^{-1}]_{m,m}} - 1, \tag{2.7}$$

for the MMSE receiver [41].

Once the BS obtains the link quality of users, it runs a scheduling algorithm to serve users according to its own scheduling policy. Then, the average throughput of user k until time slot t is determined by

$$\theta_k(t) = \frac{1}{t} \sum_{\tau=1}^{t} r_k(\tau) I_k(\tau), \tag{2.8}$$

where $r_k(t)$ is the instantaneous data rate of user k at the specific time slot t, and $I_k(t)$ is the scheduling indicator function. In other words, $I_k(t) = 1$ if the BS determines to serve user k at time slot t, and $I_k(t) = 0$ if otherwise. In case of the PF scheduling algorithm, the scheduling indicator function is given by

$$I_k(t) = \begin{cases} 1 & \text{when } k = \arg\max_{k'} \omega_{k'}(t) \\ 0 & \text{otherwise,} \end{cases} \tag{2.9}$$

where $\omega_k(t)$ is called the scheduling preference metric, because the user maximizing this value is selected to be served. The weighted-alpha scheduling rule provides a general expression for this metric, where $\omega_k(t) = r_k(t)/(\theta_k(t-1))^\alpha$ with the parameter α [42]. For special cases, the scheduler becomes the max-rate scheduler if $\alpha = 0$, whereas it works like the round-robin scheduler if $\alpha \to \infty$. If $\alpha = 1$, the scheduler becomes the PF scheduling algorithm.

One of important characteristics of a wireless propagation channel is the small-scale fading phenomenon due to multipaths. As several rays, which experience different paths and therefore has different angle of arrival, are superposed at the received antenna, the signal strength widely fluctuates even for the slight movement of the receiver. To consider such a small-scale fading phenomenon, it is assumed that the time

collection of the instantaneous data rate for user k, $\{r_k(t), t = 1, 2, ...\}$, is modeled by a stochastic process, where each element is independently generated from random variable R_k. Moreover, it is assumed that the statistic of the random variable is stationary during a long period.

2.3 Throughput Estimation for a Single Antenna Scenario under the Rayleigh Fading Environment

According to [17], the average throughput of each user converges to a specific values as time approaches infinity. The convergence point of $\theta_k(t)$ is denoted by θ_k (i.e., $\theta_k(t) \rightarrow \theta_k$ as $t \rightarrow \infty$), and it is called the long-term average throughput of user k. Then, the stationary value θ_k can be represented by

$$\theta_k = \lim_{t\to\infty} E[\theta_k(t)] = \lim_{t\to\infty} E[\frac{1}{t}\sum_{\tau=1}^{t} R_k I_k(\tau)] = \lim_{t\to\infty} \frac{1}{t}\sum_{\tau=1}^{t} E[R_k I_k(\tau)]. \quad (2.10)$$

For a sufficiently large t, the average throughput $\theta_k(t-1)$ approaches to θ_k within an ignorable level. Therefore, the time collection of the scheduling preference metric shown in the equation (2.9), $\{\omega_k(t), t = 1, 2, ...\}$ can be approximated as a stationary process, where each element is independently generated from random variable defined by $\Omega_k = R_k/\theta_k$. Using this relationship, the last expectation operation in the equation (2.10) has the following relationship:

$$\lim_{\tau\to\infty} E[R_k I_k(\tau)] = E[R_k | \Omega_k > \Omega_{k'}, \forall k' \neq k], \quad (2.11)$$

where $E[\cdot|\cdot]$ is the conditional expectation operator.

Theoretically, if a sequence of complex values $\{s_n\}_{n=1}^{\infty}$ converges to a specific value s, then the sequence of their average $\{\frac{1}{n}\sum_{n'=1}^{n} s_{n'}\}_{n=1}^{\infty}$ also converges to the same value [43]. Using this theorem and combining the equation (2.10) and (2.11), the long-term average throughput of user k is rewritten by

$$\theta_k = E[R_k | \Omega_k > \Omega_{k'}, \forall k' \neq k] = \int_0^{\infty} \theta_k \omega f_{\Omega_k}(\omega) \Pi_{k' \neq k} F_{\Omega_{k'}}(\omega) d\omega, \quad (2.12)$$

where $f_\Omega(\cdot)$ and $F_\Omega(\cdot)$ denote the probability density function (PDF) and the cumulative density function (CDF) of random variable Ω, respectively. The equation (2.12) is obtained by taking the expectation of $R_k = \theta_k \omega$ with the conditional probability of $\Omega_k > \Omega'_k, \forall k' \neq k$ which is expressed by $\Pi_{k' \neq k} F_{\Omega_{k'}}(\omega)$, where ω represents the value of random variable Ω_k.

If the instantaneous data rate of a user fluctuates around a high level (e.g., in case that the user is located close to the BS), the user undoubtedly achieves high average throughput, and therefore, the fluctuation level of the scheduling preference metric will be significantly reduced from the instantaneous data rate. On the other hand, if the instantaneous data rate of a user varies around a low level, the scheduling preference metric will not be reduced that much. According to this mechanism, the scheduling preference metric for users fluctuates around similar range, and thus, each user has an almost equal chance to be served. Figure 3.1 illustrates the CDF of the preference metric for 5 users under the single antenna scenario. Regardless of the mean rate of users, the preference metric varies around the similar interval.

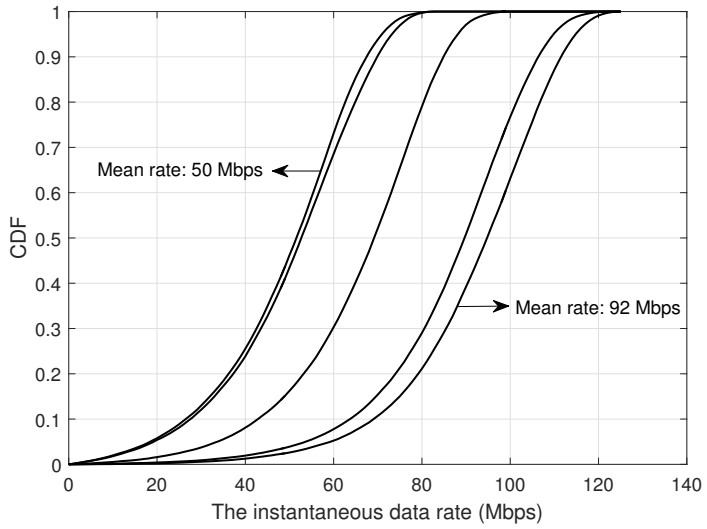
Based on the above discussion, it is assumed that the statistic of the preference metric of users is approximately identical as $F_{\Omega_{k'}} \approx F_{\Omega_k}, \forall k' \neq k$ with some marginal error. Then, the equation (2.12) can be expressed by

$$\theta_k = \int_0^\infty \theta_k \omega f_{\Omega_k}(\omega) \left(F_{\Omega_k}(\omega)\right)^{K-1} d\omega. \tag{2.13}$$

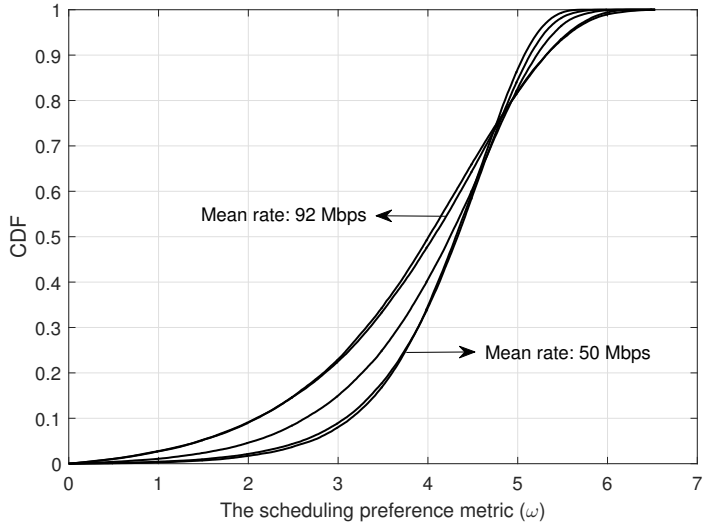
Under the Rayleigh fading environment, the SINR distribution of a specific link is characterized by the exponential distribution. Therefore, random variable Ω_k is represented by

$$\Omega_k = \frac{H(X\gamma_k)}{\theta_k}, \tag{2.14}$$

where γ_k implies the mean SINR of user k, and X is exponential random variable with mean of 1 (i.e., $f_X(x) = e^{-x}$ and $F_X(x) = 1 - e^{-x}$). $H(\cdot)$ implies the capacity function, which has increasing, concave, and differentiable properties. If Shannon's formula is used, this function is expressed by the equation (2.3). Using the relationship



(a)



(b)

Figure 2.1: Statistics of the instantaneous data rate and the scheduling preference metric under the single antenna scenario with $K = 5$ users: (a) CDF of $R_k, \forall k \in \mathcal{K}$ (b) CDF of $\Omega_k, \forall k \in \mathcal{K}$.

of (2.14), it is obtained that $f_{\Omega_k}(\omega) = e^{-x}dx/d\omega$ with $x = H^{-1}(\omega\theta_k)/\gamma_k$. Therefore, the equation (2.13) is rewritten by

$$\theta_k = \int_0^\infty H(x\gamma_k)e^{-x}(1-e^{-x})^{K-1}dx. \tag{2.15}$$

Note that the above equation implies that the long-term average throughput of user k depends only on the SINR of the user and the number of users sharing the same BS.

To further simplify the integral operation, the first order Taylor series expansion is applied to the capacity function as follows:

$$H^{(1)}(x\gamma_k) = H(x_0\gamma_k) + \gamma_k H'(x_0\gamma_k)(x - x_0), \tag{2.16}$$

where x_0 is an arbitrary constant satisfying $x_0 > 0$. By replacing the capacity function by its Taylor series form, the equation (2.15) is changed by

$$\theta_k(x_0) = \int_0^\infty H^{(1)}(x\gamma_k)e^{-x}(1-e^{-x})^{K-1}dx. \tag{2.17}$$

Because the capacity function is assumed to be concave, its linear approximation equation is always greater than the original one (i.e., $H^{(1)}(x) \geq H(x), \forall x > 0$). Therefore, $\theta_k(x_0)$ is always greater than θ_k regardless of x_0. This implies the gap between these two variables is minimized when $\theta_k(x_0)$ is minimized. In order to find an optimal x_0, the partial derivative is taken to both sides of the equation (2.17):

$$\frac{\partial\theta_k(x_0)}{\partial x_0} = \frac{\gamma_k^2 H''(x_0\gamma_k)}{K}\left(\sum_{q=1}\frac{1}{q} - x_0\right). \tag{2.18}$$

During the derivation of the equation, the following integral properties are used [44]: $\int_0^\infty e^{-x}(1-e^{-x})^{K-1}dx = \frac{1}{K}$ and $\int_0^\infty xe^{-x}(1-e^{-x})^{K-1}dx = \frac{1}{K}\sum_{q=1}^{K}\frac{1}{q}$. From the equation (2.18), I can conclude that the integral error between the equations (2.15) and (2.17) is minimized when $x_0 = \sum_{q=1}^{K}\frac{1}{q}$.

Finally, by substituting the optimal x_0 into the equation (2.17), the simplest equation for estimating the long-term average throughput of user k is obtained as

$$\theta_k = \frac{1}{K}H\left(\gamma_k\sum_{q=1}^{K}\frac{1}{q}\right). \tag{2.19}$$

The term $\frac{1}{K}$ in front of the capacity function refers to the temporal fairness among users. One of interesting observation of the above equation is that it contains the multi-user diversity gain term $G(K) = \sum_{q=1}^{K} \frac{1}{q}$. Actually, this term was derived under the assumption of linear data rate function. However, the equation (2.19) shows that this multiuser diversity gain can be used for more general capacity function.

2.4 Throughput Estimation for General Cases

2.4.1 Single User MIMO Scheduling Scenario

The throughput estimation scheme introduced in the previous section is valid only for single antenna systems under the Rayleigh fading environment. However, the statistic of instantaneous data rate becomes more diverse in a practical environment. For instance, the signal propagation environment can not be always explained by Rayleigh distribution. Moreover, the statistic of instantaneous data rate widely varies depending on how many antennas are used or which type of receiver is used. For this reason, I try to derive a general throughput estimation method with an arbitrary distribution of $\Omega_k, \forall k \in \mathcal{K}$. For ease of exposition, a single user MIMO (SU-MIMO) scheduling case is considered. Therefore, every data stream from the BS is transmitted into a target user at each time slot.

The derivation of the throughput estimation method begins from the equation (2.13). If $z = F_{\Omega_k}(\omega) = F_{R_k}(\theta_k \omega)$, then $\partial z / \partial \omega = \theta_k f_{R_k}(\omega)$. Therefore, the equation is written by

$$\theta_k = \int_0^1 F_{R_k}^{-1}(z) z^{K-1} dz. \tag{2.20}$$

To simplify the integral operation, the linear approximation technique is used again. Let $(F_{R_k}^{-1})^{(1)}(z) = F_{R_k}^{-1}(z_0) + (F_{R_k}^{-1})'(z_0)(z - z_0)$ be the first order Taylor series expansion with $z_0 \in (0, 1)$. If the inverse CDF of the equation (2.20) is replaced by

the linear approximation function, θ_k can be expressed by

$$\theta_k(z_0) = \frac{1}{K} F_{R_k}^{-1}(z_0) + \left(\frac{1}{K+1} - \frac{z_0}{K} \right) (F_{R_k}^{-1})'(z_0). \tag{2.21}$$

As the equation (2.20) shows, the contribution of $F_{R_k}^{-1}(z)$ to the integral grows as dummy variable z approaches to one, because of the term z^{K-1}. For this reason, more attention is required for the interval near one. In addition, the PDF of instantaneous data rate has a bell shaped curve for almost every fading scenario (i.e., see [41] for linear receivers), For this reason, $F_{R_k}^{-1}(z)$ can be considered to be convex in the dominant interval of the integral in the equation (2.20).

For a convex function, its linear approximation always less than the original function, and therefore, $\theta_k(z_0)$ is smaller than θ_k regardless of z_0. This means that the gap between these two variable is minimized when $\theta_k(z_0)$ is maximized. To find an optimal z_0, the partial derivative operation is taken to both sides of the equation (2.21), and the result is represented by

$$\frac{\partial \theta_k(z_0)}{\partial z_0} = \left(\frac{1}{K+1} - \frac{z_0}{K} \right) (F_{R_k}^{-1})''(z_0). \tag{2.22}$$

From the above equation, $z_0 = \frac{K}{K+1}$ is selected by an optimal value. By substituting this value into the equation (2.21), the final estimation equation can be obtained as follows:

$$\theta_k = \frac{1}{K} F_{R_k}^{-1} \left(\frac{K}{K+1} \right). \tag{2.23}$$

Using this equation, the long-term average throughput of each user can be estimated from the CDF of instantaneous data rates and the number of users sharing the same resources.

2.4.2 Multiuser MIMO Scheduling Scenario

In the MU-MIMO scheduling scheme, each data stream can be transmitted into a different user simultaneously. Since the channel state information is not available at BS side, precoding or power allocation scheme is not considered in this work. At each time

slot, users report the set of obtainable data rates from all of independently decoded data streams, and the BS assigns each transmit antenna to a single user according to PF scheduling policy. Let $r_{k,m}(t)$ denote the data rate of the m-th data stream for user k and then, the average throughput of user k until time slot t is given as

$$\theta_k(t) = \frac{1}{t}\sum_{\tau=1}^{t}\sum_{m=1}^{n_T} r_{k,m}(\tau) I_{k,m}(\tau), \tag{2.24}$$

where $I_{k,m}(t)$ is the scheduling indicator function between the m-th transmit antenna of the BS and user k at time slot t, and it is represented by

$$I_{k,m}(t) = \begin{cases} 1 & \text{when } k = \arg\max_{k'} \omega_{k',m}(t) \\ 0 & \text{otherwise,} \end{cases} \tag{2.25}$$

where $\omega_{k,m}(t) = \frac{r_{k,m}(t)}{\theta_k(t-1)}$ is the scheduling preference metric of user k from the m-th transmit antenna of the BS.

The collection of data rates from the m-th transmit antenna of the BS to user k, given by $\{r_k^m(t), t = 1, 2, ...\}$, can be modeled as a stochastic process which is characterized by random variable $R_{k,m}$. In other words, each element in this stochastic process is realized from this random variable. Without any precoding or power allocation scheme, the data streams from each transmit antenna are assumed to have a same distribution owing to the symmetry property. Thus, for the simple MU-MIMO scheduling scenario, I can express the long-term average user throughput as the summation of throughput from each transmit antenna as follows.

$$\theta_k = \frac{1}{K}\sum_{m=1}^{n_T} F_{R_{k,m}}^{-1}\left(\frac{K}{K+1}\right). \tag{2.26}$$

2.5 Implementation Issues

To estimate the long-term user throughput using the equations (2.19) for a single antenna scenario or the equations (2.23), (2.26) for more general scenarios including multiple antenna systems, the statistic of instantaneous data rates should be provided.

For this reason, users (or BSs) keep track on the variation of channel state, and choose an appropriate parameter. In details, if the equation (2.19) is used for estimating user throughput under the special case, users monitor the variation of their SINR values for several time slots, and calculate the average value for throughput estimation.

However, if the generalized versions of throughput estimation methods, shown in (2.23), (2.26), are used, the CDF of instantaneous data rate should be provided. Instead of analyzing the statistical properties for numerous environments, a practical implementation skill is applied. To this end, users keep track of the variation of data rate for several time slots, for instance N_s time slots, and consider $F_R^{-1}(\frac{K}{K+1})$ as the $\lceil \frac{N_s K}{K+1} \rceil$-th minimum value from the monitored data rate set, where $\lceil \cdot \rceil$ is the ceiling operation.

2.6 Performance Evaluation and Discussion

2.6.1 Simulation Setup

For simulation, a single cell downlink scenario is considered. The detailed simulation parameters are summarized in Table 3.1. The radius of cell is given by 2 Km, and users are randomly distributed over this region. In order to generate time varying channel, once users are placed in an arbitrary location, they begin to move toward in random directions. For multipath channel, typical urban model is assumed, and each multipath component is independently generated using Jakes simulator [45, 46]. For data rate function, Shannon capacity formula is used. The length of each time slot is given by 10 ms. In order to obtain the long-term throughput of each user, the PF scheduling algorithm runs over 60,000 time slots (i.e., 60,000 $\times$ 10 ms = 10 minutes). Every results in this section is obtained by averaging 1,000 different simulation initializations.

Table 2.1: Simulation parameters

Parameter	**Assumption**
Cell Layout	Single Cell
Carrier Frequency	2,000 MHz
System Bandwidth	10 MHz
Pathloss Model (d in km)	$P_L = 128.1 + 37.6 \log_{10}(d)$
Shadowing	Log-Normal with 8 dB Std. Dev.
Antenna Configuration	1x1 SISO, 2x2 MIMO, 4x4 MIMO
Scheduler	Proportional Fair
BS Transmit Power	43 dBm
Antenna Gain	BS: 7 dBi, User: 0 dBi
User Noise Figure	9 dB
Thermal Noise	-114 dBm/MHz
Channel Model	Typical Urban
User Velocity	3 Km/h

2.6.2 Single Antenna Scenario

To compare the performance of the proposed throughput estimation schemes, the result in [22] is also tested together. The throughput estimation equation of that work is given by

$$\theta_{n,k} = v_1 \sum_{q=0}^{K-1} \binom{K-1}{q} \frac{(-1)^q}{(q+1)} H\left(\frac{v_2}{q+1}\bar{\gamma}_{n,k}\right), \tag{2.27}$$

with two constant parameters $v_1 = 1.4 \log 2$ and $v_2 = 0.82$. The derivation of the above equation is almost similar to what did in this chapter, but the authors use the exponential integral approximation instead of the Taylor series expansion. The equation (2.27) contains a bunch of binomial coefficients that requires a high computational complexity. Moreover, it is obvious that the complexity grows with the number of users sharing the same BS.

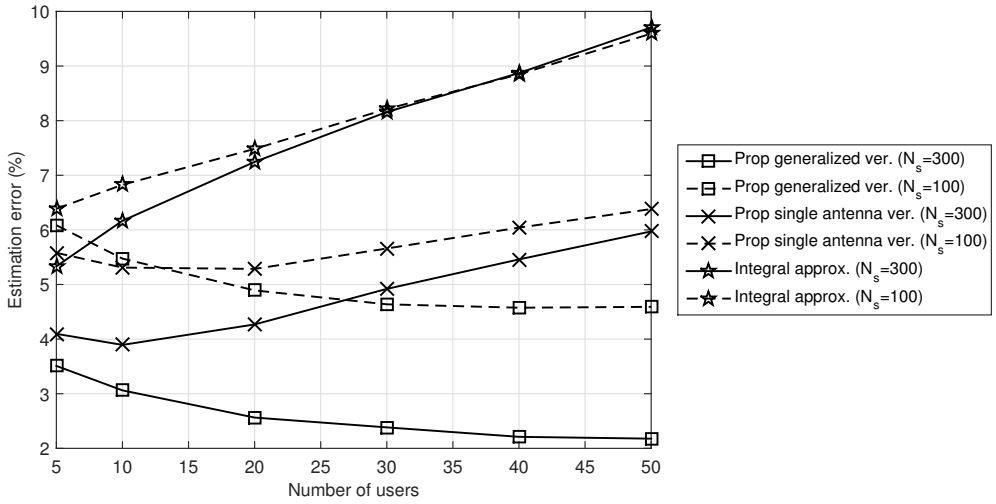
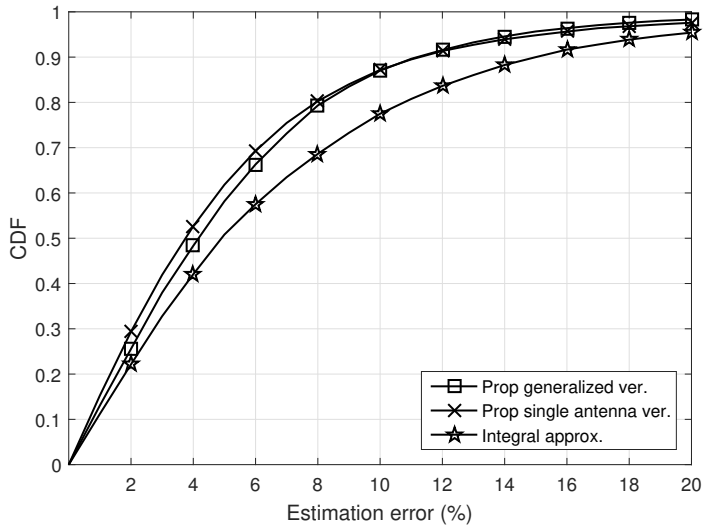


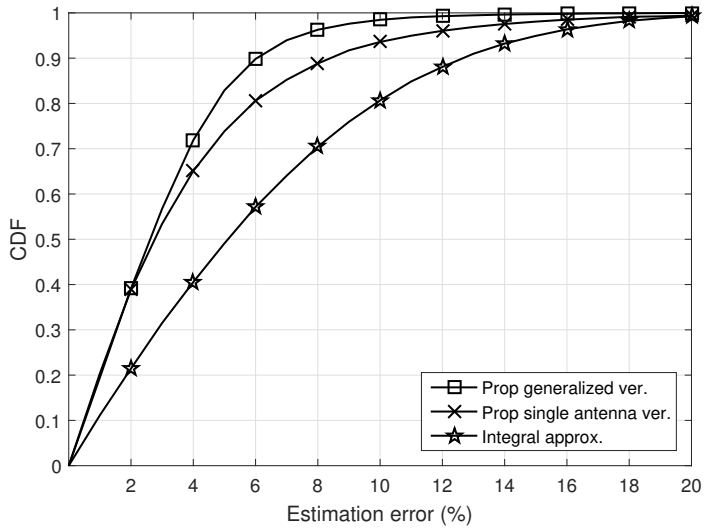
Figure 2.2: Throughput estimation error versus the number of users sharing the same BS.

Figure 2.2 illustrates the throughput estimation error for the proposed schemes and the conventional scheme shown in (2.27). It is observed that every throughput estimation scheme provides more accurate results when abundant samples are used. Among them, the generalized version of throughput estimation scheme provides the most accurate result when $N_s = 300$, where the error is measured by 2-4 %. Every throughput estimation scheme shown in the figure provides reasonable results, where the estimation error is below to 10 % no matter how many users share the same BS.

Figure 2.3 illustrates the CDF of estimation error when the BS serves 20 users. Regardless of the number of observed samples for estimating the long-term throughput, two kinds of proposed throughput estimation schemes outperform the conventional scheme. If the abundant number of samples are used for estimating user throughput, the generalized version of throughput estimation method yields the most accurate result. Moreover, the proposed method for the single antenna version is also provide an acceptable performance.



(a)



(b)

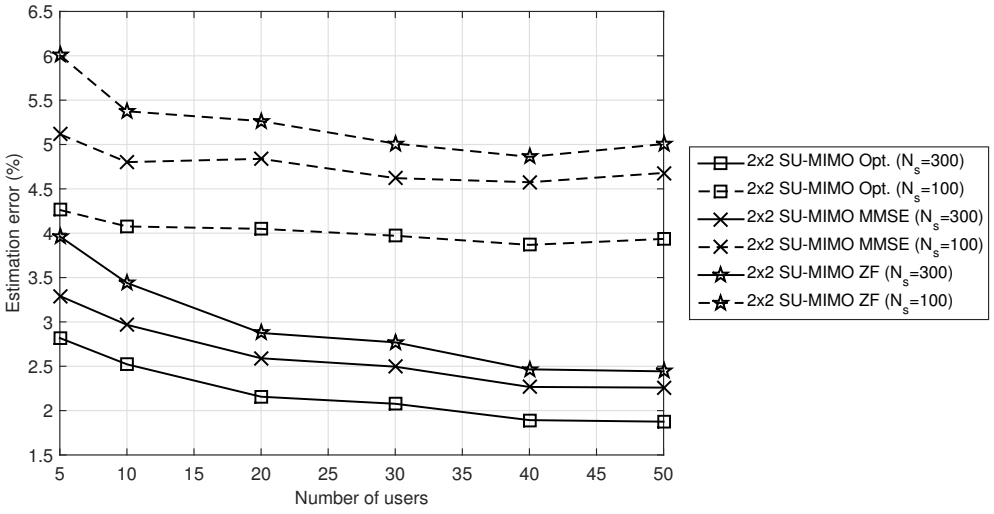
Figure 2.3: CDF of estimation error for single antenna scenario when 20 users are sharing the same BS: (a) $N_s = 100$, (b) $N_s = 300$.

2.6.3 Multiple Antenna Scenario

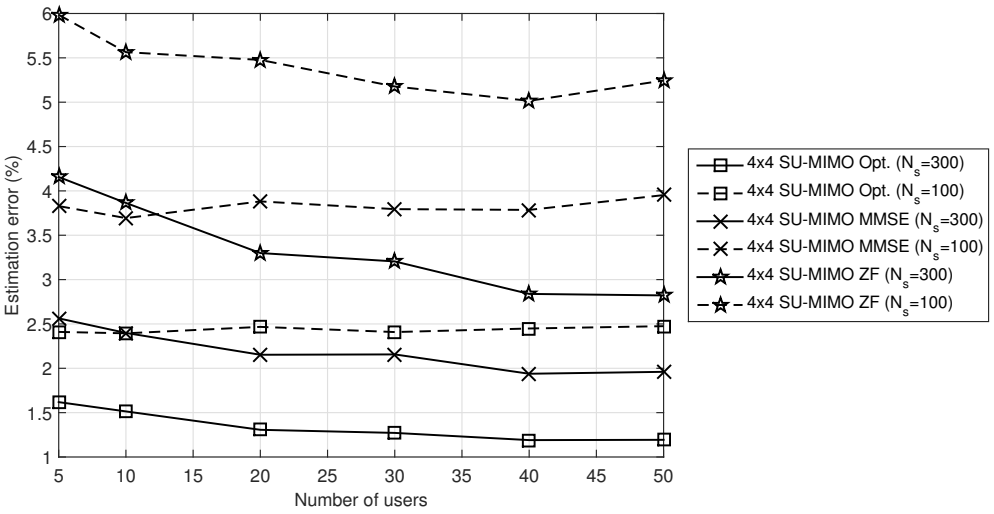
In order to evaluate the user throughput estimation performance for multiple antenna scenarios, the generalized version of throughput estimation scheme is used. The candidate for antenna configuration is 2×2 or 4×4 MIMO, and both single user and multi-user scheduling scenarios are considered. For SU-MIMO scheduling scenario, the three types of receivers, which are the optimal, MMSE, and ZF receivers, are considered. For MU-MIMO scheduling scenario, the two types of receivers, MMSE and ZF receivers, are used.

Figure 2.4 illustrates the throughput estimation error of the proposed scheme under the 2×2 MIMO antenna configuration. In detail, Figure 2.4(a) shows the result of SU-MIMO scheduling scenario, whereas Figure 2.4(b) is the result of MU-MIMO scheduling scenario. It is observed that the more instantaneous data rate samples are used, the more accurate throughput estimation is possible. Moreover, the figures show that the optimal receiver has the best throughput estimation performance regardless of the number of data rate samples. The MMSE receiver has the second best result, and the ZF receiver is followed. In every case, the estimation error is below to 6 % regardless of antenna type.

Figure 2.5 and Figure 2.6 depict the CDF of estimation error when the BS serve 10 users, where the first figure is the result of SU-MIMO scheduling scheduling scenario and the second figure is the result of MU-MIMO scheduling scenario. It is observed that the generalized version of throughput estimation scheme provides accurate results regardless of the configuration of received antenna. In addition, the estimation error tends to decrease if an abundant number of data rate samples are used.

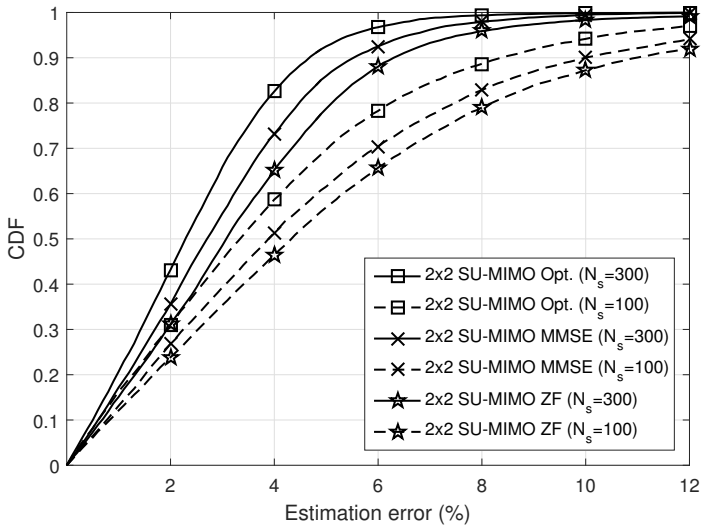


(a)

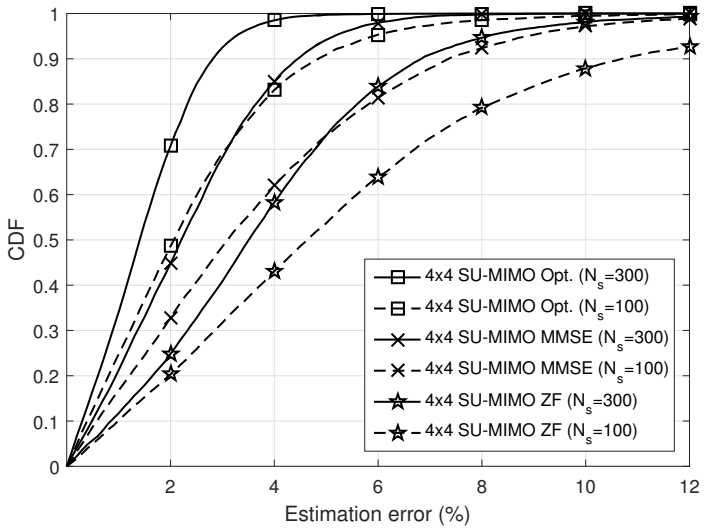


(b)

Figure 2.4: Throughput estimation error of the generalized throughput estimation scheme: (a) 2×2 SU-MIMO scheduling scenario, (b) 4×4 SU-MIMO scheduling scenario.

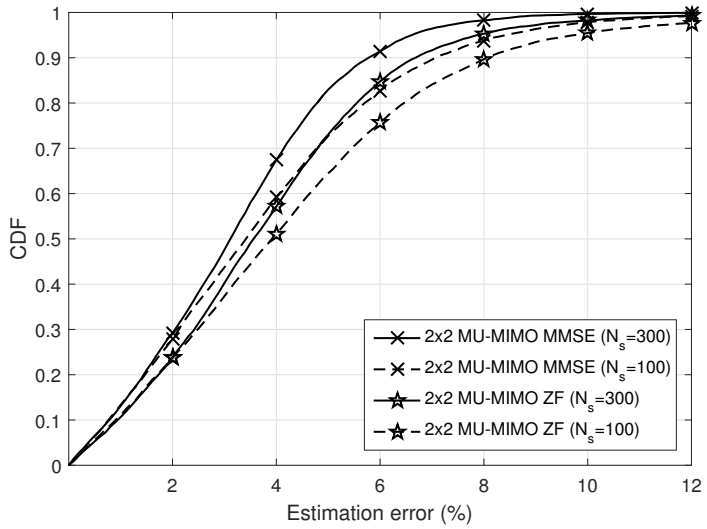


(a)

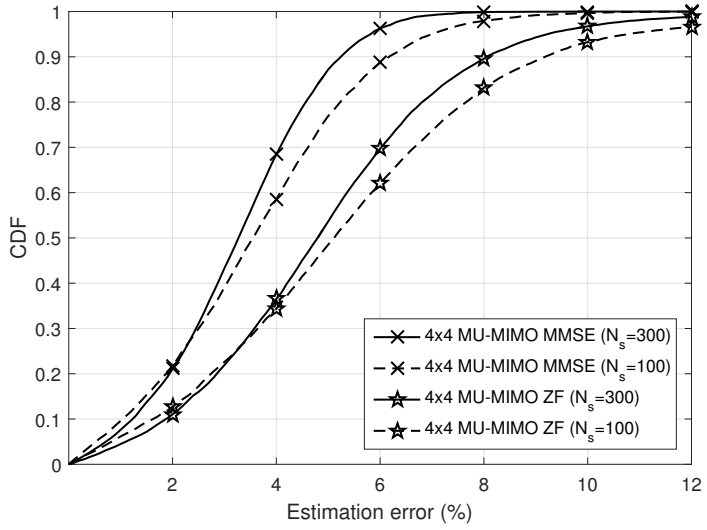


(b)

Figure 2.5: CDF of estimation error for SU-MIMO scheduling scenario where 10 users share the same BS: (a) 2×2 MIMO, (b) 4×4 MIMO.



(a)



(b)

Figure 2.6: CDF of estimation error for MU-MIMO scheduling scenario where 10 users share the same BS: (a) 2×2 MIMO, (b) 4×4 MIMO.

Chapter 3

DYNAMIC USER ASSOCIATION IN MULTI-CELL CELLULAR NETWORKS

3.1 Motivation

In the previous chapter, it is observed that the long-term average throughput of each user depends on not only the quality of communication link but also how many users sharing the same resources. If too many users are connected to a single BS, they inevitably suffer from low service quality. In that case, some users, especially those located in the cell boundary regions, could experience better service quality if they change their association from the congested BS to nearby less congested one. For this reason, cell-site selection based only on the received signal strength does not always guarantee the best service quality. Therefore, an additional parameter, such as the congestion level of each cell, should be considered.

One of major challenge in developing dynamic user association strategies is predicting how a specific event (such as the handover of a user) will impact on the overall network performance. This is closely related to estimating each individual user's throughput, because system performance is generally measured by how well the resources are allocated to them. For this, many studies simply assume the actu-

ally achievable throughput of each user as their mean rate divided by the number of users sharing the same BS. Furthermore, a few works even ignore the fluctuating nature of the wireless channel [27, 40]. To reflect the fading phenomenon, several works assume a Rayleigh fading channel over every link to derive a deterministic throughput estimation equation [29, 39].

In this chapter, a centralized load-aware cell-site selection scheme, where BSs decide the association of users, is developed. To estimate the long-term achievable throughput of each user, the proposed method introduced in the equation (2.19) is used. This method requires only the average SINR parameter for characterizing the fluctuation of wireless channel. For this reason, uplink overhead from users and message exchange between BSs for user association can be minimized. Once a central node, i.e., one of BS in the group of neighboring BSs, collects the average SINR from every BS to every user, it runs the dynamic user association algorithm to decide an optimal association. And then, users change their association according to the decision made by the central node.

3.2 System Model

A multi-cell downlink system is considered in this chapter. The set of BSs is denoted by $\mathcal{N} = \{1, ..., N\}$, and the set of users is given by $\mathcal{K} = \{1, ..., K\}$. The system band is equally partitioned into the set of sub-bands, which is represented by $\mathcal{S} = \{1, ..., S\}$. Any of two different sub-bands are designed in a non-overlapping manner, and therefore, there is no inter-sub-band interference within a cell. On the other hand, transmissions in neighboring BSs through the same sub-band still remain as interference sources. In order to provide an acceptable signal strength for users, especially those located in cell boundary region, the fractional frequency reuse (FFR) scheme can be applied. In that case, each BS is allowed to use a specific subset of $\mathcal{S}$ based on the pre-determined FFR patterns [8]. It is also important for improving system perfor-

mance to design the FFR patterns and to assign these patterns to each BS. However, such an issue is beyond interest of this chapter, the user algorithm will be developed to compatible with any types of FFR pattern are used.

To generate a realistic traffic activity, it is assumed that users randomly arrive at the networks and depart after an arbitrary amount of time. A newly arrived user is randomly placed within the coverage of multi-cell networks, and then, the user begins to move toward an arbitrary direction. User are assumed to constantly generate data traffic in their holding time, and thus, infinitely backlogged queue model is adopted. For data service, users can be associated with any BSs in the vicinity, but they can be served by at most one BS simultaneously. To let the serving BS know the link quality, users periodically measure the channel quality and report it via reverse link. Based on the channel state feedback from users, each BS serves users, those associated with itself, by using both time and frequency multiplexing schemes. Note that user are not allowed to be associated with multiple BSs simultaneously, they can measure channel qualities form all the nearby BSs.

At time slot t, the instant SINR from BS n to user k in sub-band s is expressed by

$$\gamma_{n,k}^{s}(t) = \frac{g_{n,k}^{s}(t)p_{n}^{s}}{\sum_{n' \in \mathcal{N}, n' \neq n} g_{n',k}^{s}(t)p_{n'}^{s} + N_0}, \tag{3.1}$$

where $g_{n,k}^{s}(t)$ represents the channel coefficient from BS n to user k in sub-band s, including both large-scale characteristic (i.e., pathloss and shadowing) and small-scale fading. p_{n}^{s} is the transmission power of BS n in sub-band s. Note that BSs equally allocate their transmission power to all of the available sub-bands. For this reason, if BS n is not allowed to use sub-band s, then $p_{n}^{s} = 0$ and $\gamma_{n,k}^{s}(t) = 0, \forall t, \ \forall k \in \mathcal{K}$. Finally, N_0 represents noise spectral density. Because BSs change the MCS level according to link quality, the instantaneous data rate is determined by the instant SINR level. Thus the data rate from BS n to user k through sub-band s can be expressed by

$$r_{n,k}^{s}(t) = H(\gamma_{n,k}^{s}), \tag{3.2}$$

where $H(\cdot)$ is an increasing, concave, and differentiable capacity function. As I men-

tioned in the equation (2.14), this function is usually expressed by Shannon capacity formula, $H(\gamma) = \log_2(1+\gamma)$.

For theoretical analysis, the time collection of SINR values, specifically from BS n to user k through sub-band s, $\{\gamma^s_{n,k}(t), t = 1, 2, ...\}$ is considered as a stochastic process, and each element is assumed to be realized from random variable $\Gamma^s_{n,k}$. In addition, this random variable is represented in the form of $\Gamma^s_{n,k} = X^s_{n,k}\bar{\gamma}^s_{n,k}$, where $\bar{\gamma}^s_{n,k}$ denotes average SINR and $X^s_{n,k}$ is exponential random variable with mean of one. The exponential random variable implies that the channel from the serving BS experiences the Rayleigh fading environment. Obviously, the assumption of $\Gamma^s_{n,k} = X^s_{n,k}\bar{\gamma}^s_{n,k}$ is only valid when the channel fluctuation from the serving BS is Rayleigh while other interfering channels are stationary. However, this assumption enables user throughput estimation possible, and therefore, I keep it even though the estimation error occurs. As a consequence, the time collection of instantaneous data rates, $\{r^s_{n,k}(t), t = 1, 2, ...\}$, is also considered as a stochastic process, where each element is realized from random variable $R^s_{n,k} = H(\Gamma^s_{n,k}) = H(X^s_{n,k}\bar{\gamma}^s_{n,k})$.

3.3 Problem Formulation

3.3.1 Objective and Optimal Algorithm

The objective of this work is optimizing the system utility, which is simply defined as the sum of all individual users utilities. The utility of user k is defined as $U_k(\theta_k)$, where $U_k(\cdot)$ is an increasing, concave, and differentiable utility function. And then, the system utility is expressed by $U = \sum_{k\in\mathcal{K}} U_k(\theta_k)$. It is obvious that the strategy for achieving the maximum system performance can be widely different according to the form of utility functions. Among various candidates for utility functions, the log utility function (i.e., $U_k(\theta_k) = \log(\theta_k), \forall k \in \mathcal{K}$) is used in this work. Because this types of function pays more attention to improve the performance of users suffering from low service quality, and therefore, it is adequate for the purpose of load balancing.

Time index t can be attached to θ_k in order to indicate the average throughput of user k up to the corresponding time slot, and it is updated as follows:

$$\theta_k(t+1) = (1-\beta)\theta_k(t) + \beta \sum_{n \in \mathcal{N}} \mathbf{r}_{n,k}(t) \cdot \mathbf{I}_{n,k}^T(t), \tag{3.3}$$

where $\mathbf{r}_{n,k}(t) = [r_{n,k}^1(t), ..., r_{n,k}^S(t)]$ represents the instantaneous data rate vector from BS n to user k at time slot t. Moreover, $\mathbf{I}_{n,k} = [I_{n,k}^1(t), ..., I_{n,k}^S(t)]$ denotes the scheduling indicator vector, where $I_{n,k}^s(t) = 1$ implies that BS n serves user k through sub-band s at time slot t. For this representation, $\mathbf{r}_{n,k}(t) \cdot \mathbf{I}_{n,k}^T(t)$ is the total amount of data transmitted from BS n to user k. Finally, β is the time constant for the moving average, and it is typically on the order of several thousand time slots (i.e., $\beta = 1/1000$).

In order to investigate the upper bound of the system utility, the following gradient-based scheduling process is used. At each time slot, BSs allocate each of their available resources to an appropriate target user, who contributes the maximum increase in the system utility when being served. To implement it, a central node, which has every network information such as the instantaneous data rate from every BS to every user, should make an overall scheduling decision as follows:

$$\begin{aligned}
&\max \sum_{k \in \mathcal{K}} \sum_{n \in \mathcal{N}} U_k' \left(\bar{\theta}_k(t-1)\right) \mathbf{r}_{n,k}(t) \cdot \mathbf{I}_{n,k}(t)^T \\
&\text{subject to } I_{n,k}^s(t) \in \{0,1\},\ \forall n,k,s \\
&\qquad \sum_{k \in \mathcal{K}} I_{n,k}^s(t) \leq 1,\ \forall n,s \\
&\qquad \sum_{s \in \mathcal{S}} I_{n,k}^s(t) \sum_{s' \in \mathcal{S}} I_{n,k}^{s'}(t) = 0,\ \forall k \in \mathcal{K}, \forall n \neq m.
\end{aligned} \tag{3.4}$$

The second constraint condition means that an available sub-band from a BS should be allocated to a single target user, whereas the last condition implies that users cannot be served by multiple BSs simultaneously.

Theoretically, it is known that such a myopic scheduling strategy approximately achieves the optimal result [19]. Even though this algorithm provide a notable perfor-

mance, it is difficult to implement in a practical system. First of all, the concept of serving BS is not exist in the algorithm, and thus, users can be served by a BS in this time slot and by another BS in the next time slot. For this reason, they should report the instantaneous data rate from all the BSs in the vicinity, which are potential candidates for being served. Such feedback transmissions naturally yields a tremendous signalling overhead as well as excessive message change among BSs. The second issue is related to the calculation complexity. Even though the central node has enough information, solving the optimization problem in (3.4) within a short time is not an easy task. Finally, users should frequently change their association from a BS to another BS, according to the solution.

Because of these difficulties, the result of the above optimization problem will be used to evaluate how close the proposed algorithm achieves the upper bound of the system. Moreover, to reduce the complexity of solving the problem in (3.4), the last constraint is removed during simulation. Therefore, the scheduling indicator variables are simply determined by

$$I^s_{n,k}(t) = \begin{cases} 1 & \text{when } k = \arg\max_{k'} \omega^s_{m,k'}(t) \\ 0 & \text{otherwise,} \end{cases} \tag{3.5}$$

where the scheduling preference metric is given by $\omega^s_{m,k}(t) = U'_k(\theta_k(t-1))r^s_{n,k}(t)$. It is obvious that the result in (3.5) provides better system performance than (3.4), because one constraint condition is deleted from the original optimization problem. For this reason, the result with scheduling decision in (3.5) will be used as a benchmark of the proposed scheme.

3.3.2 User Association Problem

Unlike, the global scheduling processes shown in (3.4) or (3.5), users generally become associated with a single BS for a while, and are served by a local scheduler which is executed in a independent manner. Let $a_{n,k} \in \{0, 1\}$ be the association variable

between BS n and user k, where $a_{n,k} = 1$ indicates that the user is associated with the BS.

In order to estimate the long-term throughput of each user, the equation (2.14) is used in here, because this equation works well with just two parameters: the average SINR and the number of users sharing the same resources. However, both the equations (2.14) and (2.23) are designed for a single sub-band scenario, they should be appropriately modified for multiple sub-band systems. To this end, I consider that the long-term throughput of a user results from the sum of throughput achieved via each sub-band, and use the proposed throughput estimation method to calculate the amount of achieved throughput for each sub-band. Then, the achieved throughput for user k through sub-band s can be expressed by

$$\theta_k^s = \psi_{n,k}(K_n, \bar{\gamma}_{n,k}^s) = \frac{1}{K_n} H\left(\bar{\gamma}_{n,k}^s G(K_n)\right), \tag{3.6}$$

where $K_n = \sum_{k\in\mathcal{K}} a_{n,k}$ represents the number of active users at BS n, and $\psi_{n,k}(\cdot,\cdot)$ is a function of the amount of throughput from BS n to user k. It is clear that if the user is not associated with the BS (i.e., $a_{n,k} = 0$), then $\psi_{n,k}(K_n, \bar{\gamma}_{n,k}^s) = 0, \forall s \in \mathcal{S}$. The multiuser diversity gain is expressed by a function $G(K) = \sum_{q=1}^{K} \frac{1}{q}$ for simplicity. Finally, the long-term average throughput of user k is expressed by $\theta_k = \sum_{s\in\mathcal{S}} \theta_n^s$.

Using the equation (3.6), the user association problem for optimizing the system utility is formulated as follows:

$$\begin{aligned}
&\max \sum_{k\in\mathcal{K}} U_k\Big(\sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}} \psi_{n,k}(Y_n, \bar{\gamma}_{n,k}^s)\Big) \\
&\text{subject to } a_{n,k} \in \{1,0\},\ \forall n \in \mathcal{N},\ \forall k \in \mathcal{K}, \\
&\qquad \sum_{n\in\mathcal{N}} a_{n,k} = 1,\ \forall k \in \mathcal{K}, \\
&\qquad \sum_{k\in\mathcal{K}} a_{n,k} = Y_n,\ \forall n \in \mathcal{N}.
\end{aligned} \tag{3.7}$$

The second constraint implies that users can be associated with only one BS. Every combination of user association is finite, and thus, the above user association problem

has at least one optimal solution. However, obtaining it within a short time may not seem to be possible, because the size of solution set grows exponentially with the number of BSs multiplied by the number of users. Theoretically, this problem is known to be reducible to the three-dimensional matching problem that is NP-complete [28]. For this reason, algorithms for obtaining an optimal solution within a polynomial time do not exist.

3.4 Centralized Dynamic User Association Algorithm

Using the equation (3.6), BSs can predict the long-term throughput of users if the average SINR of every sub-band is provided. To this end, users continuously measures the SINR level from not only their serving BS but also nearby potential target BSs, and report every average value via reverse link in every short time interval (e.g., several hundreds of time slots). Once BSs collect these information, a central node finally determines the overall association between BSs and users. The detailed operation is developed based on the following propositions.

Proposition 3.1 (Effect of a single handover event)**.** Suppose that user i is currently associated with BS n. If this user changes the association to another BS m while other users keep their association, then the system utility will be changed by

$$\Delta U_{i,n\to m} = \sum_{k\in\mathcal{K}_m^*}\sum_{s\in\mathcal{S}}\frac{1}{K_m^*}H\left(\bar{\gamma}_{m,k}^s G(K_m^*)\right) + \sum_{k\in\mathcal{K}_n^*}\sum_{s\in\mathcal{S}}\frac{1}{K_n^*}H\left(\bar{\gamma}_{n,k}^s G(K_n^*)\right)$$
$$- \sum_{k\in\mathcal{K}_m}\sum_{s\in\mathcal{S}}\frac{1}{K_m}H\left(\bar{\gamma}_{m,k}^s G(K_m)\right) - \sum_{k\in\mathcal{K}_n}\sum_{s\in\mathcal{S}}\frac{1}{K_n}H\left(\bar{\gamma}_{n,k}^s G(K_n)\right), \tag{3.8}$$

where parameters with star mark $(\cdot)^*$ imply the values after handover of user i is done. In detail, $\mathcal{K}_m^* = \mathcal{K}_m + \{i\}$ and $\mathcal{K}_m^* = \mathcal{K}_m - \{i\}$. And the number of users at BS n and m is $K_m^* = K_m + 1$ and $K_n^* = K_n - 1$.

Proposition 3.2 (Effect of a new arrival user)**.** Suppose that recently arrival user i becomes associated with BS n. Then the system utility difference will be expressed by

$$\Delta U_{i,new\to n} = \sum_{k\in\mathcal{K}_n^*}\sum_{s\in\mathcal{S}}\frac{1}{K_n^*}H\left(\bar{\gamma}_{n,k}^s G(K_n^*)\right) - \sum_{k\in\mathcal{K}_n}\sum_{s\in\mathcal{S}}\frac{1}{K_n}H\left(\bar{\gamma}_{n,k}^s G(K_n)\right), \quad (3.9)$$

where $\mathcal{K}_n^* = \mathcal{K}_n + \{i\}$ and $K_n^* = K_n + 1$.

The above propositions are nothing but the difference between system utilities before and after association (or reassociation) of user i occurs. For this reason, the detailed proofs are omitted.

Using these propositions, Algorithm 3.1 for the central node to determine the overall association can be developed. An important thing to consider is that these propositions estimates the system utility change only if a single user change its association while others keep their association. If multiple users change their association simultaneously, the equations (3.8), (3.9) may yield wrong results. For this reason, the algorithm is designed to perform iterative processes, where a single association change is taken place per each iteration. Such a process is continued until no more association changes occur (i.e., $flag = 0$). Note that this iterative algorithm is executed at the central node in every handover interval.

If the central node obtains optimal associations, it transmits the results to every involved BS. Obviously, this is a kind of centralized algorithm where the association control is achieved at a central node. However, our algorithm can be implemented even in a distributed manner with small modifications. For instance, each BS could broadcast the number of active users, as additional information for associations, so that each user can estimate the receivable throughput as well as the utilities from all of the nearby BSs. The detailed operation will be depicted in the next chapter.

Algorithm 3.1 Centralized user association algorithm

INPUT

Current user association: $a_{n,k}, \forall n \in \mathcal{N}, k \in \mathcal{K}$

Average SINR values: $\bar{\gamma}_{n,k}^{s}, \forall n \in \mathcal{N}, k \in \mathcal{K}, s \in \mathcal{S}$

INITIALIZATION

Initialize the set of users: $\mathcal{K}_{new}$, $\mathcal{K}_n, \forall n \in \mathcal{N}$

Initialize the number of active users: $Y_n, \forall n \in \mathcal{N}$

$flag \leftarrow 1$

ITERATION

while $flag$ **do**

$flag \leftarrow 0$

if $\mathcal{K}_{new} \neq \phi$ **then**

Find the maximum $\Delta U_{i,new \rightarrow n}$ among $i \in \mathcal{K}_{new}, n \in \mathcal{N}$

$a_{n,i} \leftarrow 1$, $flag \leftarrow 1$

$\mathcal{K}_{new} \leftarrow \mathcal{K}_{new} - \{i\}$, $\mathcal{K}_n \leftarrow \mathcal{K}_n \cup \{i\}$, $Y_n \leftarrow Y_n + 1$

else

Find the maximum $\Delta U_{i,n \rightarrow m}$ among $i \in \mathcal{K}$, $n, m \in \mathcal{N}$, and $a_{n,i} = 1$

if $\Delta U_{i,n \rightarrow m} > 0$ **then**

$a_{n,i} \leftarrow 0$, $a_{m,i} \leftarrow 1$, $flag \leftarrow 1$

$\mathcal{K}_n \leftarrow \mathcal{K}_n - \{i\}$, $\mathcal{K}_m \leftarrow \mathcal{K}_m \cup \{i\}$, $Y_n \leftarrow Y_n - 1$, $Y_m \leftarrow Y_m + 1$

end if

end if

end while

3.5 Performance Evaluation and Discussion

3.5.1 Simulation Setup

For system-level simulation, an ideal cellular deployment scenario which consists of 21 hexagonal cells is assumed. Figure 3.1 represents the baseline scenario, where three different types of cells are deployed and every BS are located at the center of the cell. In order to remove all the border effects, a wrap-around technique is applied.

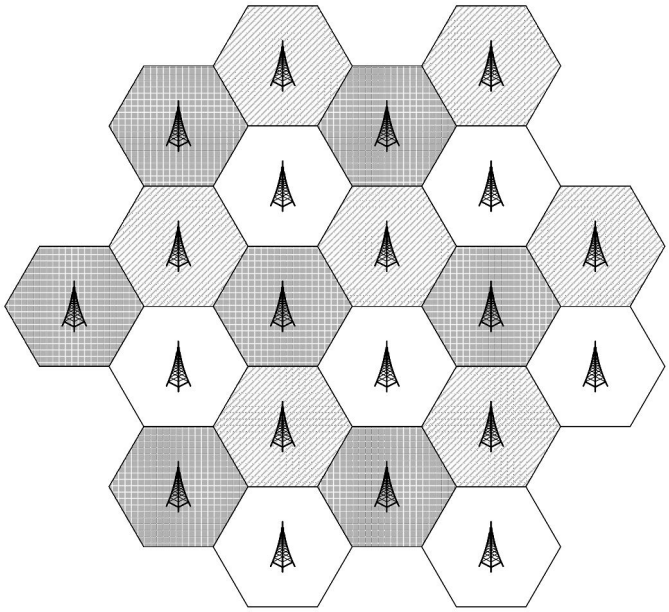


Figure 3.1: The baseline cell deployment scenario, where three different types of cells are exist and BSs are located at the center of the cell.

The total system bandwidth is divided into 12 sub-bands. To increase cell-edge performance, the FFR scheme is applied, where 3/4 of sub-bands are shared by all the BSs and remaining sub-bands are allocated to each type of cell with a reuse factor of 1/3. In other words, each BS can utilize 10 sub-bands. Users are assumed to experience the frequency-selective Rayleigh fading channels from every BS. To this end, each multipath component is generated using Jakes simulator by assuming the user velocity

Table 3.1: Simulation parameters

Parameter	**Assumption**
Cell Layout	Hexagonal 21 Cells
Inter-Site Distance	1,000 m
Carrier Frequency	2,000 MHz
Sub-Band Bandwidth	720 kHz
Number of Sub-Bands	12
Pathloss Model (d in km)	$P_L = 128.1 + 37.6 \log_{10}(d)$
Shadowing	Log-Normal with 8 dB Std. Dev.
Antenna Configuration	1x1 SISO
Scheduler	Proportional Fair
BS Transmit Power	43 dBm
Antenna Gain	BS: 5 dBi, User: 0 dBi
User Noise Figure	9 dB
Thermal Noise	-114 dBm/MHz
Channel Model	Typical Urban
User Velocity	3 Km/h

of 3Km/h. Table 3.1 summarizes simulation parameters. Other parameters not shown in here are referred in the IEEE 802.16m evaluation methodology document [47].

To generate realistic traffic activities, dynamic user arrival and departure model is used. Let $\rho_l = \frac{\lambda_l}{\mu}$ denote the traffic intensity of l-th cell, where λ_l is the Poisson arrival rate of users, and $\frac{1}{\mu} = 20$ (seconds) is the exponentially distributed holding time of each user. According to this process, the number of active users at BS l dynamically fluctuates around the traffic intensity parameter ρ_l. For this reason, I can make an unbalanced loading scenario by assigning traffic intensity parameter to each BS differently. Once users arrive at the network, they move in a random direction. The association changes occurred in every 100 ms.

3.5.2 Throughput Estimation Error in Multi-cell Environments

At first, I verify the accuracy of the proposed throughput estimation formula. I randomly distribute a fixed number of users into the central cell, and run the PF scheduling algorithm over 60,000 time slots, so that the average throughput of each user converges to a fixed value. Also, users are just moving around their original location in that their mean SINR values do not vary largely. The performance metric is the relative error to the actual throughput that is obtained through the simulation. For the purpose of performance comparison, I also test the conventional throughput estimation method shown in the equation (2.27).

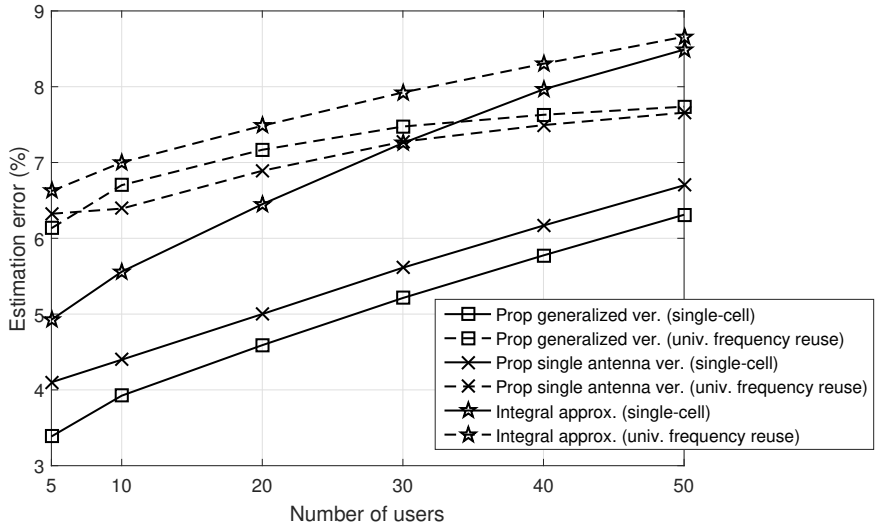


Figure 3.2: Throughput estimation error for the proposed schemes and the conventional scheme under single cell and multi-cell environments.

Figure 3.2 illustrates the relative error versus the number of active users. The dashed lines present the results of single cell case, where neighboring interfering cells are absent. Meanwhile, the solid lines depict the outcomes of multi-cell environment where every BSs share the same resources. Each value is obtained by averaging the outcomes from 100 different simulation initializations. In both environments (i.e., single and multi-cell cases), the proposed throughput estimation methods yield more ac-

curate results than the conventional scheme. Also it is observed that estimation errors are low in a single cell environment.

3.5.3 Load Balancing Effect

For the purpose of performance comparison, the SINR-based cell selection scheme is considered, where each user tries to be associated with the BS providing the best signal strength. Because the objective of this work is maximizing the sum of log utilities, the system performance can be measured in the form of the geometric average of throughput $GAT = (\Pi_{k\in\mathcal{K}}\bar{\theta}_k)^{1/K}$. Additionally, to measure the performance of users suffering from low service quality, the average 5th percentile user throughput is observed.

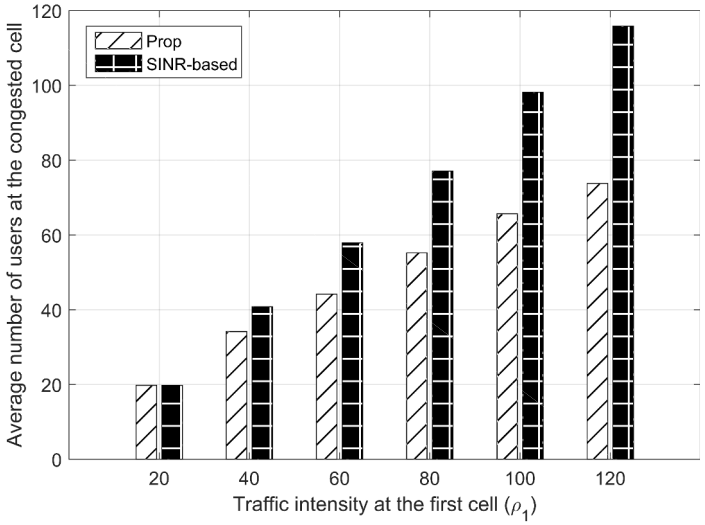
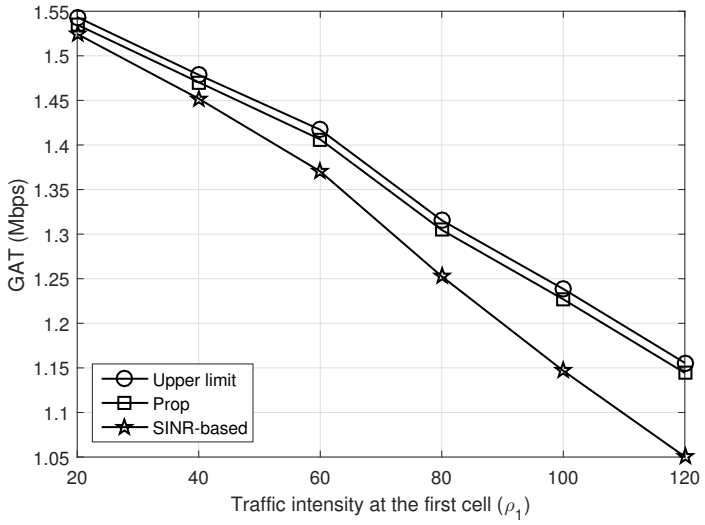
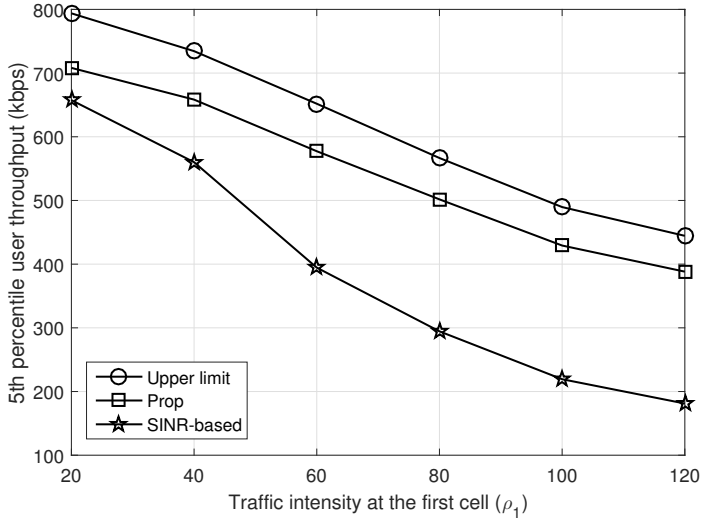


Figure 3.3: Average number of active users at the congested cell

Figure 3.3 illustrates the number of active users at the congested BS. The proposed user association scheme redirects users at the congested BS into neighboring BSs for the purpose of traffic offload. The next figure shows the results of the load balancing effect. Figure 3.4(a) shows the GAT performance the traffic load at the central cell varies from 20 to 120 whereas that of other BSs is fixed to 20 (i.e., $\rho_1 = \rho_2 = 20$).



(a)



(b)

Figure 3.4: System performance versus traffic intensity at the congested cell: (a) Geometric average of throughput (GAT), (b) Average 5th percentile user throughput. The traffic intensity at other macro cells is assumed to be 30 (i.e., $\rho_n = 30, \forall n \neq 1$).

Regardless of the traffic load at the congested cell, the proposed scheme attains the upper limit of the system. Figure 3.4(b) represents the 5th percentile user throughput. It is observed that the proposed scheme enhance the cell-edge performance by $7.5 - 114\%$ according to the traffic load of the central cell. Even when the traffic intensity of every cell is identical (i.e., $\rho_0 = 20$), the temporal load of each BS is possibly different due to the random arrival and departure of users. For this reason, the proposed scheme can achieve the performance improvements.

Chapter 4

DECENTRALIZED USER ASSOCIATION METHOD IN HETEROGENEOUS CELLULAR NETWORKS

4.1 Motivation

In the previous chapter, the centralized load balancing algorithm in multi-cell cellular networks is introduced. This scheme redirects users at a congested BS into another neighboring less congested BS, for the purpose of improving the overall system performance. Such a load balancing algorithm is also necessary for associating users in heterogeneous cellular networks, where small-sized cells, such as pico and femto cells, coexist with the macro BSs [48,49]. As the name indicates, small cell BSs are designed to occupy much less room for install and consume much less power as well as operating cost, compared to the conventional macro cell BSs.

In this chapter, a decentralized user association framework is proposed, where each user autonomously chooses an optimal BS by considering the overall system performance and BSs merely accept or reject association requests from users. The reason why decentralized strategy is adopted in this chapter is that the centralized user association requires users to feedback the channel status from all the BSs in the vicinity, and consequently, it yields uplink overhead as well as excessive message exchange among

BSs. Moreover, unlike the centralized method developed in Chapter 3, the generalized version of throughput estimation method is used in this chapter, in order to guarantee the compatibility for various scenarios including multiple antenna configuration. In that case, the fluctuation of wireless channel is not characterized by a single parameter anymore, and thus, the overhead for the centralized strategy will be increased inevitably.

Instead, if users themselves can estimate the achievable throughput and corresponding service quality from all the nearby candidate BSs, they does not need to report excessive parameters for association. To help users estimate the achievable throughput, BSs should periodically broadcast a few parameters. By receiving these parameters, users can estimate the throughput and make a handover decision. According to this guideline, two decentralized user association algorithms are developed. In the first scheme, BSs are involved the handover process, where they transfer the handover requests from users into neighboring BSs. By slightly modifying this, in the seconde method, users directly send the handover requests into the target BSs.

4.2 System Model

I consider a heterogeneous cellular system, where various types of infrastructures such as macro and small cell BSs coexist in the network. The set of BSs is denoted by $\mathcal{N} = \{1, ..., N\}$, and it can be partitioned into several subsets according to how many different types of BSs are deployed. For instance, if macro and pico BSs coexist in the network, the set of BS can be represented by $\mathcal{N} = \mathcal{N}_{macro} \cup \mathcal{N}_{pico}$, where two partitions denote the set of macro BSs and the pico BSs, respectively. The set of users is denoted by $\mathcal{K} = \{1, ..., K\}$. The total system band is equally divided into the set of sub-bands, which is expressed by $\mathcal{S} = \{1, ..., S\}$. Furthermore, time is slotted, so that each BS serves users using both time and frequency multiplexing schemes. In order to manage the inter-cell and inter-tier interference, the FFR scheme is applied in

this work, and therefore, each BS can utilize a subset of sub-bands according to the pre-determined FFR pattern.

Because this work focuses on the load balancing between BSs regardless of their specific types, open access mode is assumed. For this reason, users can freely access any types of BSs. An important thing to consider is the different physical sizes and capabilities across different types of BSs. For example, small cell BSs are generally designed to consume limited power in both operation and signal transmission, and therefore, have relatively small downlink coverage as compared to the macro BSs. In contrast, the uplink coverage is determined by the transmission power of users; hence it will not be different according to the BS type. Such a difference between downlink and uplink coverage area in heterogeneous cellular network makes the user association problem more complex. For this reason, I focus on the user association in downlink traffic, because a large portion of data traffic is still generated in this direction.

Like the previous chapter, it is assumed that users randomly arrive at the network and departure after an arbitrary amount of time. During their stay time, they move toward a random way point while constantly generating downlink data traffic. For this reason, a full buffer model with an infinitely backlogged queue is applied. At any moment in time, users are able to connect any types of BS, but no more than one simultaneously. However, users can measure channel status from other BSs in the vicinity and even receive a few fundamental parameters for association. For the purpose of transmission rate adaptation and scheduling, users periodically report channel state from the serving BS.

For a specific time slot t, let $r^s_{n,k}(t)$ denote the instantaneous data rate from BS n to user k through sub-band s. Basically, this value is determined by (2.3) in case of single antenna scenario, and (2.4)-(2.6) in case of multiple antenna scenario according to the receiver type. It is known that several users can be served by using a single wireless resources by transmitting multiple, independent data streams using multiple transmit antenna. However, for ease of exposition, I assume a single user scheduling scenario,

where every data possible data streams within a sub-band is transmitted to a target user. To mathematically analyze the time fluctuation of data rate due to a multipath propagation environment, it is assumed that the time collection of instantaneous data rate $\{r^s_{n,k}(t), t = 1, ...\}$ is considered as a stochastic process, where each element is realized from random variable $R^s_{n,k}$.

4.3 Problem Formulation

The objective of this chapter is optimizing system utility which is defined by $U = \sum_{k\in\mathcal{K}} U_k(\theta_k)$, where $U_k(\cdot)$ is an increasing, concave, and differentiable utility function for user k, and θ_k means the long-term average throughput of the user. Like the previous chapter, the log utility function is used for all users, because this kind of utility function is well-suited for the purpose of load balancing. The average throughput is updated according to (3.3), and the upper limit of the system can be obtained by (3.4) or (3.5). For reducing the calculation complexity, the global scheduling process of (3.5) will be utilized during simulation.

Let $a_{n,k} \in \{0, 1\}$ be the association variable between BS n and user k, where the element 1 implies that the association is established. Because each user can be associated with at most one BS at any moment, I have $\sum_{n\in\mathcal{N}} a_{n,k} = 1, \forall k \in \mathcal{K}$. To separate users according to their serving BS, I define $\mathcal{K}_n = \{k \in \mathcal{K} | a_{n,k} = 1\}$ as the set of users associated with BS n. Then, the long-term average throughput of user $k \in \mathcal{K}_n$ is determined by the number of users sharing the same BS, denoted by Y, and by that user's link quality. This relationship can be expressed as follows.

$$\bar{\theta}_k = \psi_n(Y, \mathbf{R}_{n,k}), \tag{4.1}$$

where $\mathbf{R}_{n,k} = [R^1_{n,k}, ..., R^s_{n,k}]$ represents the random rate vector, and $\psi_n(\cdot, \cdot)$ is the throughput function of BS n.

According to equation (4.1), the performance of each user varies with how many users share the same BS. In addition, the system utility, which is simply defined as

the sum of each individual user's utility, is also determined by the user association. In other words, the objective to maximize system performance is achieved by finding the best user association, through the following optimization problem:

$$\begin{aligned} &\max \sum_{n\in\mathcal{N}} \sum_{k\in\mathcal{K}_n} U_k(\psi_n(Y_n, \mathbf{R}_{n,k})) \\ &\text{subject to } a_{n,k} \in \{0,1\},\ \forall n,k, \\ &\qquad \sum_{n\in\mathcal{N}} a_{n,k} = 1,\ \forall k, \\ &\qquad \sum_{k\in\mathcal{K}} a_{n,k} = Y_n,\ \forall n, \end{aligned} \tag{4.2}$$

where Y_n indicates the number of users associated with BS n. Because the sets of BSs and users are both finite, there exists at least one optimal solution. However, obtaining it within a short period of time is not an easy task, because every possible combination of the association variables grows exponentially with the number of BSs multiplied by the number of users. Theoretically, our optimization problem, equation (4.2), is almost identical to that proposed by [28], and this problem is known to be reducible to the three-dimensional matching problem that is *NP-complete*. Hence, algorithms for finding the optimal solution in polynomial time do not exist unless *P=NP*. For this reason, I introduce heuristic real-time association schemes in the next section.

4.4 Decentralized User Association Algorithm

4.4.1 Overview

A key feature of the proposed schemes is expressed by estimating the overall system utility change in a distributed manner. In other words, users themselves are able to predict how their association change will influence not only their own service quality but also the overall network performance. Figure 4.1 illustrates the very basic operations enabling such a process.

At first, each BS periodically broadcasts some system parameters to notify the

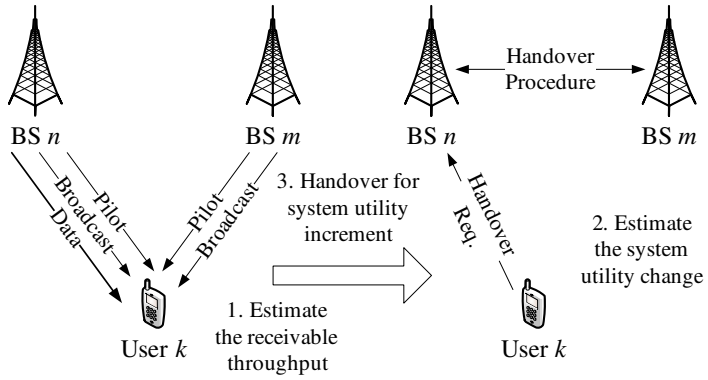


Figure 4.1: Overview of the proposed load-aware handover framework.

users of its current status. By receiving this information and monitoring pilot channels, each user estimates the potentially achievable throughput as well as the corresponding amount of obtainable utility from all the nearby BSs, in case it becomes associated with one of them. At the same time, they also calculate how much utility will be lost by leaving the current serving BS. The reassociation of a user also influences the performance of others connected to the BSs involved in the handover procedure. In detail, those that belong to the target BS will experience service quality degradation because of the increased number of resource-sharing users, whereas those associated with the original serving BS will have an advantage from the departure of one user. For this reason, users should take into consideration every incidental effect before making a handover decision. Unfortunately, there is no way for each user to know its influence against other users. For this reason, such effects are calculated at the BS side, and then should be broadcast to the users.

The detailed design of the algorithms, including throughput estimation, broadcast signal construction, and handover protocol among BSs, are described in the next subsection.

4.4.2 User Scheduling and Throughput Estimation

In the previous chapter, it is investigated how the optimal scheduling algorithm controls every signal transmission in the network (Please refer to the equations (3.4) and (3.5)). This method is also used in this chapter in order to evaluate the upper limit of the system. A similar scheduling process is for a practical system, but a major difference is the set of users that each BS considers to serve. In other words, the optimal scheduling algorithm cares about every user in the vicinity, while the practical scheduler serves only those users actually associated with itself. Thus, in this case, the scheduling decision indicator, from BS n to user k through sub-band s, is given by

$$I_{n,k}^{s}(t) = \begin{cases} 1 & \text{if } k = \arg\max_{k' \in \mathcal{K}_n} \omega_{n,k'}^{s}(t) \\ 0 & \text{otherwise,} \end{cases} \tag{4.3}$$

where the scheduling preference metric is expressed by $\omega_{n,k}^{s} = \frac{r_{n,k}^{s}(t)}{\theta_k(t-1)}$.

For estimating user throughput, this work adopts the generalized version of throughput estimation method, which is expressed by the equation (2.26). By simply expanding the result derived under the single sub-band scenario to the multiple sub-band scenario, the achievable throughput of user $k \in \mathcal{K}_n$, when BS n serves Y users, is represented by

$$\theta_k = \psi_n(Y, \mathbf{R}_{n,k}) = \sum_{s \in \mathcal{S}} \frac{1}{Y} F_{R_{n,k}^{s}}^{-1} \left(\frac{Y}{Y+1} \right), \tag{4.4}$$

where $F_R(\cdot)$ denotes the cumulative density function (CDF) of random variable R. As more users share the same BS, each of them is rarely selected as a scheduling target unless its instant rate approaches approximately the peak level during fluctuation. Therefore, the instant rate of each scheduling target user tends to increase with the number of resource-sharing users.

4.4.3 Broadcast Signal Design

In the remaining parts of this paper, I mainly address the utility, rather than considering the long-term throughput alone. For ease of exposition, I additionally define the utility

of user $k \in \mathcal{K}_n$ as

$$U_{n,k}(Y) = U_k\left(\psi_n(Y, \mathbf{R}_{n,k})\right). \tag{4.5}$$

Because the throughput function decreases with Y and every utility function has increasing property, the composite function decreases as the number of users sharing BS n increases.

A major role of the broadcast signal is to help users estimate the system utility change. On the basis of the following propositions, I can obtain design intuitions.

Proposition 4.1 (A single user handover)**.** Assume user k is currently associated with BS n. If the user changes association to another nearby BS m while other users keep their associations, then the system utility will be changed by

$$\begin{aligned}\Delta U_{k,n\to m} = {} & U_{m,k}(Y_m + 1) - U_{n,k}(Y_n - 1) \\ & + \sum_{i\in\mathcal{K}_m} \left[U_{m,i}(Y_m + 1) - U_{m,i}(Y_m)\right] \\ & + \sum_{i\in\mathcal{K}_n} \left[U_{n,i}(Y_n - 1) - U_{n,i}(Y_n)\right].\end{aligned} \tag{4.6}$$

Obviously, sets of $\mathcal{K}_m$ and $\mathcal{K}_n$ will be changed when the handover of user k is completed. In detail, if time index terms t and $t+1$ respectively denote states before and after the handover, the relationship between notations in (4.6) can be expressed by $\mathcal{K}_m(t+1) = \mathcal{K}_m(t) \cup \{k\}$, $\mathcal{K}_n(t+1) = \mathcal{K}_n(t) - \{k\}$, $Y_m(t+1) = Y_m(t) + 1$, and $Y_n(t+1) = Y_n(t) - 1$. However, for ease of exposition, I omit the time index term and make every notation indicate the state before the handover (e.g., $\mathcal{K}_m$=$\mathcal{K}_m(t)$), because equation (4.6) will be used to predict the system utility change before an actual handover event occurs.

Proof. The derivation of the above equation is merely rearrangement work, and the proof is shown in Appendix A. □

Proposition 4.2 (A new user arrival)**.** Assume recently arrived user k becomes associated with BS n. Then, the system utility is changed by

$$\Delta U_{k,new\to n} = U_{n,k}(Y_n + 1) + \sum_{i\in\mathcal{K}_n} \left[U_{n,i}(Y_n + 1) - U_{n,i}(Y_n)\right]. \tag{4.7}$$

Like the previous case, every notation indicates the state before the association of user k is completed.

Proof. Proposition 4.2 is a special case of Proposition 4.1. □

Let us examine equation (4.6) line by line. The first line represents the utility change of user k because of the handover, and it can be calculated at the user side as long as the number of active users at BS n and m is provided. The second line represents the overall effect on the users belonging to the target BS as user k is incoming, whereas the last line indicates the sum of utility changes of users at BS n as one user is leaving. The second and third lines cannot be calculated in a distributed manner, because there is no way for each user to know the other users' channel statuses. In contrast, each serving BS has enough information from users' CQI report to calculate these two lines.

On the basis of Propositions 4.1 and 4.2, I can define three basic parameters that should be broadcast: One for the user number indicator (UNI) and two versions of the utility difference indicator, named UDI1 and UDI2. As the name suggests, UNI simply indicates the number of active users at the BS. In addition, UDI1 and UDI2 directly refer to the second and the third line of equation (4.6), and these parameters with respect to BS n are respectively denoted by

$$\begin{aligned}\Delta U^{(n,1)} &= \sum_{i\in\mathcal{K}_n} \left[U_{n,i}(Y_n) - U_{n,i}(Y_n + 1)\right], \\ \Delta U^{(n,2)} &= \sum_{i\in\mathcal{K}_n} \left[U_{n,i}(Y_n - 1) - U_{n,i}(Y_n)\right].\end{aligned} \tag{4.8}$$

Both UDI1 and UDI2 are defined to be always positive.

Assume user k is currently associated with BS n. If there is a nearby target BS that makes the equation of Proposition 4.1 positive, then changing the association will cause an increase in system utility. Therefore, a very primitive way to define the handover condition is introduced as follows.

Definition (Original handover condition)**.** The handover condition for user k from the present serving BS n to another target BS m is defined by

$$U_{m,k}(Y_m + 1) - U_{n,k}(Y_n - 1) - \Delta U^{(m,1)} + \Delta U^{(n,2)} > 0. \tag{4.9}$$

The most important thing to consider in developing a handover algorithm is the stability of the system. Here is an example of an unstable scenario: If there exists an under-loaded BS surrounded by many congested BSs, many users may try to change their association from the crowded ones to the less crowded one simultaneously. In that event, the original under-loaded BS becomes congested, and some users may try to return their previous serving BS again. To prevent such an oscillation phenomenon, the following handover procedure is suggested.

PROCEDURE (LOAD-AWARE HANDOVER)

Step 1. Each BS calculates UDI1 and UDI2 and transmits these parameters with UNI through the broadcast channel.

Step 2. (For active users) On the basis of the handover condition, users search a potential handover target among all the BSs in the vicinity. If there exists at least one target BS satisfying the handover condition, they inform their serving BS where a reassociation to the best one among all the candidates is required. At the same time, users also report how much the system utility will be changed because of their handover.

Step 2. (For recently arrived users) On the basis of Proposition 4.2, the newly arrived users calculate the system utility change in case they will be associated with one of the nearby BSs. Then, they directly send the association request to the best one that provides the maximum calculation result.

Step 3. Once serving BSs receive the handover requests, they select only the one user that expects to contribute the maximum utility increase when the reassociation is completed. Then, they transfer the request of the selected user to the corresponding target BS in company with the expected value of the utility change.

Step 4. For purposes of stable operation, each target BS also accepts only one user according to the following order. The recently arrived user has the highest priority, and the active user expecting the maximum utility increase is next.

Step 5. Do step 1 to 4 periodically (e.g., every 100 ms).

END PROCEDURE

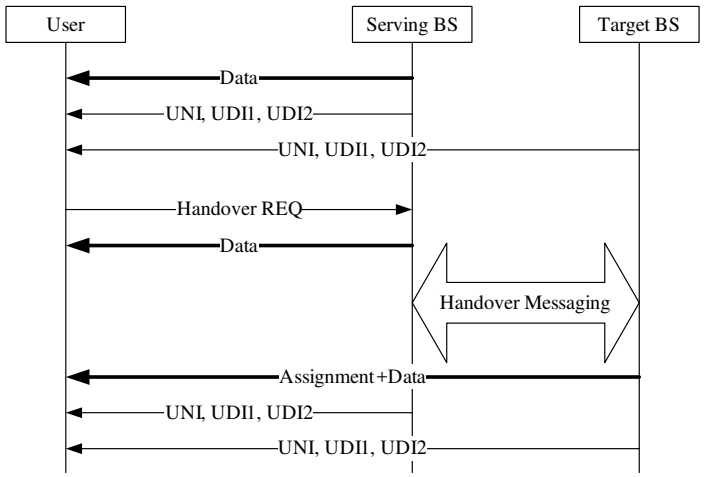


Figure 4.2: Flow chart of the proposed load-aware handover procedure.

Figure 4.2 illustrates the flow chart of the proposed framework. Even though I struggled to restrict the number of handover events per BS to be less than one to reduce the oscillation phenomenon, a few unexpected scenarios might occur. For instance, a user changes association from one to the other BS, while another user does so exactly in the opposite direction. In this case, the number of active users at every BS remains the same, and consequently, the estimation based on Proposition 4.1 may bring a wrong result. Another scenario is that a user changes association from one BS to another, while the other user changes association from another to the other BS. To suppress

such unexpected scenarios, the original handover condition needs to be reinforced.

Definition (Modified handover condition)**.** I improve the handover condition for user k, from BS n to m, as follows.

$$\begin{aligned} & U_{m,k}(Y_m + 1) - U_{n,k}(Y_n - 1) \\ & \quad - \max(\Delta U^{(m,1)}, \Delta U^{(m,2)}) \\ & \quad + \min(\Delta U^{(n,1)}, \Delta U^{(n,2)}) > \log h_{th}, \end{aligned} \tag{4.10}$$

where $h_{th} \geq 1$ implies handover threshold.

Proposition 4.3 (Handover with the modified condition)**.** If the original handover condition is replaced with the modified version, the proposed load-aware handover procedure always guarantees a system utility increase.

Proof. The proof is represented in Appendix B. □

According to the above proposition, the proposed handover procedure with the modified condition works well as I intended, because the system utility is always increasing unless the active users terminate their wireless access. In fact, Proposition 4.3 does not imply that the global optimal solution is obtained. Nevertheless, the potential of the proposed scheme is that it achieves the maximum system utility rise for every iteration, because it accepts the best request among conflicting handover requests. Numerical investigation will determine the proximity with which the proposed method attains the upper limit of the system.

The role of the handover threshold is to provide more stability, where assigning $h_{th} = a$ makes users hesitate to change their association until the obtainable throughput from the target BS becomes a times from the baseline scenario (i.e., $h_{th} = 1$). However, if the threshold level is set too high, users cannot quickly respond to the variation of the network, and the overall performance will decrease. The tradeoff relationship between the handover threshold and the system performance will be discussed in the performance evaluation section.

4.5 Fully Decentralized Algorithm

In the previous section, I proposed a load-aware handover procedure, where users autonomously choose the best BS among all the nearby candidates. The proposed scheme does not require additional feedback for load balancing purposes, and the backhaul communication is used only for carrying handover requests as the conventional system does. Now I slightly modify the proposed scheme, so that users directly send handover requests to the target BS, instead of transferring it via their serving BS. Such an assumption is indeed preferable for various scenarios. For instance, using this algorithm, users can freely alternate their wireless access networks among totally different generations of cellular systems, such as 3G, 4G, and even future wireless systems.

PROCEDURE (DIRECT HANDOVER REQUEST)

Step 1. Each BS transmits UNI, UDI1, and UDI2 through the broadcast channel.

Step 2. (For active users) On the basis of the modified handover condition and predetermined handover threshold h_{th}, users search nearby target BSs satisfying the condition. If there exists at least one candidate for reassociation, they directly send the handover request message to the best one providing the maximum performance increase.

Step 2. (For recently arrived users) On the basis of Proposition 4.2, the newly arrived users calculate the system utility difference against every BS in the vicinity, and also directly request the association to the best one.

Step 3. Each BS then accepts every incoming users.

Step 4. Do step 1 to 3 periodically.

END PROCEDURE

Unlike the load-aware handover procedure introduced in the previous section, many users can change their associations simultaneously, due to the lack of an intermediate coordinator. For this reason, I control the number of handover events by choosing an appropriate handover threshold.

4.6 Performance Evaluation and Discussion

4.6.1 Simulation Setup

In the experiments, two specific types of BSs, macro and femto BSs, are used. Figure 4.3 illustrates the baseline deployment scenario, where seven sites, each with three sectors, are placed on a hexagonal grid.

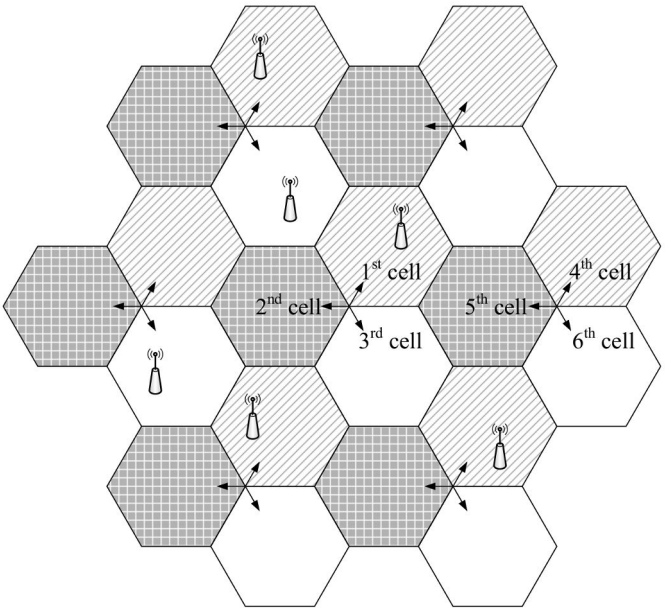


Figure 4.3: The baseline cell deployment scenario, where three different types of macro cells share the same site.

Even though sectors sharing the same site are actually covered by a single set of equipment, I assume that the three separate macro BSs are deployed at each site, because user scheduling and association control for each sector are essentially executed in an independent manner. For this reason, I use the terms cell and sector interchangeably. The wrap-around technique is adopted in the original cell layout to remove all border effects. To distinguish the BS type, I define $\mathcal{N}_{macro} = \{1, ..., 21\} \subset \mathcal{N}$ as the set of macro BSs, and $\mathcal{N} - \mathcal{N}_{macro}$ as the set of femto BSs. Index $n \in \mathcal{N}_{macro}$ is

also used to indicate the cell covered by the corresponding macro BS. For interference management purposes, I simply apply the FFR scheme, where every macro BS shares 12 sub-bands, and the remaining 12 sub-bands are assigned to each cell with a reuse factor of 1/3. Moreover, to mitigate inter-tier interference, femto BSs are allowed to use the parts of the sub-bands that are not used in the macro cell where they belong. The major parameters are summarized in Table 4.1.

To generate realistic traffic activities, a dynamic user arrival and departure model is used during our simulation. For a given $n \in \mathcal{N}_{macro}$, I define $\rho_n = \lambda_n/\mu$ as the traffic intensity at the n-th cell, where λ_n represents the Poisson arrival rate of users in this cell, and $1/\mu = 60$ seconds is an exponentially distributed stay time for every new arriving user. After users are located in an arbitrary position within the cell, they begin to move in a random direction while constantly generating data traffic. According to this model, the number of active users within the n-th cell fluctuates around ρ_n. Therefore, I can plausibly make various loading scenarios, including a highly unbalanced case, by assigning this value to each cell differently.

The performance of the proposed user association schemes is verified with extensive simulations under diverse environments. To investigate the data offloading capability of small cell BSs, I first consider an unbalanced loading scenario. In this case, I forcibly make an arbitrary cell congested and deploy a few femto BSs only within this cell. Another possible scenario is that the same number of femto BSs is installed in each cell by assuming the traffic intensity at every cell is identical. Finally, I also consider a realistic scenario where the traffic intensity of each cell varies over time.

4.6.2 Unbalanced Traffic Intensity

For an unbalanced scenario, the first cell is selected as a potential candidate for being congested and the traffic intensity for this cell is changed from 30 to 120, while that for other cells is fixed at 30. To distribute the load at the congested macro BS, 10 femto BSs are deployed over the first cell. Throughout the figures in this chapter,

Table 4.1: Simulation parameters

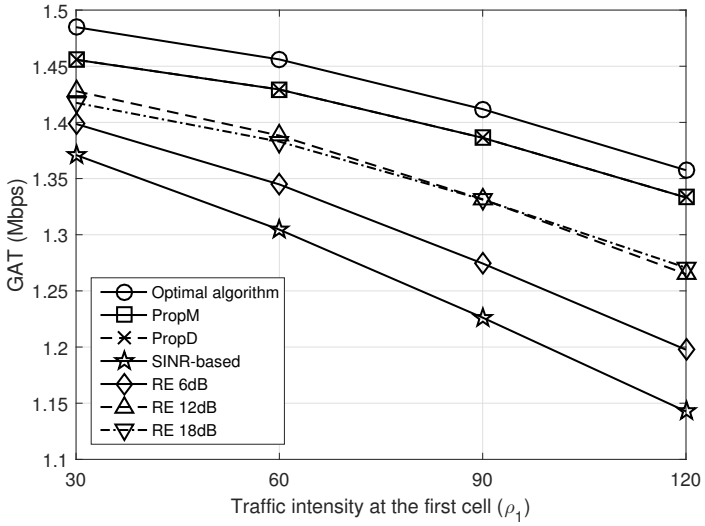
Parameter	**Assumption**
Cell Layout	7 sites each with 3 sectors
Inter-Site Distance	1,000 m
Carrier Frequency	2,000 MHz
Sub-Band Bandwidth	360 kHz
Number of Sub-Bands	24
Pathloss Model (d in km)	$P_L = 128.1 + 37.6 \log_{10}(d)$ (macro) $P_L = 136.74 + 39.2 \log_{10}(d)$ (femto)
Shadowing	Log-Normal dist. with 8 dB Std. (macro) 10 dB Std. (femto)
BS Transmit Power	46 dBm (macro), 46 dBm (femto)
Antenna Configuration	2x2 MIMO, ZF receiver
Antenna Gain	BS: 5 dBi, User: 0 dBi
Scheduler	Proportional Fair
User Noise Figure	9 dB
Thermal Noise	-114 dBm/MHz
Channel Model	Typical Urban
User Velocity	3 Km/h

PropM and *PropD* respectively refer to the proposed handover scheme with the modified handover condition and its fully decentralized version. Also, *SINR-based* means the traditional signal strength based cell-site selection rule, whereas *RE 6 dB*, *12 dB*, and *18 dB* represent the range expansion methods with the corresponding bias levels.

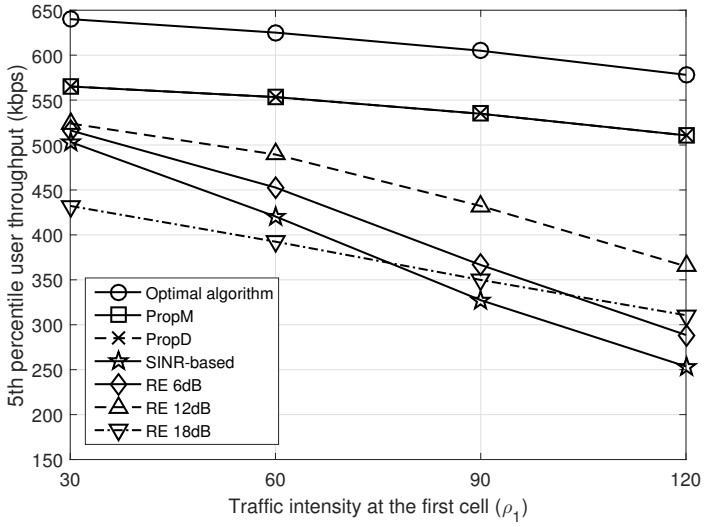
Figure 4.4(a) shows that both proposed schemes surpass the conventional schemes, regardless of how severely crowed the first cell is. The GAT gain is measured as $1.9 - 5.1\%$ from the best conventional scheme (i.e., RE 12 dB). This is not a trivial result, because it is equivalent to an increase in the throughput of every user in the network by that amount. Moreover, under such an unbalanced loading scenario, the proposed schemes pay special attention to the users at the congested cell that have relatively low service quality. Therefore, most users associated with the under-loaded BSs are not highly affected whether the load balancing algorithm is applied or not. This is the reason why the GAT metric does not show an impressive result.

On the other hand, the average throughput of the 5th percentile users, depicted in Figure 4.4(b), provides a more interesting result. Like the previous result, both of the two proposed algorithms surpass the conventional schemes, while providing a $7.3 - 29\%$ performance gain over the best conventional one (i.e., RE 12 dB). Another important observation from both graphes in Figure 4.4 is the performance gap between the optimal algorithm and the proposed schemes. This can be explained by the multi-user scheduling gain. When the optimal algorithm runs, each BS can take full advantage of multi-user diversity by serving every user in the vicinity. However, in practical systems, each BS takes care of only the limited set of users associated with itself, and the achievable diversity order becomes low. Hence, the performance gap always exists even if users are properly assigned to the BSs.

Figure 4.5 compares the average number of users at the congested, under-loaded, and femto BSs. The traffic intensity at the ordinary cells is fixed by 30, and that for the congested cell is assumed to be 60 for Figure 4.5(a) and 90 for Figure4.5(b). In both cases, 10 femto BSs are deployed at the congested cell for the purpose of traffic

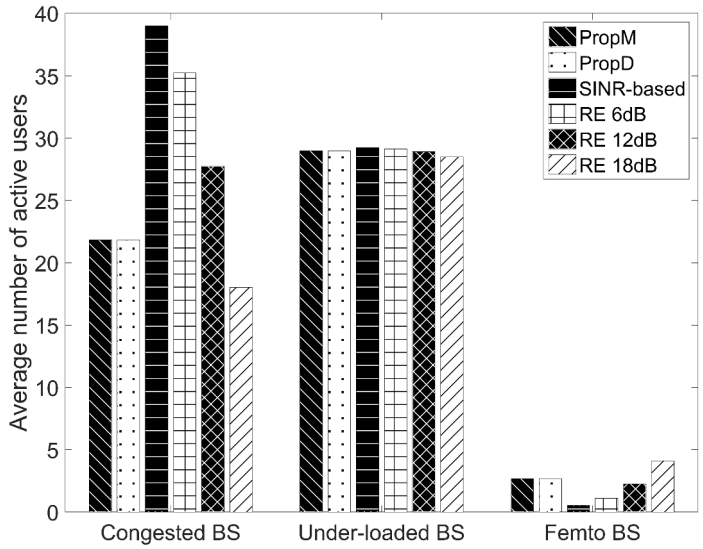


(a)

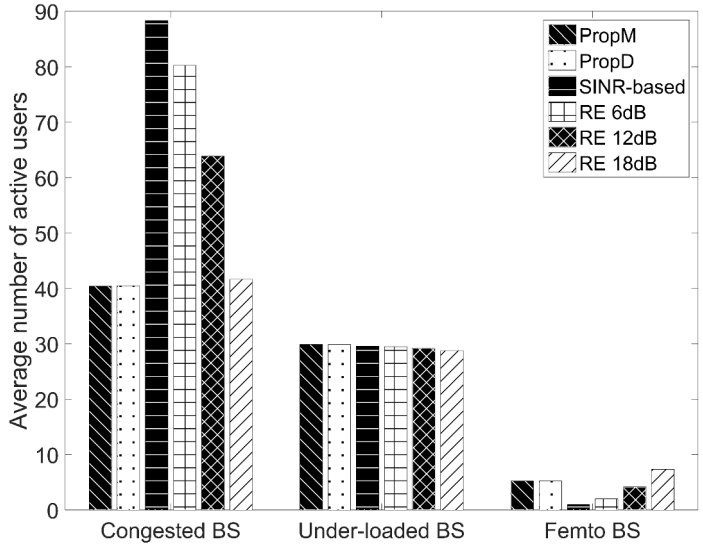


(b)

Figure 4.4: System performance versus traffic intensity at the congested cell: (a) Geometric average of throughput (GAT), (b) Average 5th percentile user throughput. The traffic intensity at other macro cells is assumed to be 30 (i.e., $\rho_n = 30, \forall n \neq 1$).



(a)



(b)

Figure 4.5: Average number of active users at each type of BS according to traffic intensity of the congested cell: (a) $\rho_1 = 60$, $\rho_n = 30, \forall n \neq 1$ (b) $\rho_1 = 90$, $\rho_n = 30, \forall n \neq 1$. 10 femto BSs are deployed at the congested cell.

offloading. Under the SINR-based cell selection scheme, almost every user prefers the macro BS. As the offset for signal strength from femto BSs increases, users begin to be associated with them. The proposed schemes also undertake a similar operation, but the only difference is that they distribute users at the congested BS to femto BSs and neighboring under-loaded BSs as well.

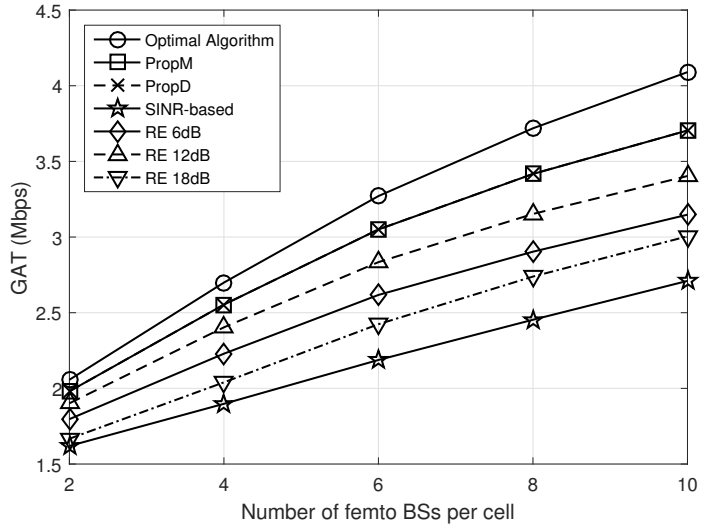
4.6.3 Equal Traffic Intensity

To simulate an equal loading scenario, I assign the traffic intensity at every cell at 30 (i.e., $\rho_n = 30, \forall n \in \mathcal{N}_{macro}$), and then deploy the same number of femto BSs per cell.

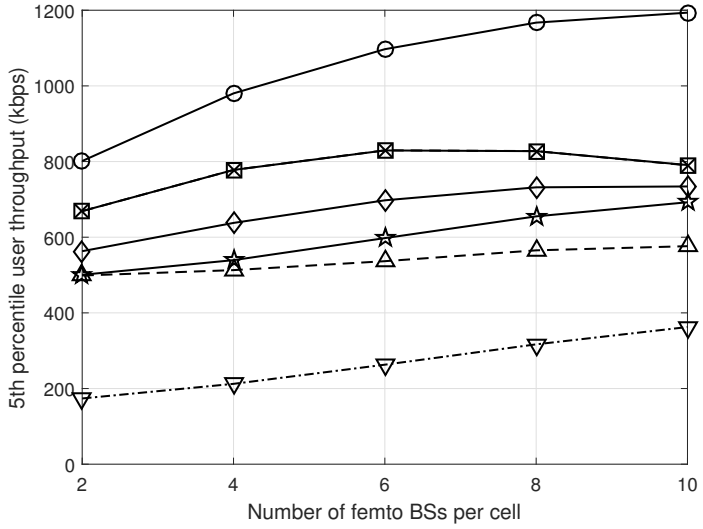
Figure 4.6(a) shows the GAT performance, where the proposed schemes attain $3.7 - 9.4\%$ of the upper limit, while achieving a $4.2 - 8.8\%$ gain over the best conventional scheme (i.e., RE 12 dB). Like the unbalanced loading case, there is a performance gap between the optimal algorithm and the proposed schemes. And the gap is increased with the number of femto BSs per cell. Similar result is observed in Figure 4.6(b). Even though the proposed schemes outperforms the best conventional scheme (i.e., RE 6dB) by $7.7 - 21\%$, the gap from the optimal algorithm is widening with the number of femto BSs.

Such performance gaps between the optimal scheduling algorithm and the proposed scheme can be explained by multi-user diversity gain. When the optimal scheduling algorithm runs, each BS fully exploits the multi-user diversity regardless of how many femto BSs are deployed. However, as Figure 4.7(a) shows, the number of active users at each BS decreases with the number of femto BSs, and thus the achievable scheduling gain is reduced accordingly. This is the reason why the performance gap between the optimal algorithm and the proposed schemes widens with the number of deployed femto BSs.

In Figure 4.7(a), femto BSs that do not serve any user are excluded. Instead, the average percentage of femto BSs in an idle mode is shown in Fig. 4.7(b). Under the traditional signal strength based cell-site selection method, users rarely try to connect

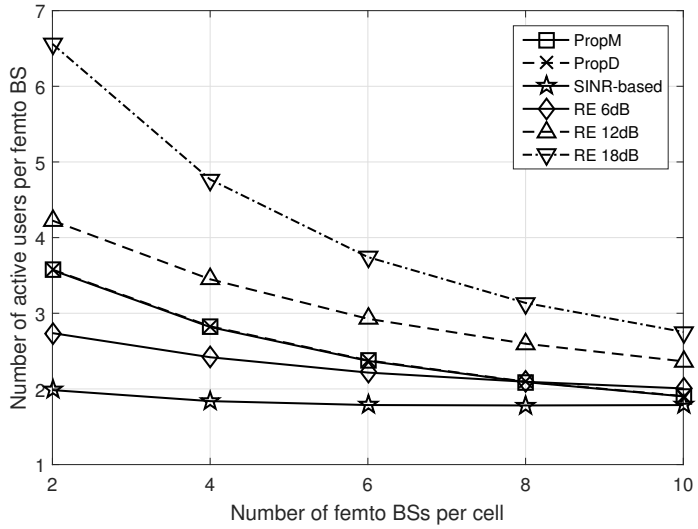


(a)

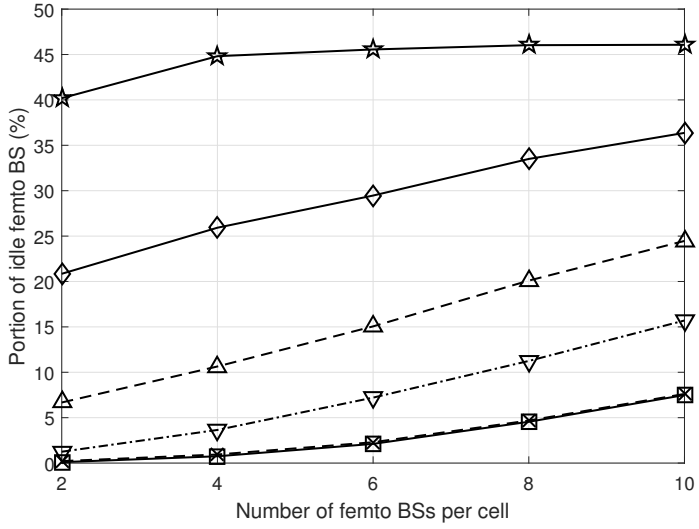


(b)

Figure 4.6: System performance versus the number of femto BSs per macro cell: (a) Geometric average of throughput (GAT), (b) Average 5th percentile user throughput. The traffic intensity at every macro cell is assumed to be 30.



(a)



(b)

Figure 4.7: System performance for the equally loading scenario (i.e., $\rho_n = 30, \forall n \in \mathcal{N}_{macro}$): (a) Number of active users per femto BS, (b) Average percentage of femto BSs in an idle mode.

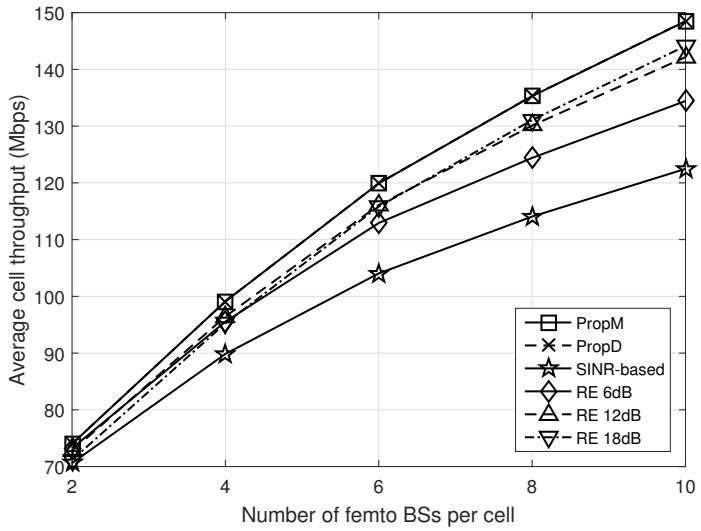
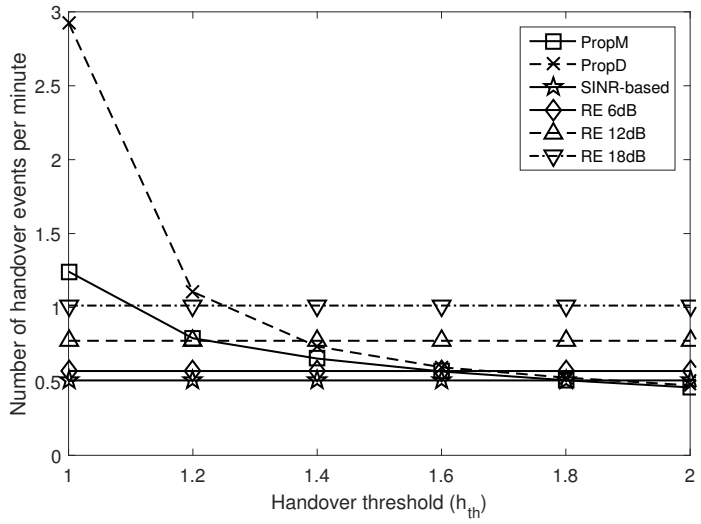


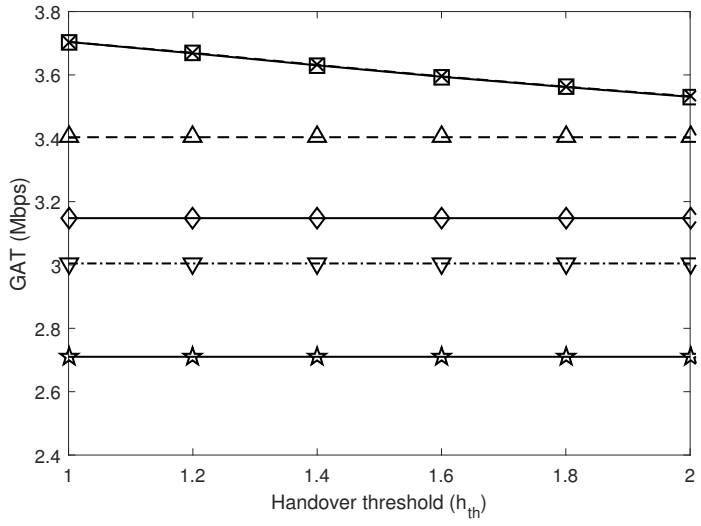
Figure 4.8: Average cell throughput for the equally loading scenario (i.e., $\rho_n = 30, \forall n \in \mathcal{N}_{macro}$).

to the femto BS, because it provides relatively low signal quality. In this case, almost $40 - 45\%$ of the femto BSs are in an idle mode, and even the rest of them serve only a small number of users. When the biasing method is applied, each femto BS begins to attract more users, making them virtually experience better signal quality. For this reason, the number of active users at each femto BS increases with the bias level, while the percentage of idle-state femto BSs decreases as well. The proposed schemes also offload users into femto BSs effectively, in order to reduce the idle mode operation.

The average cell throughput, which is obtained by dividing the total system throughput by the number of macro cells, is also an important performance metric. However, optimizing the system throughput is not the objective of this study. Furthermore, cell throughput is sometimes sacrificed to take care of users with low service quality. Nevertheless, Figure 4.8 shows that the proposed algorithms outperform other existing schemes. This is because they effectively utilize the pre-installed infrastructures so that the percentage in the idle mode is minimized. In any case, the cell throughput increases with the number of deployed femto BSs.



(a)



(b)

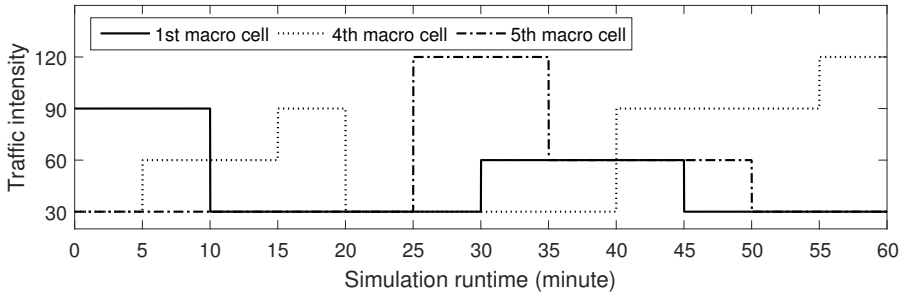
Figure 4.9: Two opposite system performances versus the handover threshold level: (a) Geometric average of throughput (GAT), (b) Number of per-user handover events per minute.

Another important feature that should be observed is the impact of the handover threshold. Figure 4.9 illustrates two opposite system performances versus various threshold values. In these cases, the traffic intensity at every cell is set as before, and 10 femto BSs are deployed per cell. The more strict the handover threshold level, the more users hesitate to change the association. Therefore, they could not quickly respond to environment changes, and this is naturally followed by a degradation in GAT performance. The positive effect is related with the system stability. As the threshold level increases, the per-user average number of handover events within a unit time interval is decreasing. Interestingly, Figure 4.9(b) shows that the original SINR-based scheme is the most stable, but this is because users rarely change their association to femto BSs.

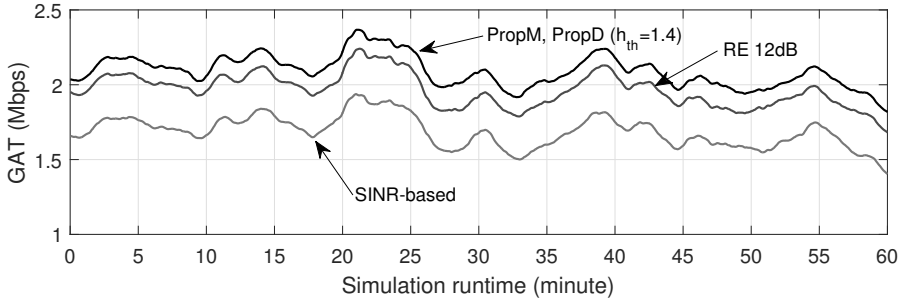
4.6.4 Dynamic Scenarios

I arbitrarily choose the first, fifth and sixth cells as candidates for the dynamic setting. Figure 4.10(a) illustrates our scenario, where the traffic intensity for the selected cells changes over time, while that for the remaining cells is fixed at 30. As the x-axis of the graph indicates, the total simulation runtime is extended to 1 h in this case. To reflect a realistic deployment aspect, a random number of femto BSs up to 10 is installed over each cell. Through the above process, a total of 109 femto BSs are deployed. For a clear performance comparison, the results for only the following algorithms are shown. Among various handover threshold levels for the proposed schemes, I choose $h_{th} = 1.4$, because it provides a reasonable balance between GAT performance and stability. The range expansion method with a bias level of 12 dB and the original SINR-based cell-site selection rule are also included.

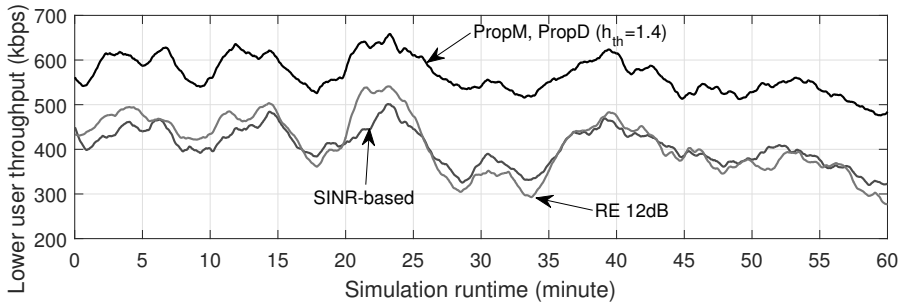
First, GAT performance is represented in Figure 4.10(b). Throughout the simulation runtime, the proposed schemes always outperform the best conventional scheme by $2.8-12\%$ with an average of 6.5%. Like the previous unbalanced loading scenarios, this result is meaningful, because it is equivalent to every user in the network equally experiencing that amount of performance enhancement. The average throughput of the



(a)

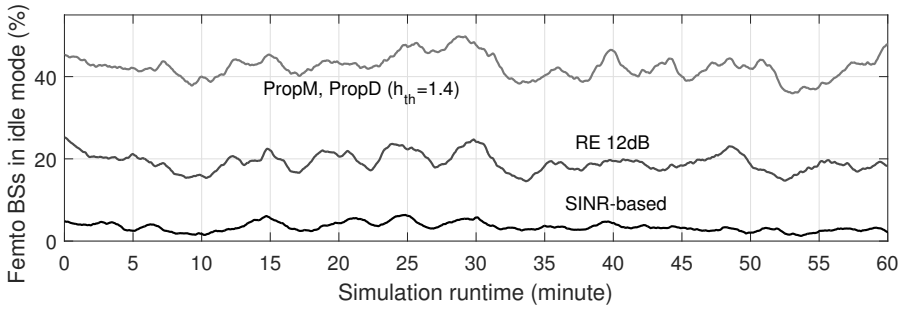


(b)

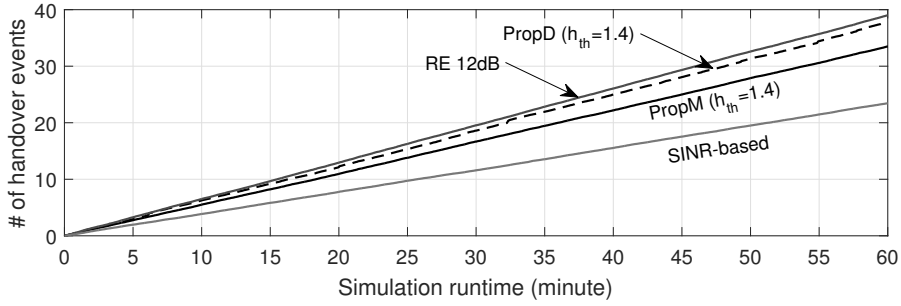


(c)

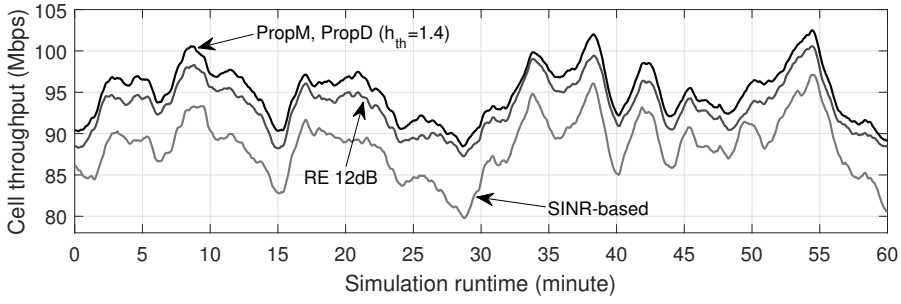
Figure 4.10: System performance under a dynamic setting: (a) Dynamic traffic load scenario, (b) Geometric average of throughput (GAT), (c) Average 5th percentile user throughput



(a)



(b)



(c)

Figure 4.11: System performance under the dynamic setting: (a) Average percentage of femto BSs in an idle mode, (b) Accumulated number of per-user handover events, (c) Average cell throughput.

5th percentile users, depicted in Figure 4.10(c), represents a more impressive result. Similar to GAT performance, the proposed schemes always provide superior results compared to the other existing schemes, where the performance gain from the range expansion scheme is $5.0 - 85\%$ with an average of 41%, and that from the original SINR-based scheme is $11 - 119\%$ with an average of 40%.

Compared to its remarkable GAT performance, the range expansion scheme does not play a meaningful role for improving system utility, because it simply offloads users into femto BSs by biasing the signal qualities from them without considering any form of utility. In contrast, our schemes make users consider the system utility, which is designed so that users with low service quality will be protected. Another remarkable effect is shown in Figure 4.11(a). Even though the range expansion scheme offloads users into the femto BSs, on the average, $15 - 30\%$ of the femto BSs do not serve any user. In contrast, the proposed schemes effectively utilize pre-installed infrastructure, where the average percentage of idle-state femto BS is almost below 10%.

Figure 4.11(b) represents the accumulated number of handover events per user. As the graph shows, the SINR-based cell-site selection rule provides the best stability, where each users in the network experiences an average of 22 association changes during 1 hour. However, under the original association method, users rarely connect to femto BSs, and almost every association change is due to the inter-cell handover. Finally, the average cell throughput performance is represented in Figure 4.11(c). The graph shows that the proposed schemes outperform other existing schemes in almost everywhere.

Chapter 5

JOINT OPTIMIZATION OF USER ASSOCIATION & INTER-TIER INTERFERENCE MANAGEMENT IN HETEROGENEOUS CELLULAR NETWORKS

5.1 Motivation

From the previous chapter, it is studied that the user association is an important issue to improve the overall system performance. For instance, in heterogeneous cellular networks, users associated with the macro BSs are sometimes forcibly redirected into nearby small cell BSs. In that case, such users may suffers from low services quality because the user association algorithm generally consider the overall performance rather than each individual user's performance. Moreover, the macro BSs radiate much more transmission power than small cell BSs, inter-tier interference is a major factor to degrade the performance of users at the small cell BSs.

To guarantee the performance of users, especially those forcibly offloaded from the macro BS, enhanced inter-cell interference coordination (eICIC) scheme was introduced [50–52]. In this scheme, macro BSs periodically mute data transmission for some subframes, which is called almost blank subframes (ABS), in order to remove inter-tier interference. Several attempts have been made to minimize the inter-tier inter-

ference [53, 54]. These schemes try to find an optimal ABS portion in order to achieve the maximum system performance.

This chapter studies a joint optimization problem for both user association and inter-tier interference management. Like previous chapters, the time-varying nature of wireless propagation channels is considered during the development of the algorithm. To this end, the generalized version of throughput estimation method will be used. The proposed algorithm is divided into two parts, where the first part try to change the ABS portion dynamically in direction of improving the overall performance. The second part also attempts to associate BSs and users to achieve the same purpose.

5.2 System Model

I consider a heterogeneous cellular system, where several pico BSs overlie in the macro cell coverage. Sets of macro and pico BSs are respectively denoted by $\mathcal{N}_m$ and $\mathcal{N}_p$, and then the total set of BS is given by $\mathcal{N} = \mathcal{N}_m \cup \mathcal{N}_p = \{1, ..., N\}$. A set of users is also defined by $\mathcal{K} = \{1, ..., K\}$. Open access operation is assumed, thus, users can be associated with any types of BSs but not more than one simultaneously. For the purpose of wireless resource allocation, the system frequency band is equally partitioned into a set of subbands $\mathcal{S} = \{1, ..., S\}$. In addition, time is divided into a set of subframes $\mathcal{T} = \{1, 2, ...\}$, for the same purpose. A frequency subband during a single subframe will constitute a fundamental unit of data transmission. Essentially, I do not confine this work to a single antenna system, but assume a single user scheduling scenario, where a subband should be allocated to a single target user at each subframe.

The fractional frequency reuse (FFR) scheme provides an effective way to mitigate inter-cell and inter-tier interference, and it can therefore be applied in this study. However, the optimization of FFR patterns is beyond the scope of this work. Instead, I will mainly focus on a time-domain interference management method, which works well with an arbitrary FFR pattern. To this end, macro BSs periodically mute the data

transmission to remove the interference effecting pico BSs. Such an operation is synchronized among all macro BSs, and the portion of ABSs is denoted by α. It is essential to distinguish two different types of subframes: the set of almost blank ones $\mathcal{T}_{ABS}$ and that of regular ones $\mathcal{T}_{Regular}$.

For a specific subframe $t \in \mathcal{T}(= \mathcal{T}_{ABS} \cup \mathcal{T}_{Regular})$, I define $r^s_{n,k}(t)$ as the instantaneous data rate from the BS n to the user k through the subband s. If only a single data stream is transmitted to a user, then this value is directly related to the signal to interference-plus-noise ratio. If multiple data streams are simultaneously transmitted to a user, then the instant rate is expressed by the sum of the rates of all individual streams.

One important issue regarding realistic signal propagation is the multipath fading phenomenon. When several rays with different propagation paths are superposed at a receiver, the signal strength widely fluctuates, even with a slight movement of the receiver. To reflect this time-varying property, the collection of instantaneous data rates $\{r^s_{n,k}(t), t = 1, ...\}$ can be considered as a stochastic process that is characterized by two different random variables. Specifically, $r^s_{n,k}(t)$ is assumed to be generated from random variables $R^s_{n,k,ABS}$ when $t \in \mathcal{T}_{ABS}$, and $R^s_{n,k,Regular}$ when $t \in \mathcal{T}_{Regular}$.

5.3 Problem Formulation

The objective of this work is to optimize network-wide utility, which is defined as $U = \sum_{k \in \mathcal{K}} U_k(\bar{\theta}_k)$. Here, $\bar{\theta}_k$ represents the long-term average throughput of user k, and $U_k(\cdot)$ is a utility function for the corresponding user. Among the various candidates for utility functions, I select log utility (i.e., $U_k(\bar{\theta}_k) = \log(\bar{\theta}_k),\ \forall k \in \mathcal{K}$) because it pays more attention to users with a low service quality, and is therefore adequate for the purpose of achieving fairness among users.

In order to achieve the maximum possible network-wide utility, BSs at each subframe should serve users that contribute the maximum utility increase when being

served. According to this scheduling policy, the scheduling target user k^* for BS n, with subband s and subframe t, is selected by making the following comparison:

$$k^* = \arg\max_{k \in \mathcal{K}_n} \frac{r^s_{n,k}(t)}{\bar{\theta}_k}, \tag{5.1}$$

where $\mathcal{K}_n$ is the set of users associated with the BS n. Equation (5.1) implies that a user whose instant rate approaches the peak level during a fluctuation has a high priority for scheduling. As a result of this mechanism, the cell throughput increases according to the number of resource sharing users, and this phenomenon is termed multi-user diversity gain.

To guarantee the performance for users at pico BSs, some studies have attempted to define a concept of victim users, who suffer from severe inter-tier interference, and thus exclusively utilize ABSs [53]. However, by performing system-level simulations, I notice that partitioning the set of users at pico BSs into two subsets degrades the multi-user diversity gain.

For this reason, I assume that pico BSs allocate both ABSs and regular subframes to users without any discrimination. By modifying the throughput estimation method described in Chapter 2, the long-term throughput for a user k, who is currently associated with a BS n, can be expressed by

$$\begin{aligned}\bar{\theta}_k &= \psi_n(\alpha, Y, \mathbf{R}_{n,k,ABS}, \mathbf{R}_{n,k,Regular}) \\ &= \alpha \sum_{s \in \mathcal{S}} \frac{1}{Y} F^{-1}_{R^s_{n,k,ABS}} \left(\frac{Y}{Y+1} \right) \\ &\quad + (1-\alpha) \sum_{s \in \mathcal{S}} \frac{1}{Y} F^{-1}_{R^s_{n,k,Regular}} \left(\frac{Y}{Y+1} \right),\end{aligned} \tag{5.2}$$

where random variables are given by $\mathbf{R}_{n,k,Regular} = [R^1_{n,k,Regular}, ..., R^S_{n,k,Regular}]$ and $\mathbf{R}_{n,k,ABS} = [R^1_{n,k,ABS}, ..., R^S_{n,k,ABS}]$. $F_R(\cdot)$ is the cumulative density function (CDF) of the random variable R, and Y implies the number users at BS n. The terms multiplied with α and $(1-\alpha)$ respectively denote the expected amount of throughput through ABSs and regular subframes.

I define $a_{n,k}$ as the association variable between the BS n and the user k, where a value of $a_{n,k} = 1$ implies that the association between them is established. Using (5.2), the maximization of network-wide utility can be formulated as the following optimization problem:

$$\begin{aligned} \max \sum_{n\in\mathcal{N}} \sum_{k\in\mathcal{K}_n} & U_k(\psi_n(\alpha, Y_n, \mathbf{R}_{n,k,ABS}, \mathbf{R}_{n,k,Regular})) \\ \text{subject to} \quad & 0 \le \alpha \le 1, \\ & a_{n,k} \in \{0,1\},\ \forall n,k, \\ & \sum_{n\in\mathcal{N}} a_{n,k} = 1,\ \forall k, \\ & \sum_{k\in\mathcal{K}} a_{n,k} = Y_n,\ \forall n, \end{aligned} \tag{5.3}$$

where $\mathcal{K}_n = \{k \in \mathcal{K} | a_{n,k} = 1\}$, and the constraint condition $\sum_{n\in\mathcal{N}} a_{n,k} = 1$ follows from the single user scheduling assumption. Theoretically, this optimization problem is known to be *NP-complex*, even for a fixed α [35]. Therefore, finding an optimal solution within a polynomial time may not be possible.

5.4 Joint Optimization Algorithm

For user association, the proposed throughput estimation based user association algorithm shown in the previous chapter is applied in this chapter. The only change is the expression of utility, where the utility of user $k \in \mathcal{K}_n$ is expressed by

$$U_{n,k}(Y) = U_k\left(\psi_n(\alpha, Y, \mathbf{R}_{n,k,ABS}, \mathbf{R}_{n,k,Regular})\right). \tag{5.4}$$

Now, I analyze the influence of changing the ABS portion on the system utility. Because a dramatic change of this parameter induces the network to become severely disordered, an iterative algorithm is desirable. For a fixed user association, the instant change of the network-wide utility with respect to a small change of the ABS portion can be represented by

$$\frac{\partial U}{\partial \alpha} = \sum_{k\in\mathcal{K}} \frac{\frac{\partial \bar{\theta}_k}{\partial \alpha}}{\bar{\theta}_k}. \tag{5.5}$$

By using equation (5.2), the partial derivative of the long-term throughput in (5.5) can be derived as

$$\frac{\partial \bar{\theta}_k}{\partial \alpha} = \sum_{s \in \mathcal{S}} \frac{1}{Y} F^{-1}_{R^s_{n,k,ABS}} \left(\frac{Y}{Y+1} \right) - \sum_{s \in \mathcal{S}} \frac{1}{Y} F^{-1}_{R^s_{n,k,Regular}} \left(\frac{Y}{Y+1} \right). \tag{5.6}$$

By applying a slight variation of the ABS portion by $\Delta\alpha$, the overall system utility will be approximately changed by $\partial U/\partial \alpha \cdot \Delta\alpha$. Therefore, an algorithm to find an optimal value for α can be developed as follows.

PROCEDURE (ITERATIVE ABS TUNING)

Step 1. BSs divide the calculation work of (5.5) in order that BS n calculates $\partial U_n/\partial \alpha$ with $U_n = \sum_{k \in \mathcal{K}_n} U_k(\bar{\theta}_k)$.

Step 2. A central node collects of all the calculation results, and finally computes $\partial U/\partial \alpha = \sum_{n \in \mathcal{N}} \partial U_n/\partial \alpha$.

Step 3. For a pre-defined value of $\Delta\alpha$, every macro BS in the network simultaneously changes the ABS portion to α^* according to the following conditions:

$$\alpha^* = \begin{cases} \alpha + \Delta\alpha & \text{if } \partial U/\partial \alpha \cdot \Delta\alpha > \log \alpha_{th} \\ \alpha - \Delta\alpha & \text{if } \partial U/\partial \alpha \cdot \Delta\alpha < -\log \alpha_{th} \\ \alpha & \text{otherwise.} \end{cases} \tag{5.7}$$

Step 4. Repeat steps 1 to 3 periodically.

END PROCEDURE

Similar to handover threshold, $\alpha_{th} \geq 1$ represents the threshold level for an ABS portion change, and this condition is essential for securing stability. Once this threshold value is determined, the ABS portion is not changed until the product the throughput of all users is expected to become α_{th} times greater than in the baseline scenario (i.e., $\alpha_{th} = 1$). Based on this intuition, I choose the threshold value for changing the ABS

portion as $\alpha_{th} = a^K$, which the change in network-wide performance is equivalent to the condition of that every user experiences an a-fold enhancement in performance.

5.5 Performance Evaluation and Discussion

5.5.1 Simulation Setup

As a system-level simulation, I consider a baseline deployment scenario, where seven sites, having three sectors each, are placed on a hexagonal grid. I consider each sector as an individual macro cell, and run scheduling and handover processes independently. The warp-around technique is used to remove all border effects. The detailed parameters are summarized in Table I.

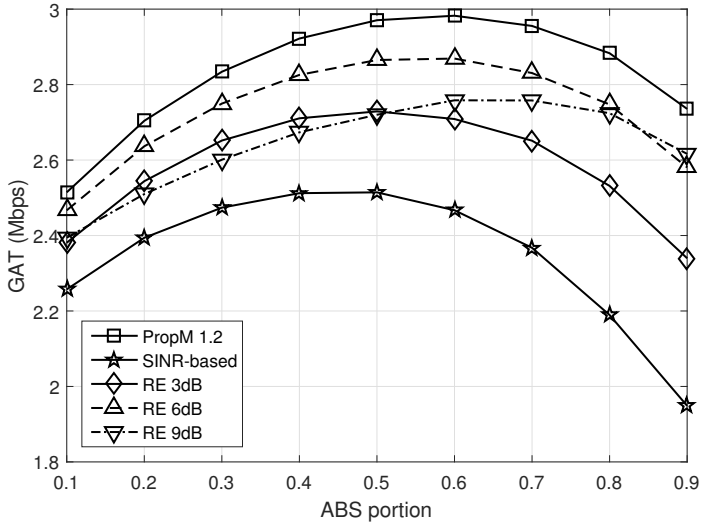
To generate realistic traffic activities within the coverage of a macro BS, I assign $\rho_n = \lambda_n/\mu$ to each BS $n \in \mathcal{N}_m$, where λ_n is the Poisson arrival rate of users in the relevant macro cell, and every new arriving user remains for an exponentially distributed random time, with a mean of $1/\mu = 60$ s. After users have been generated in an arbitrary position, they move in a random direction. Each element in a 2×2 small-scale fading channel is generated using Jake's simulator.

5.5.2 Simulation Results

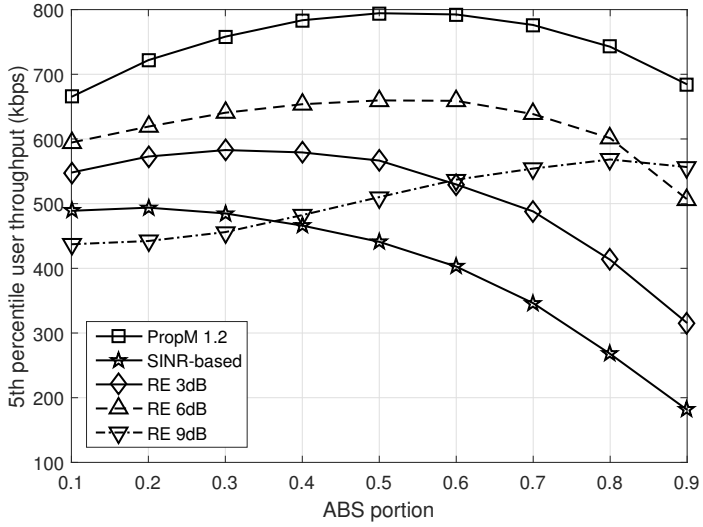
In order to investigate the effect of the eICIC operation, I assign the traffic intensity for every cell to 30, and deploy two pico BSs per cell. Figure 5.1 illustrates the GAT and the average of 5th percentile user throughput when the ABS portions are fixed. Among various results, *PropM 1.2* refers to the proposed user association algorithm with the handover threshold value of $h_{th} = 1.2$. In this scheme, users estimate the throughput based on the data rate variation during the last one second, and the handover procedure is run every 100 ms. The signal strength based BS selection scheme (*SINR-based*) and range expansion methods by using different bias levels (i.e., *RE 3dB, 6dB, 9dB*) are also depicted.

Table 5.1: Simulation Parameters

Parameter	**Assumption**
Cell Deployment	7 sites each with 3 sectors
Inter-Site Distance	1 Km
Carrier Frequency	2 GHz
Number of SubBands	6 (each has 180 kHz bandwidth)
FFR Pattern	Universal frequency reuse
Subframe Interval	1 ms
Pathloss Model (d in Km)	$128.15 + 37.6 \log_{10}(d)$ dB (macro) $136.74 + 39.2 \log_{10}(d)$ dB (pico)
Std. Dev. for Shadowing	8 dB (macro), 10 dB (pico)
Antenna Configuration	2x2 MIMO with ZF receiver
Scheduler	Proportional Fair
BS Transmit Power	43 dBm (macro), 30 dBm (pico)
Antenna Gain	12 dBi (macro), 5 dBi (pico), 0 dBi (user)
User Noise Figure	9 dB
Thermal Noise	-114 dBm/MHz
Channel Model	Typical Urban
User Velocity	3 Km/h



(a)



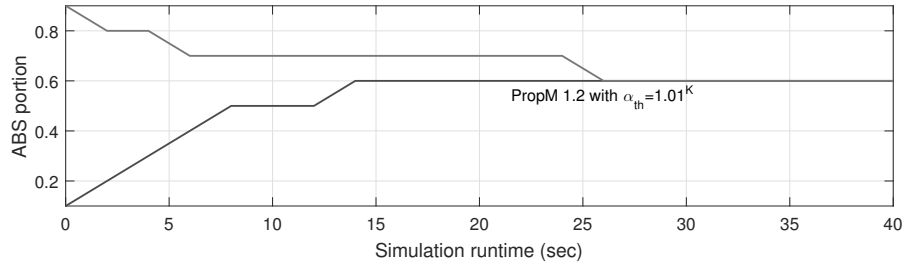
(b)

Figure 5.1: System performance versus fixed ABS portion: (a)Geometric average of throughput (GAT), (b) Average 5th percentile user throughput

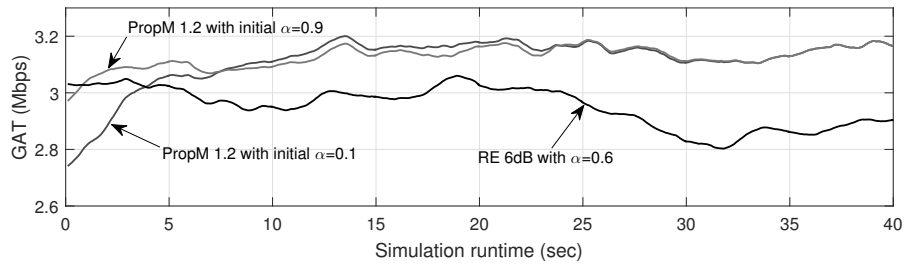
Among the various ABS portion values, $\alpha = 0.6$ provides the best performance in the case of our proposed algorithm, and the performance gain over the best conventional scheme (i.e., RE 6 dB) is measured at 4.0%. This is equivalent to saying that every user in the network experiences a gain in throughput performance of 4.0%. With $\alpha = 0.6$, the average fifth percentile user throughputs for the proposed algorithm and RE with 6 dB, which generally refer to the cell-edge performance, are measured as 792.5 kbps and 659.0 kbps, respectively. This indicates that the performance of the proposed scheme is 20% higher.

Figure 5.2 illustrates the results of the proposed ABS tuning algorithm cooperating with the user association scheme. Each procedure for changing the portion of ABSs is run every 2 s, and the threshold value of $\alpha_{th} = 1.01^K$ is assumed. This means that the ABS portion changes when every user in the network expects 1% of performance enhancement. Regardless of the initial value of α, it approaches to an acceptable range, for example 0.5-0.7, within a very few iterations. After that, the proposed algorithms always outperforms the best conventional scheme.

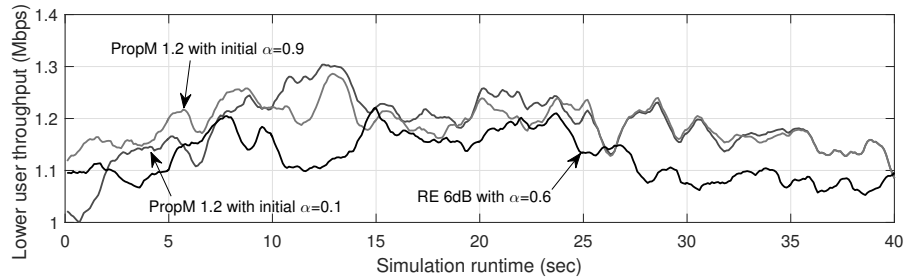
Finally, I verify the proposed algorithms under more dynamic scenarios, where the system configuration changes over time. To this end, I change the number of pico BSs per cell for every 50 s, in order to generate on/off activity of pico BSs. Figure 5.3 depicts the dynamic scenario, the variation of ABS portion, and GAT performance. Our algorithms always provide superior results compared to the conventional scheme using a fixed ABS portion, where the gain in GAT is 1.8-13.6% with an average of 6.1%, and that in cell-edge performance is measured by 3.1-71.0% with an average of 21.5%.



(a)

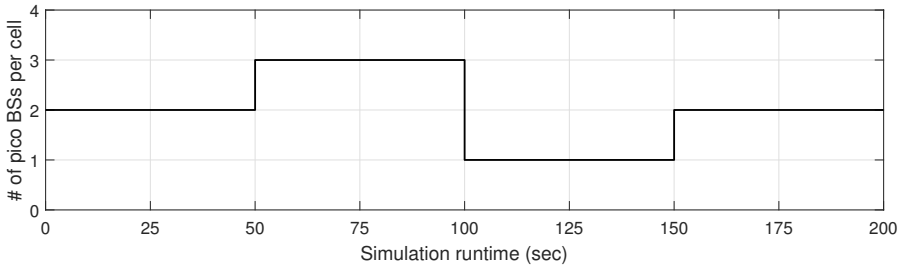


(b)

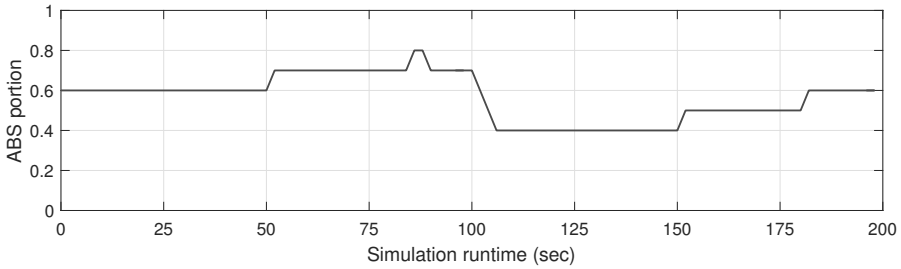


(c)

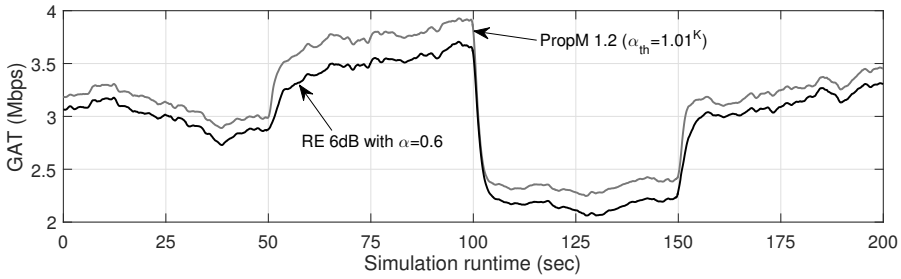
Figure 5.2: Change of ABS portion and corresponding system performances: (a) ABS portion (b) Geometric average or throughput (GAT), (c) Average 5th percentile user throughput. The proposed ABS tuning algorithm is initialized with two different values $\alpha = 0.1$ and 0.9



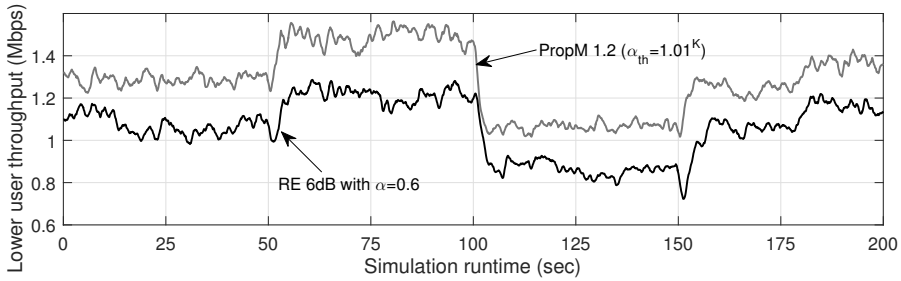
(a)



(b)



(c)



(d)

Figure 5.3: Performances of the joint optimization algorithm (a) Dynamic scenario, (b) ABS portion, (c) Geometric average or throughput (GAT), (d) Average 5th percentile user throughput

Chapter 6

CONCLUSION

In this dissertation, several strategies for enhancing the utilization of pre-installed infrastructures in cellular systems were proposed. All of the proposed algorithms were basically designed to reflect real propagation environments (e.g., small-scale fading phenomenon). At the beginning of the study, user throughput estimation methods for the PF scheduling algorithms were investigated. Based on the intuition that the scheduling preference metric of the PF scheduler for every user varies around a similar interval, I derived an integral form of throughput estimation equation by assuming that the CDF of scheduling preference metric of every user is identical. This assumption enables that all users predict the long-term throughput of themselves without other users' channel information. In order to further simplify the integral calculation, I derived two types of closed-form throughput estimation equations for Rayleigh fading and general fading environments. Even though the final estimation equations are simple because of a series of approximation processes, they provided accurate estimation results.

Based on the proposed throughput estimation methods, both centralized and decentralized user association algorithms were proposed. The centralized algorithm determines associations between every BS and every user at a central node, and therefore, excessive uplink feedback and message exchange between BSs are required. On the

other hand, the decentralized version of algorithm does not yield such overheads, and the only things to do is broadcasting three basic parameters from each BS. Such a feature is especially desirable for heterogeneous cellular networks, where there may exist many candidate BSs in the vicinity of each user. Anyway, both user association algorithms enhanced the system performance in terms of log utility. As a result, the performance of users, located at cell-edge regions thereby experiencing low service quality, was significantly improved. Especially, in heterogeneous cellular environments, it is observed that the decentralized algorithm increases not only low user performance but also the overall network capacity, because this scheme redirects users into small cell BSs in an idle mode.

Finally, a joint optimization of inter-tier interference management and user association in heterogeneous cellular networks was investigated. In order to remove interference from a high powered BS to users at a low powered BS, macro BSs periodically mute their data transmission. I modified user throughput estimation equation to support such an inter-tier interference management technique called eICIC, and proposed an algorithm for the joint optimization problem. It is verified that the proposed scheme significantly increases the utilization of heterogeneous cellular network by efficiently operating the pre-installed infrastructures.

Appendix A
Proof of Proposition 5.1

When the reassociation of user k is successfully completed, the number of users connected to BSs n and m is respectively given as $Y_n - 1$ and $Y_m + 1$. Then, user k expects to obtain $U_{m,k}(Y_m + 1)$ amount of utility from the target BS, because it now begins to be served by that one. Meanwhile, it is disconnected from the original serving BS, thereby loosing the utility of $U_{n,k}(Y_n)$. In addition, any user $i \in \mathcal{K}_m$ suffers from the service quality degradation, because of the increased number of users sharing the same BS. The corresponding utility change is expressed by $U_{m,i}(Y_m + 1) - U_{m,i}(Y_m)$. In contrast, any user $i \in \mathcal{K}_n - \{k\}$ is able to benefit by one user's leaving, and the amount of utility increase is given by $U_{n,i}(Y_n - 1) - U_{n,i}(Y_n)$.

By summarizing all the individual users' performance variations, I can obtain the overall system utility variation by

$$\begin{aligned}\Delta U_{k,n\to m} = {} & U_{m,k}(Y_m + 1) - U_{n,k}(Y_n) \\ & + \sum_{i\in\mathcal{K}_m} \left[U_{m,i}(Y_m + 1) - U_{m,i}(Y_m)\right] \\ & + \sum_{i\in\mathcal{K}_n - \{k\}} \left[U_{n,i}(Y_n - 1) - U_{n,i}(Y_n)\right]. \end{aligned} \tag{1}$$

Finally, the equation in Proposition 5.1 can be obtained by simply rearranging the above equation so that the summation index of the last line belongs to $\mathcal{K}_n$, instead of $\mathcal{K}_n - \{k\}$.

Appendix B
Proof of Proposition 5.3

According to the proposed handover procedure described in Chapter 4.4, many users can change their association simultaneously even within a single iteration. For mathematical tractability, I assume that the handover events within an iteration occurs in an arbitrary order, and show that the modified handover condition is designed to increase the system utility at each phase. Let us consider a handover event, and assume that it is user k's turn to change the association from BS n to m. In this situation, the possible number of users at target BS m can be given by either Y_m in the case where nobody has left that BS, or $Y_m - 1$ when another user has left at a previous phase. In a similar manner, the number of users at the original serving BS n can be given by either Y_n or $Y_n + 1$.

Now I investigate the worst-case scenario. The potential utility gain of user k, as becoming associated with BS m, could be either $U_{m,k}(Y_m)$ or $U_{m,k}(Y_m + 1)$. Because the utility decreases with the number of resource-sharing members, user k is able to have at least $U_{m,k}(Y_m + 1)$ amount of utility.

The handover of this user also degrades the performance of users at the target BS. The following two possible cases should be taken into consideration.

i) Nobody left BS m previously: Any user $i \in \mathcal{K}_m$ suffers from $U_{m,i}(Y_m + 1) - U_{m,i}(Y_m)$ amount of utility decrease. The overall effect is exactly same as $-\Delta U^{(m,1)}$.

ii) User j left BS m in a previous phase: Any user $i \in \mathcal{K}_m - \{j\}$ experiences util-

ity degradation by $U_{m,i}(Y_m) - U_{m,i}(Y_m - 1)$. Then, the total utility difference is expressed by

$$\begin{aligned}
&\sum_{i \in \mathcal{K}_m - \{j\}} [U_{m,i}(Y_m) - U_{m,i}(Y_m - 1)] \\
&\quad = -\Delta U_m^{(m,2)} + U_{m,j}(Y_m - 1) - U_{m,j}(Y_m) \\
&\quad \geq -\Delta U_m^{(m,2)}. \qquad (2)
\end{aligned}$$

According to the results from i) and ii), the maximum amount of performance degradation of users at BS m, due to user k, is represented by $-\max(\Delta U^{(m,1)}, \Delta U^{(m,2)})$.

An extra consideration is required for calculating the effect on users associated with the original serving BS, because user k belongs to $\mathcal{K}_n$. The following two possible scenarios also should be taken into consideration.

iii) Nobody joined at BS n previously: Any user $i \in \mathcal{K}_n - \{k\}$ experiences the utility increase of $U_{n,i}(Y_n - 1) - U_{n,i}(Y_n)$, while user k loses $U_{n,k}(Y_n)$ amount of utility. Then, the overall effect is calculated as

$$\begin{aligned}
&\sum_{i \in \mathcal{K}_n - \{k\}} [U_{n,i}(Y_n - 1) - U_{n,i}(Y_n)] - U_{n,k}(Y_n) \\
&\quad = \Delta U_n^{(n,2)} - U_{n,k}(Y_n - 1). \qquad (3)
\end{aligned}$$

iv) User j already joined at BS n in a previous phase: Any user $i \in \mathcal{K}_n + \{j\} - \{k\}$ takes advantage of the utility increase with the amount of $U_{n,i}(Y_n) - U_{n,i}(Y_n + 1)$, while user k loses the utility of $U_{n,k}(Y_n + 1)$. Hence, the overall utility change is represented by

$$\sum_{i \in \mathcal{K}_n + \{j\} - \{k\}} [U_{n,i}(Y_n) - U_{n,i}(Y_n + 1)] - U_{n,k}(Y_n + 1)$$
$$\geq \sum_{i \in \mathcal{K}_n - \{k\}} [U_{n,i}(Y_n) - U_{n,i}(Y_n + 1)] - U_{n,k}(Y_n + 1)$$
$$= \Delta U_n^{(n,1)} - U_{n,k}(Y_n). \tag{4}$$

The first inequality is because $U_{n,j}(Y_n) - U_{n,j}(Y_n + 1) \geq 0$. From iii) and iv), the minimum amount of utility gain for users at BS n is given by $\min(\Delta U^{(n,1)}, \Delta U^{(n,2)}) - U_{n,k}(Y_n - 1)$.

By summarizing every effect in the worst case (i.e., the amount of utility gained and lost for user k as well as both advantages and disadvantages for users at the BSs involved in the reassociation process), the minimum amount of utility change in this phase is exactly the same as the left side of the inequality in equation (4.10). For this reason, the modified handover condition with threshold level $h_{th} \geq 1$ always guarantees the system utility increase.

Bibliography

[1] H. L. Bertoni, *Radio Propagation for Modern Wireless Systems*, Prentice Hall PTR, 1999.

[2] T. Rappaport, *Wireless Communications: Principles and Practice*, Prentice Hall PTR, 2001.

[3] T. K. Sarkar, Z. Ji, K. Kim, A. Medouri, and M. Salazar-Palma, "A survey of various propagation models for mobile communication," *IEEE Antennas and Propagation Magazine*, vol. 45, no. 3, pp. 51–82, June 2003.

[4] Ministry of Science, ICT and Future Planning, http://www.msip.go.kr

[5] H. Lei, X. Zhang, and D. Yang, "A novel frequency reuse scheme for multi-cell ofdma systems," in *Proceedings of IEEE 66th Vehicular Technology Conference*, September 2007, pp. 347–351.

[6] M. Assaad, "Optimal fractional frequency reuse (FFR) in multicellular OFDMA system," in *Proceedings of IEEE 68th Vehicular Technology Conference*, September 2008, pp. 1–5.

[7] G. Boudreau, J. Panicker, N. Guo, R. Chang, N. Wang, and S. Vrzic, "Interference coordination and cancellation for 4g networks," *IEEE Communications Magazine*, vol. 47, no. 4, pp. 74–81, April 2009.

[8] M. M. Wang, J. Borran, T. Ji, T. Richardson, and M. Dong, "Interference management and handoff techniques," *IEEE Vehicular Technology Magazine* vol. 4, no. 4, pp. 64–75, December 2009.

[9] R. Y. Chang, Z. Tao, J. Zhang, and C. C. Kuo, "A graph approach to dynamic fractional frequency reuse (FFR) in multi-cell OFDMA networks," in *Proceedings of IEEE International Conference on Communications*, June 2009, pp. 1–6.

[10] A. L. Stolyar and H. Viswanathan, "Self-organizing dynamic fractional frequency reuse for best-effort traffic through distributed inter-cell coordination," in *Proceedings of IEEE International Conference on Computer Communications*, April 2009, pp. 1287–1295.

[11] B. Rengarajan, A. L. Stolyar, and H. Viswanathan, "Self-organizing dynamic fractional frequency reuse on the uplink of ofdma systems," in *Proceedings of 44th Annual Conference on Information Sciences and Systems*, March 2010, pp. 1–6.

[12] A. Sang, X. Wang, M. Madihian, and R. D. Gitlin, "A flexible downlink scheduling scheme in cellular packet data systems," *IEEE Transactions on Wireless Communications*, vol. 5, no. 3, pp. 568–577, March 2006.

[13] F. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. 8, pp. 33–37, 1997.

[14] X. Liu, E. K. P. Chong, and N. B. Shroff, "A framework for opportunistic scheduling in wireless networks," *Computer Networks*, vol. 41, no. 4, pp. 451–474, March 2003.

[15] A. Asadi and V. Mancuso, "A survey on opportunistic scheduling in wireless communications," *IEEE Communications Surveys Tutorials*, vol. 15, no. 4, pp. 1671–1688, Fourth 2013.

[16] H. Jin, B. C. Jung, and V. C. M. Leung, "Fundamental limits of cdf-based scheduling: Throughput, fairness, and feedback overhead," *IEEE/ACM Transactions on Networking*, vol. 23, no. 3, pp. 894–907, June 2015

[17] H. J. Kushner and P. A. Whiting, "Convergence of proportional-fair sharing algorithms under general conditions," *IEEE Transactions on Wireless Communications*, vol. 3, no. 4, pp. 1250–1259, July 2004.

[18] R. Agrawal and V. Subramanian, "Optimality of certain channel aware scheduling policies," in *Proceedings of Annual Allerton Conference on Communication, Control and Computing*, 2002.

[19] A. L. Stolyar, "On the asymptotic optimality of the gradient scheduling algorithm for multiuser throughput allocation," *Operations Research*, vol. 53, no. 1, pp. 12–25, january 2005.

[20] D. Avidor, S. Mukherjee, J. Ling, and C. Papadias, "On some properties of the proportional fair scheduling policy," in *Proceedings of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 2, September 2004, pp. 853–858.

[21] S. Borst, "User-level performance of channel-aware scheduling algorithms in wireless data networks," *IEEE/ACM Transactions on Networking*, vol. 13, no. 3, pp. 636–647, June 2005.

[22] J.-G. Choi and S. Bahk, "Cell-throughput analysis of the proportional fair scheduler in the single-cell environment," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 2, pp. 766–778, March 2007.

[23] A. J. Goldsmith and S.-G. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Transactions on Communications*, vol. 45, no. 10, pp. 1218–1230, October 1997.

[24] S. Catreux, P. F. Driessen, and L. Greenstein, "Data throughputs using multiple-input multiple-output (MIMO) techniques in a noise-limited cellular environment," *IEEE Transactions on Wireless Communications*, vol. 1, no. 2, pp. 226–235, April 2002.

[25] S. V. Hanly, "An algorithm for combined cell-site selection and power control to maximize cellular spread spectrum capacity," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 7, pp. 1332–1340, Sepember 1995.

[26] S. Das, H. Viswanathan, and G. Rittenhouse, "Dynamic load balancing through coordinated scheduling in packet data systems," in *Proceedings of International Conference on Computer Communications*, vol. 1, 2003, pp. 786–796.

[27] A. Sang, X. Wang, M. Madihian, and R. D. Gitlin, "Coordinated load balancing, handoff/cell-site selection, and scheduling in multi-cell packet data systems," in *Proceedings of the 10th Annual International Conference on Mobile Computing and Networking*, 2004, pp. 302–314.

[28] T. Bu, L. Li, and R. Ramjee, "Generalized proportional fair scheduling in third generation wireless data networks," in *Proceedings of International Conference on Computer Communications*, 2006, pp. 1–12.

[29] K. Son, S. Chong, and G. Veciana, "Dynamic association for load balancing and interference avoidance in multi-cell networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 7, pp. 3566–3576, July 2009.

[30] M. Hong, A. Garcia, and J. Barrera, "Joint distributed access point selection and power allocation in cognitive radio networks," in *Proceedings of International Conference on Computer Communications*, April 2011, pp. 2516–2524.

[31] H. Kim, G. de Veciana, X. Yang, and M. Venkatachalam, "Distributed α-optimal user association and cell load balancing in wireless networks," *IEEE/ACM Transactions on Networking*, vol. 20, no. 1, pp. 177–190, February 2012.

[32] S. Lee, K. Son, H. Gong, and Y. Yi, "Base station association in wireless cellular networks: An emulation based approach," *IEEE Transactions on Wireless Communications*, vol. 11, no. 8, pp. 2720–2729, August 2012.

[33] M. Hong and A. Garcia, "Mechanism design for base station association and resource allocation in downlink OFDMA network," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 11, pp. 2238–2250, December 2012.

[34] J. Choi, J.-W. Choi, and S.-C. Kim, "An improved throughput estimation method and dynamic user association in multi-cell networks," in *Proceedings IEEE International Symposium on Personal Indoor and Mobile Radio Communications*, 2013, pp. 3349–3353.

[35] J. Choi, W. H. Lee, Y. H. Kim, J. H. Lee, and S. C. Kim, "Throughput estimation based distributed base station selection in heterogeneous networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 11, pp. 6137–6149, November 2015.

[36] R. Madan, J. Borran, A. Sampath, N. Bhushan, A. Khandekar, and T. Ji, "Cell association and interference coordination in heterogeneous LTE-A cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 9, pp. 1479–1489, December 2010.

[37] C. S. Chen, F. Baccelli, and L. Roullet, "Joint optimization of radio resources in small and macro cell networks," in *Proceedings of IEEE Vehicular Technology Conference*, 2011, pp. 1–5.

[38] S. Corroy, L. Falconetti, and R. Mathar, "Dynamic cell association for downlink sum rate maximization in multi-cell heterogeneous networks," in *Proceedings of IEEE International Conference on Communications*, June 2012, pp. 2457–2461.

[39] H.-S. Jo, Y. J. Sang, P. Xia, and J. Andrews, "Heterogeneous cellular networks with flexible cell association: A comprehensive downlink sinr analysis," *IEEE*

Transactions on Wireless Communications, vol. 11, no. 10, pp. 3484–3495, October 2012.

[40] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2706–2716, June 2013.

[41] R. H. Y. Louie, M. R. Mckay, and I. B. Collings, "Maximum sum-rate of mimo multiuser scheduling with linear receivers," *IEEE Transactions on Communications*, vol. 57, no. 11, pp. 3500–3510, November 2009.

[42] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Transactions on Networking*, vol. 8, no. 5, pp. 556–567, October 2000.

[43] W. Rudin, *Principles of mathematical analysis*, New York: McGraw-Hill Book Co., 1976.

[44] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series and Products*, New York: Academic Press, 2000.

[45] R. H. Clarke, "A statistical theory of mobile-radio reception," *Bell System Technical Journal*, vol. 47, no. 6, pp. 957–1000, 1968.

[46] W. C. Jakes and D. C. Cox, Eds., *Microwave Mobile Communications*, Wiley-IEEE Press, 1994.

[47] IEEE standard for local and metropolitan area networks part 16: Air interface for broadband wireless access systems: Advanced air interface, IEEE Std. 802.16m, 2011.

[48] A. Ghosh, N. Mangalvedhe, R. Ratasuk, B. Mondal, M. Cudak, E. Visotsky, T. Thomas, J. Andrews, P. Xia, H. Jo, H. Dhillon, and T. Novlan, "Heterogeneous

cellular networks: From theory to practice," *IEEE Communications Magazine*, vol. 50, no. 6, pp. 54–64, June 2012.

[49] Qualcomm, "LTE heterogeneous networks," July 2013. http://www.qualcomm.com/media/documents/files/lte-heterogeneous-networks.pdf

[50] M. Shirakabe, A. Morimoto, and N. Miki, "Performance evaluation of inter-cell interference coordination and cell range expansion in heterogeneous networks for lte-advanced downlink," in *Proceedings of International Symposium on Wireless Communication Systems*, November 2011, pp. 844–848.

[51] K. I. Pedersen, Y. Wang, B. Soret, and F. Frederiksen, "eICIC functionality and performance for lte hetnet co-channel deployments," in *Proceedings of IEEE Vehicular Technology Conference*, September 2012, pp. 1–5.

[52] Y. Wang and K. I. Pedersen, "Performance analysis of enhanced inter-cell interference coordination in lte-advanced heterogeneous networks," in *Proceedings of IEEE Vehicular Technology Conference*, May 2012, pp. 1–5.

[53] J. Pang, J. Wang, D. Wang, G. Shen, Q. Jiang, and J. Liu, "Optimized time-domain resource partitioning for enhanced inter-cell interference coordination in heterogeneous networks," in *Proceedings of IEEE Wireless Communications and Networking Conference*, April 2012, pp. 1613–1617.

[54] D.-H. Sung and J. Baras, "Utility-based almost blank subframe optimization in heterogeneous cellular networks," in *IEEE Global Communications Conference*, December 2014, pp. 3622–3627.

초 록

LTE (Long-Term Evolution), LTE-Advanced 등의 4세대 이동통신 시스템에서는 데이터 서비스의 속도와 품질이 이전 세대 대비 대폭 개선되었다. 데이터 서비스가 용이해짐에 따라 고화질 비디오 스트리밍 서비스, 클라우드 기반 스토리지 서비스 등 모바일 트래픽을 많이 발생시키는 어플리케이션이 널리 보급되고 있고, 과도한 트래픽으로 인하여 망 과부화 현상 등이 발생하기도 한다. 폭발적으로 증가하는 모바일 데이터 트래픽을 수용하기 위하여 서비스 사업자는 단위면적 당 기지국의 수를 늘리는 등 무선 자원의 공간 재활용률을 높이려는 시도를 하고 있다. 또한 기존의 매크로 기지국 외에 피코, 펨토 기지국 등 설치 비용이 적고 비교적 쉽게 설치할 수 있는 소형 기지국을 도입하여 매크로 기지국의 부하를 경감시키는 등의 논의도 진행 중이다.

본 논문에서는 이동통신 망의 효율적인 운용 방안에 대하여 제안한다. 특히 혼잡한 기지국에 연동되어 낮은 품질의 서비스를 이용하고 있는 사용자를 인접 기지국으로 분산시켜 해당 사용자뿐만 아니라 네트워크 전체의 성능을 개선하는 부하 분산 기법에 대하여 다룬다. 그 첫 번째 단계로 여러 사용자가 하나의 기지국을 공유하는 경우 각 사용자가 받을 수 있는 서비스 품질을 예측하는 방법을 제안한다. 이 과정에서 실제 전파 환경을 반영하기 위하여 다중 경로로 인한 수신신호의 시변 특성을 고려하고, 신호 세기의 변화를 확률 변수로 모델링 하여 2개의 전송률 예측 기법을 제안한다. 제안한 기법은 실제 시스템에서 간단하게 구현이 가능하다는 장점이 있으며, 예측 오차 역시 대부분의 경우에 대하여 10% 미만으로 비교적 정확하다.

다음으로는 전송률 예측 기법을 이용하여 다중 셀 환경에서 사용자 연동 문제를

다룬다. 우선 어느 하나의 기지국에서 네트워크 상의 모든 채널 정보를 수집하여 사용자 개개인에게 최적의 기지국을 할당하는 중앙 집중 방식의 사용자 연동 기법을 제시한다. 하지만 계산에 필요한 모든 정보를 중앙의 기지국에 전달하기 위해서는 많은 상향링크 피드백과 기지국간의 통신이 필요하다. 더욱이 이러한 오버헤드는 단위면적당 기지국의 수에 비례하여 증가한다. 따라서 사용자가 직접 최적의 기지국을 선택하는 분산적 사용자 연동 알고리즘을 추가로 제안한다. 제안한 기법은 기존의 다중 셀 시스템 외에도 주위에 이용 가능한 기지국의 수가 더욱 많아지는 이종 망 환경에서 특히 좋은 성능을 발휘한다.

마지막으로 이종 망 환경에서 매크로 기지국과 소형 기지국 간의 간섭 제어 알고리즘을 제안한다. 매크로 기지국의 경우 비교적 많은 전력을 사용하기 때문에 소형 기지국에 연동된 사용자에게 많은 간섭을 발생시킨다. 이러한 문제를 해결하기 위하여 eICIC (enhanced Inter Cell Interference Coordination) 기법이 제안되었으며, 매크로 기지국이 일정 시간 동안 데이터를 전송하지 않음으로써 소형 기지국에 연동된 사용자들에게 미치는 간섭을 줄일 수 있다. 본 논문에서는 사용자의 연동과 이종 기지국간의 간섭문제를 동시에 해결하는 방법을 제안한다.

주요어: 이동통신 시스템, 이종 망 시스템, 무선 자원 관리, 사용자 연동, 간섭제어
학번: 2012-30947