Load Balancing by Dynamic Base Station Relay Station Associations in Cellular Networks

Zexi Yang, *Student Member, IEEE*, and Zhisheng Niu, *Fellow, IEEE*

***Abstract*—In this paper, we propose a relay-assisted load balancing scheme in cellular networks. The relay stations can be dynamically associated with different base stations. The load transferring from over-loaded cells to neighboring under-loaded cells is realized by dynamically changing the base station-relay station associations. A distributed algorithm, in which each relay station only needs to exchange information with its neighboring base stations and makes a re-association decision by itself, is proposed. Simulation results show that the overall call blocking probability can be reduced significantly when our distributed algorithm is applied.**

***Index Terms*—Cellular networks, load balancing, dynamic relay station association.**

I. Introduction

NEXT generation wireless networks will have more smaller cells to provide ubiquitous high data-rate services. This trend results in more dynamic variation of traffic load in both time domain and space domain. However, most existing cellular networks are designed by assuming a fixed traffic load. To solve this problem, dynamic load balancing schemes are proposed. We can broadly classify the dynamic load balancing schemes in the literature into several categories: (1) Strategies based on channel borrowing from under-loaded cells[2]; (2)Strategies based on base station (BS) selection[3]; (3)Strategies based on power control and cell breathing[4]; (4)Strategies based on relay-assisted traffic transferring[6][7].

Cooperative relaying has been adopted in some standards to extend service coverage and increase cell-edge mobile users' (MUs') throughput[1]. The links between BS and relay stations (RSs), RSs and MUs, and BS and MUs, are referred to as relay links, access links and direct links, respectively. BSs and RSs are usually deployed at elevated locations, e.g., on the roof of buildings or at the top of towers. Thus, it is highly possible that an RS has line of sight channels with more than one neighboring BSs, especially when the cellular networks become denser. In this case, the traffic load can be balanced by making RSs associated with different BSs. Several works focusing on relay-assisted dynamic load balancing already exist[6][7]. Most existing works construct an overlay ad-hoc network that helps transfer traffic, and divide channels into two parts—one for the ad-hoc overlay network, the other for the cellular network. In this paper, we study the load balancing problem with in-band RSs to reduce the overall call blocking probability. We first propose an approximated method to calculate the blocking probability reduced by re-association of each RS, then describe the problem as a weighted maximum independent set problem. We also propose a distributed algorithm in which each RS only needs to exchange information with neighboring BSs and makes a re-association decision by itself. In simulation results, we show that the overall call blocking probability can be reduced significantly when our scheme is applied. It also shows that the performance gap between the distributed algorithm and a centralized greedy algorithm is rather small.

The rest of the paper is organized as follows: Section II provides the system model; The algorithms are proposed in Section III; Simulation results are demonstrated in Section IV; Section V concludes this work.

II. System Model

Consider the forward links of a cellular network with N cells. The BSs set is denoted by $\{\mathrm{BS}_1, \mathrm{BS}_2, \cdots, \mathrm{BS}_N\}$. Assume a pre-defined frequency reuse scheme is applied. Denote the set of BSs that work in the same frequency band with BS_n by $\mathcal{B}_n$. The RSs set is denoted by $\{\mathrm{RS}_1, \mathrm{RS}_2, \cdots, \mathrm{RS}_M\}$. An RS is allowed to be associated with BSs whose signal strength is greater than a threshold. Denote the set of BSs with which RS_m is allowed to be associated by $\mathcal{B}^{(m)}$. In each cell, the resource is divided into orthogonal blocks. Denote the number of resource blocks in each cell by J. Denote the transmission powers of BSs and RSs by $\mathrm{Pt}^{(b)}$ and $\mathrm{Pt}^{(r)}$. Assume that $\mathrm{Pt}^{(b)}$ and $\mathrm{Pt}^{(r)}$ are fixed, and equally distributed in different resource blocks. For the channel model, we only consider location-dependent channel gains, including path loss fading and shadowing. Denote the channel gain between BS_n and RS_m by $G^{(b)}_{n,m}$, and the channel gain between RS_k and RS_m by $G^{(r)}_{k,m}$. Consider homogeneous real-time MUs. By the concept of effective bandwidth for wireless MUs proposed in [8], the required data rate for the MUs can be calculated and denoted by r_c. MUs randomly arrive and depart the network, and can be associated with either a BS or an RS. Assume the arrival process is a Poisson process, and the time duration that each MU stays in the network is an exponential distributed random variable. Denote the arrival rate at location x and time t by $\lambda(x,t)$, and the average call duration by $\frac{1}{\mu}$. To avoid intra-cell interference, resource blocks are allocated orthogonally to different MUs in one cell, and the MUs in one cell are scheduled in a round-robin way. The interference is generated from other co-channel cells. Since the transmitter in one cell is determined by the scheduling scheme and RS selection results, it is very complex and difficult to accurately

Manuscript received October 30, 2012. The associate editor coordinating the review of this letter and approving it for publication was N. Mehta.

Z. Yang and Z. Niu are with Tsinghua National Laboratory for Information Science and Technology, Dept. of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: yzx03@mails.tsinghua.edu.cn).

This work is sponsored in part by the National Basic Research Program of China (2012CB316001), the Nature Science Foundation of China (60925002, 61021001), and Hitachi Ltd.

Digital Object Identifier 10.1109/WCL.2012.121812.120797

2162-2337/13$31.00 © 2013 IEEE

estimate the interference. To make a conservative estimating of interference, assume that the interference from a co-channel cell is generated by the BS/RS that, if transmitted, induces the strongest interference. Thus, the signal-to-interference-plus-noise ratio (SINR) for the relay link between $\mathrm{BS}_{n(m)}$ and RS_m can be expressed as:

$$\Gamma^{(b)}_{n(m),m} = \frac{\mathrm{Pt}^{(b)} G^{(b)}_{n(m),m}}{N_0 + \sum\limits_{k \in \mathcal{B}_{n(m)}} \max\left\{\mathrm{Pt}^{(b)} G^{(b)}_{k,m}, \max\limits_{j \in \mathcal{R}^{(k)}} \{\mathrm{Pt}^{(r)} G^{(r)}_{j,m}\}\right\}}, \tag{1}$$

where N_0 is the noise power, and $\mathcal{R}^{(k)}$ is the set of RSs that are associated with BS_k. Denote the channel gain between BS_n and an MU by $G^{(b)}_n$, and the channel gain between RS_m and this MU by $G^{(r)}_m$. The SINR between BS_n and the MU, and RS_m and the MU can be expressed as:

$$\Gamma^{(b)}_n = \frac{\mathrm{Pt}^{(b)} G^{(b)}_n}{N_0 + \sum\limits_{k \in \mathcal{B}_n} \max\left\{\mathrm{Pt}^{(b)} G^{(b)}_k, \max\limits_{j \in \mathcal{R}^{(k)}} \{\mathrm{Pt}^{(r)} G^{(r)}_j\}\right\}}, \tag{2}$$

$$\Gamma^{(r)}_m = \frac{\mathrm{Pt}^{(r)} G^{(r)}_m}{N_0 + \sum\limits_{k \in \mathcal{B}_{n(m)}} \max\left\{\mathrm{Pt}^{(b)} G^{(b)}_k, \max\limits_{j \in \mathcal{R}^{(k)}} \{\mathrm{Pt}^{(r)} G^{(r)}_j\}\right\}}. \tag{3}$$

The transmission data rates between $\mathrm{BS}_{n(m)}$ and RS_m, BS_n and the MU, and RS_m and the MU are approximately represented by Shannon's formula, and denoted by $r^{(b)}_{n(m),m}$, $r^{(b)}_n$, and $r^{(r)}_m$. Assume that the relay scheme of decode-and-forward is adopted, and there is no signal combination at the receiver side. Other relay schemes can be easily extended. If an MU is associated with RS_m, then when this MU is served, each resource block used by the MU is divided into two phases, and the division aims at maximizing the equivalent data rate. In the first phase, $\mathrm{BS}_{n(m)}$ transmits to RS_m, and in the second phase, RS_m transmits to the MU. Thus, the equivalent data rate when the MU is associated with RS_m is $r^{(e)}_m = \frac{r^{(r)}_m r^{(b)}_{n(m),m}}{r^{(r)}_m + r^{(b)}_{n(m),m}}$.

III. Relay Assisted Load Balancing Scheme

Assume that each MU is associated with the BS or RS that has the strongest signal strength. When a new MU arrives, its required resource blocks number can be denoted by $k^{(s)} = \lceil \frac{r_c}{r^{(s)}} \rceil$, where $r^{(s)}$ is the data rate for the MU in one resource block. If $k^{(s)}$ is smaller than or equal to the unallocated resource blocks in that cell, the MU can be accepted. Otherwise, it is blocked. The objective of RSs association at time t is to minimize the overall call blocking probability. Denote the traffic arrival rate for the n^{th} cell by Λ_n, and the coverage area of the n^{th} cell by $\mathcal{L}_n$. Thus, Λ_n and $\mathcal{L}_n$ depend on the RSs associations. The problem is formulated as (4). The subscript t is skipped for ease of exposure.

$$\begin{aligned} &\min_{\mathrm{I}_{m,n}} \sum_{n=1}^{N} \Lambda_n \mathrm{Pr}_n \\ &\text{s.t.} \quad \sum_{n=1}^{N} \mathrm{I}_{m,n} = 1, \qquad \forall m \\ &\qquad \mathrm{I}_{m,n} = 0, \qquad \forall n \notin \mathcal{B}^{(m)} \\ &\qquad \Lambda_n = \textstyle\int_{x \in \mathcal{L}_n} \lambda(x) \mathrm{d}x, \quad \forall n. \end{aligned} \tag{4}$$

In (4), Pr_n is the call blocking probability for the n^{th} cell, and $\mathrm{I}_{m,n} \in \{0,1\}$ indicates whether RS_m is associated with BS_n. The first and the second constraints guarantee that each RS_m must be associated with one BS in $\mathcal{B}^{(m)}$.

One of the difficulties in solving (4) is that Pr_n does not have analytical expression. Thus, we propose an approximated method to calculate the call blocking probability for one cell. Without loss of generality, consider the n^{th} cell in the rest of this paragraph. For a new arrival MU, denote the probability that its required resource blocks number equals to j by P_j. Thus, we have $\sum\limits_{j=1}^{\infty} P_j = 1$. First consider a simplified case in which the required resource blocks numbers for all MUs are the same, i.e., we have $P_{j_0} = 1$ for some j_0. In this simplified case, the problem can be modeled as an M/M/s(0) queue, and the call blocking probability can be expressed as:

$$\frac{a^s}{s!} \left[\sum_{k=0}^{s} \frac{a^k}{k!}\right]^{-1}, \tag{5}$$

where $s = \lfloor \frac{J}{j_0} \rfloor$, and $a = \frac{\Lambda_n}{\mu}$. For the general case in which required resource blocks numbers for MUs are different, define the average required resource blocks number as $\overline{j_0} = \sum\limits_{j=1}^{\infty} P_j j$. By substituting j_0 with $\overline{j_0}$ and from (5), we obtain an approximated analytical expression of call blocking probability for the n^{th} cell. In the rest of this section, the blocking probability we consider is the one calculated by our approximated method.

Define the coverage of a cell as the sum of the BS's coverage and the coverage of the RSs that are associated with the BS. Name the set of cells whose coverage and blocking probabilities might be influenced by the association of RS_m as Affected Cells Cluster (ACC) of RS_m, and denote it by $\mathcal{A}_m$. Assume a protocol that can guarantee the information exchange between RS_m and BSs in $\mathcal{A}_m$ exists. Define the *weighted blocking probability* of $\mathcal{A}_m$ as $\mathrm{Pr}^{(\mathcal{A})}_m = \sum\limits_{n \in \mathcal{A}_m} \Lambda_n \mathrm{Pr}_n$. Thus, $\mathrm{Pr}^{(\mathcal{A})}_m$ is a measure of the call blocking probability level of $\mathcal{A}_m$. Given the traffic distribution and the associations of other RSs, an optimal association of RS_m exists, which can minimize $\mathrm{Pr}^{(\mathcal{A})}_m$. Define the subtract of the current $\mathrm{Pr}^{(\mathcal{A})}_m$ and the minimized $\mathrm{Pr}^{(\mathcal{A})}_m$ as the *weighted blocking probability gain* of RS_m, and denote it as $G^{(\mathcal{A})}_m$. Thus, we have $G^{(\mathcal{A})}_m \geq 0$.

Construct an undirect graph $G(V, E)$. Each vertex v_m in G corresponds to RS_m in the network. An edge between v_i and v_j exists if and only if $\mathcal{A}_i$ and $\mathcal{A}_j$ overlap with each other. We say that v_i and v_j are adjacent if they have an edge between them. Since the calculation of the weighted blocking probability gain of one RS is based on the condition that other RSs keep their current associations, for one RS, its weighted blocking probability gain may be different if the associations of some other RSs are changed. From the definition of ACC, it can be derived that if v_i and v_j are not adjacent to each other, then the calculation of the weighted blocking probability gain of RS_i is not affected by the association of RS_j, and vice versa. Thus, we state that the weighted blocking probability gain of RS_i and RS_j are independent from each other if v_i and v_j are not adjacent. The vertex v_m and all its adjacent vertices form a vertices set denoted by $\mathcal{V}_m$. Each vertex

v_m has a non-negative weight $G_m^{(\mathcal{A})}$. Since it is hard to predict the call blocking probability if RSs that correspond to adjacent vertices make associations simultaneously, we aim to select the RSs whose weighted blocking probability gains are independent from each other, and maximize the reduction of overall call blocking probability. This is equivalent to the weighted maximum independent set (WMIS) problem, which is a well-known NP-hard problem. There are several greedy algorithms that can solve the WMIS problem. However, a central node is required for these algorithms, and the weighted blocking probability gains and association decisions must be exchanged between the central node and the RSs. As the network scale enlarges, the overhead of centralized algorithm increases. Thus, we propose a distributed RS re-association scheme as follows:

1) **Information Gathering and Distribution:** Each BS_n measures the traffic load in its cell, in terms of the arrival rate and the required resource block numbers of the MUs. Each BS_n also records the list of RSs which are currently associated with it. Each BS_n distributes this information to the RSs which have BS_n in their ACCs.
2) **Weighted Blocking Probability Gain Calculation:** With the information from BSs and the channel state information, and based on the proposed method that approximately calculates call blocking probability, each RS_m traverses all its possible associations, finds the optimal association that can minimize $\text{Pr}_m^{(\mathcal{A})}$, and calculates $G_m^{(\mathcal{A})}$.
3) **Information Exchange:** Each RS_m reports $G_m^{(\mathcal{A})}$ to the BSs in $\mathcal{A}_m$. For each BS_n, it will receive $G_m^{(\mathcal{A})}$ from multiple RSs, and it will select the maximal positive one and respond to that RS with a positive signal.
4) **RS Re-association and MU Re-association:** Each RS_m that receives the positive signal from all the BSs in $\mathcal{A}_m$ is re-associated to minimize $\text{Pr}_m^{(\mathcal{A})}$.

The proposed RS re-association scheme is executed periodically. The interval between two executions is related to the variance of traffic, and is set in practical implementations. Theorem 1 states the characteristics of the RSs that are re-associated with our distributed scheme.

Theorem 1: If $G_m^{(\mathcal{A})} > 0$ and $G_m^{(\mathcal{A})} = \max_{v_s \in \mathcal{V}_m} G_s^{(\mathcal{A})}$, then RS_m will be re-associated to minimize $\text{Pr}_m^{(\mathcal{A})}$.

Proof: Assume $G_m^{(\mathcal{A})} > 0$ and $G_m^{(\mathcal{A})} = \max_{v_s \in \mathcal{V}_m} G_s^{(\mathcal{A})}$. We have $\forall \text{BS}_n \in \mathcal{A}_m$, BS_n will receive $G_m^{(\mathcal{A})}$ reported by RS_m. For any other RS_s that also reports $G_s^{(\mathcal{A})}$ to BS_n, we must have $\text{BS}_n \in \mathcal{A}_s$, and $\mathcal{A}_m \cap \mathcal{A}_s \neq \emptyset$. By the definition of $G(V, E)$, we have that v_m and v_s have an edge between them, and $v_s \in \mathcal{V}_m$. Since $G_m^{(\mathcal{A})} = \max_{v_s \in \mathcal{V}_m} G_s^{(\mathcal{A})}$, we have that $G_m^{(\mathcal{A})}$ is the maximal weighted blocking probability gain received by BS_n, and BS_n will respond with a positive signal to RS_m. From step 3 in our proposed scheme, RS_m will be re-associated to another BS to minimize $\text{Pr}_m^{(\mathcal{A})}$.■

Moreover, it can be easily proved that, if there exists RS_m that satisfies $G_m^{(\mathcal{A})} > 0$, then at least one RS is re-associated.

IV. Simulation Results

Consider a cellular network with 19 cells. The radius of one cell is 500m. In each cell, three RSs are deployed. Consider two RSs deployment methods as Fig. 1 shows. The first method is to deploy RSs evenly, and the distance between a BS and an RS in one cell is 300 meters. In the second method, each RS is distributed randomly in one certain area. The universal frequency reuse with fractional frequency reuse (FFR) is adopted. The MUs that are associated directly with BSs are interior MUs and the MUs that are associated with RSs are cell-edge MUs. All the interior MUs are allocated a common bandwidth while the cell-edge MUs' bandwidth is partitioned across cell based on a reuse factor of three. The allocation of bandwidth between interior MUs and cell-edge MUs follows the traffic loads of the two parts of MUs[11][1]. Transmission powers of BSs and RSs are set as 43dBm and 37dBm. Noise power is -100 dBm. We adopt the path loss channel model proposed in [10], which has been verified by practical measurements in some urban area in Japan. For each RS_m, both $\mathcal{B}^{(m)}$ and its ACC $\mathcal{A}_m$ constitute the three nearest BSs. The aggregate bandwidth of the system is 1MHz. Simulation duration T_s is set as 1 hour. The time interval between two algorithms executions is 6 minutes.

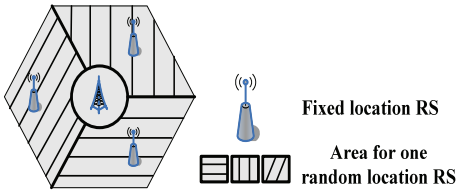
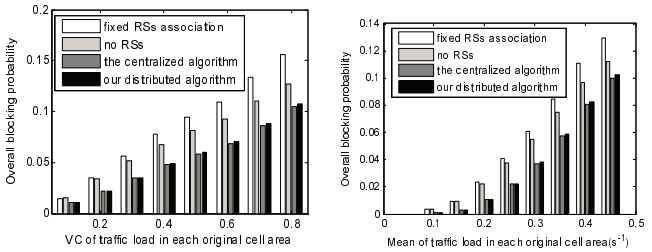


Fig. 1. Simulation topology: RSs can be deployed either evenly or randomly.

MUs randomly arrive in the network with rate requirement of 20K bit/s, and the average call duration is 50 seconds. Assume that the traffic is uniformly distributed within one original cell area[2], and the traffic loads in different original cell areas are distinct. The traffic load in each original cell area is described by an arrival rate. As ref. [9] has pointed out that the spatial statistic of cellular traffic can be described by a lognormal distribution, we want to generate traffic loads that satisfy the following conditions: (1) At any time, the traffic load of one original cell area should be a random variable with lognormal distribution; (2) At any time, the traffic loads of different original cell areas are independent from each other; (3) The traffic load of one original cell area should vary continuously. Denote the arrival rate in the original n^{th} cell area at time t by $\Lambda_n(t)$. First generate i.i.d. normal distribution variables $\alpha_0^{(n)} \sim \mathcal{N}(\eta, \sigma^2)$ and $\alpha_{T_s}^{(n)} \sim \mathcal{N}(\eta, \sigma^2)$, $n \in \{1, 2, \cdots, N\}$. Then generate $\alpha_t^{(n)}, n \in \{1, 2, \cdots, N\}$ following: $\alpha_t^{(n)} = \frac{t'\alpha_{T_s}^{(n)} + (1-t')\alpha_0^{(n)} - \eta}{2t'^2 + 1 - 2t'} + \eta$, where $t' = \frac{t}{T_s}$. It can be easily verified that $\Lambda_n(t) = e^{\alpha_t^{(n)}}, n \in \{1, 2, \cdots, N\}, t \in [0, T_s]$ satisfies the above conditions. The average arrival rate in each original cell area η is used to denote the intensity of the traffic load, and the variance coefficient (VC) defined as $\frac{\sigma^2}{\eta^2}$ is used to denote

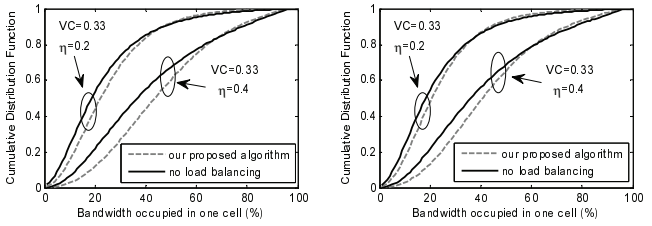
[1]Our previous analysis can be easily extended to the FFR scenario. In each cell, the blocking probability for the BS and for all the RSs can be calculated respectively. The blocking probability of the cell is the average of them weighted by arrival rates.

[2]The original cell area has the hexagonal shape as Fig. 1 shows. When our scheme is applied, the real coverage of each cell will change as the associations of RSs change.



(a) (b)

Fig. 2. Comparison of call blocking probabilities when different schemes are applied. The RSs are evenly deployed. (a) Impact of the VC of traffic load in each original cell area. The average arrival rate is set as 0.3 s^{-1}. (b) Impact of the average arrival rate in each original cell area. The VC of traffic load is set as 0.33.



(a) (b)

Fig. 3. Cumulative distribution function of bandwidth occupied in each cell. (a) The RSs are evenly deployed. (b) The RSs are randomly deployed.

the variation of the traffic load.

To verify the performance of our proposed distributed algorithm, we compare it with the following centralized greedy algorithm that solves the WMIS problem. Select the vertex with maximal weight in the independent set, and delete this vertex and all of its neighboring vertices from the graph. Repeat this process for the remaining subgraph until the subgraph becomes empty. The RSs selected to make re-associations with our distributed algorithm are a subset of that with the centralized algorithm. We also incorporate the scenario without RSs into comparison. Simulation results in Fig. 2 show that with different traffic loads, our proposed algorithm can significantly reduce the overall blocking probability. It also shows that the overall blocking probabilities, when our proposed distributed algorithm and the centralized greedy algorithm are applied, are very close. It should be noted that compared with the scenario that each RS is associated with the closest BS, the scenario without RSs has lower blocking probability. This is because we adopt the RSs selection scheme that each MU is associated with the BS or RS that has the strongest signal strength. The RSs selection scheme is easy to be implemented, but may not be optimal for maximizing throughput. Fig. 2 also shows that compared with the scenario without RSs, our proposed scheme still can reduce blocking probability by more 10%. If signal combination and spatial reuse[12] are incorporated, the blocking probability for the scenarios with RSs can be further reduced. To observe the traffic load distribution when our algorithm is applied, we make snapshots of the network every 6 minutes and get the cumulative distribution function of bandwidth occupied in each cell. The results in Fig. 3 show that when our algorithm is applied, those cells that are light-loaded afford more traffic load due to traffic transferring from over-loaded cells. The calculation results show that compared with the scenario without load balancing algorithms, our distributed algorithm can improve the bandwidth utilization by about 10% with the fixed RSs deployment method and about 6% with the random RSs deployment method.

V. Conclusion

We investigate the load balancing problem by RSs association in a cellular network to reduce the overall call blocking probability. We propose an approximated method to calculate the call blocking probability reduced by re-association of each RS, and describe the problem as a weighted maximum independent set problem. With our proposed distributed algorithm, each RS only needs to exchange information with neighboring BSs and make a re-association decision by itself. Simulation results show that in a cellular network where traffic distributes non-uniformly, the overall call blocking probability can be significantly reduced by applying our proposed schemes.

References

[1] 3GPP TR 36.814, "Further advancements for E-UTRA physical layer aspects," v9.0.0, Mar. 2010.

[2] S. Patra, *et al.*, "Improved genetic algorithm for channel allocation with channel borrowing in mobile computing," *IEEE Trans. Mobile Computing*, vol. 5, no. 7, pp. 884–892, July 2006.

[3] K. Son, *et al.*, "Dynamic association for load balancing and interference avoidance in multi-cell networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3566–3576, July 2009.

[4] S. V. Hanly, "An algorithm for combined cell-site selection and power control to maximize cellular spread spectrum capacity," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 7, pp. 1332–1340, Sept. 1995.

[5] K. Pedersen, *et al.*, "An overview of downlink radio resource management for UTRAN long-term evolution," *IEEE Commun. Mag.*, vol. 47, no. 7, pp. 86–93, July 2009.

[6] X. Wu, *et al.*, "MACA: an efficient channel allocation scheme in cellular networks," in *Proc. 2000 IEEE Globecom*, pp. 1385–1389.

[7] H. Wu, *et al.*, "Integrated cellular and ad hoc relaying systems: iCAR," *IEEE J. Sel Areas Commun.*, vol. 19, no. 10, pp. 2105–2115, Oct. 2001.

[8] J. Tang, *et al.*, "Effective bandwidth-based QoS provisioning for real-time audio/video streaming over MIMO-OFDM wireless networks," in *Proc. 2005 IEEE IPDPS*.

[9] U. Gotzner, *et al.*, "Spatial traffic distribution in cellular networks," *1998 IEEE Vehicular Technology Conference*.

[10] T. Oda, *et al.*, "Advanced LOS path-loss model in microcellular mobile communications," *IEEE Trans. Veh. Technol.*, vol. 49, no. 6, Nov. 2000.

[11] T. Novlan, *et al.*, "Analytical evaluation of fractional frequency reuse for OFDMA cellular networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 12, pp. 4294–4305, Dec. 2011.

[12] Z. Yang, *et al.*, "Throughput improvement by joint relay selection and link scheduling in relay-assisted cellular networks," *IEEE Trans. Veh. Technol.*, vol. 61, no. 6, pp. 2824–2835, July 2012.