UDC 621.391.1

FAIRNESS FOR OPTIMIZATION IN COOPERATIVE WIRELESS CELLULAR NETWORKS

Maxime Guirnyk

Royal Institute of Technology, Stockholm, Sweden

Cooperative communication can improve the performance of cellular mobile networks. The optimization techniques based on duality theory, decomposition and subgradient method were applied for such wireless cellular cooperative systems. It was shown that the joint optimization and resource allocation problem can be solved efficiently within a network utility maximization framework. A concept of proportional fairness was used to achieve fair distribution of quality of service among users. Simulation results confirm the validity of the theoretical work.

Introduction

Cooperative communication can improve the performance of cellular mobile networks in terms of data rate, power saving, interference mitigation, etc. The basic technique is relaying, i.e. retransmission of the signal via a partner node. For example, consider a cooperative transmission scheme shown in Fig. 1. Users 1 and 2 try to transmit their codewords to the base transceiver station (BTS). Since the transmission is realized wirelessly via broadcasting, users can receive each other's codewords and simply retransmit them to the base station. Thus, the base transceiver station will have two copies of each codeword and can decode them with more reliability.

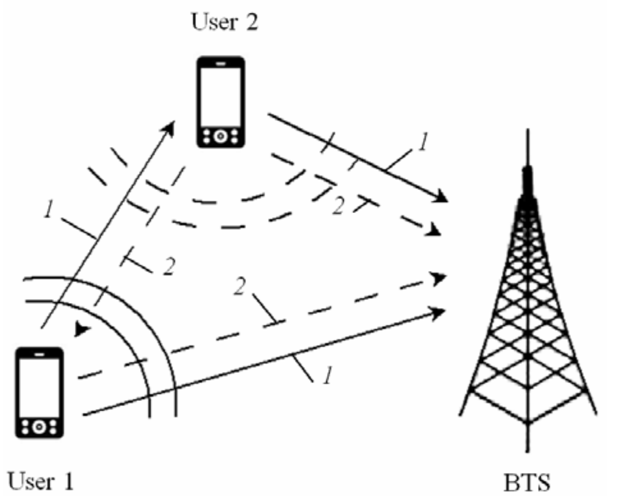

Fig. 1. Cooperative communication between two users and a base station: (*1*) transmission trajectory of user 1; (2) transmission trajectory of user 2.

In [1] several relaying strategies available for the cooperative communication are described. Using decode-and-forward (DF) strategy user tries to decode the partner's bits and then retransmits the decoded bits to the destination. Therefore each user has a partner to provide the diversity, which leads to increased reliability of communication [2]. This strategy is useful when a relaying partner is able to decode a message, and it is located closer to the source than to the destination [3]. Another cooperative transmission strategy is amplify-and-forward (AF) strategy. Each user receives a noisy version of the signal transmitted by its partner, amplifies and retransmits it. The destination combines the information sent by the user and his partner, and makes the final decision on the transmitted bits [1]. In the simplest case, the destination only decides which version is better to choose. AF strategy is more suitable when the relay is located closer to the destination and its received signal-to-noise ratio (SNR) may not be high enough to allow decoding [3].

In a power- and bandwidth-limited network the choice of the relaying strategy also depends on the amount of the expanded power at the source and at the relay nodes. Since each node can act as a source, a destination and a relay simultaneously, the resource allocation problem is coupled with the selection of relays and their retransmission strategies.

While using the existing algorithms of the resource allocation a problem of fairness among nodes may appear. In a cell of a wireless cellular system all the nodes demand for communication with different intensity which depends on the services used. If one part of the users is close to the BTS and another one is quite far, we may observe that the terminals situated closer the base transceiver station always would have the preference to use the channel. The goal of this paper is to provide an effective algorithm of the resource allocation among the transmitting and relaying terminals, choosing the relay, its relaying strategy, and providing the fairness among the users.

System model

The target application is a cooperative wireless cellular network shown in Fig. 2. We assume that orthogonal frequency-division multiple access (OFDMA) [4] is used in the network. The multiple access in the OFDMA scheme is achieved by assigning subsets of subcarriers (OFDM frequency tones) to individual users. It allows simultaneous data transmission from several users using the same spectrum. Thus, each user terminal can act as a source and a relay at the same time, but using different frequency tones. The bandwidth allocation is simplified to a simple assignment of source-destination pair among OFDM tones, similarly to [3]. The network is assumed to be centralized, and therefore the BTS determines the cooperation between the nodes.

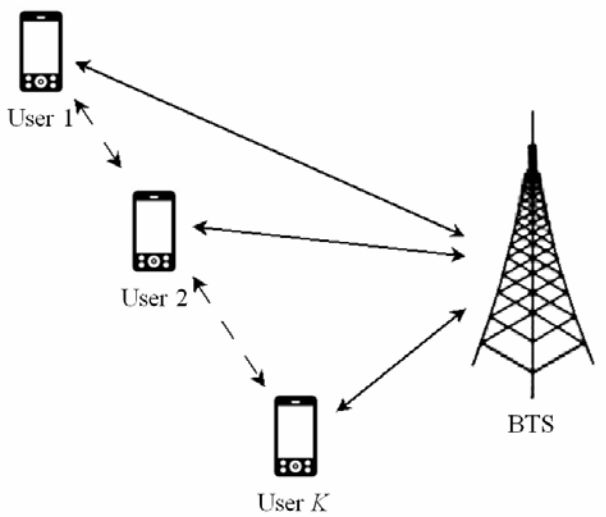

Fig. 2. Cooperative wireless cellular network.

We assume a slow-fading environment with full channel side information (CSI) available at the BTS. The coherence bandwidth of the channel is greater than the bandwidth of several frequency tones and, hence, fading between tones which are far away from each other is uncorrelated.

Taking into account path loss, the channel gain can be modeled as

$$h = H / G,$$

where H is a random variable with the Rayleigh distribution, which is used to represent the Rayleigh fading in the channel; G represents the path loss.

Let us assume that there are K users in the network illustrated in Fig. 2 and denote the base transceiver station as node $K+1$. There are N OFDM tones available and we assume that there is no inter-carrier interference. Also we restrict each tone to have a single active data stream at each moment. However, each of $M = 2K$ streams can use more than one tone. Each transmission has at most one relay used and the BTS can never be a relay. We assume that links source-destination, source-relay and relay-destination use the same frequency tone.

Let $\mathbf{P}$ denote a $(K+1) \times N$ matrix such that its entry P_{kn} is the power used by node k in tone n. Let $\mathbf{R}$ be a $2K \times N$ matrix such that its entry R_{mn} is the data rate of stream m in tone n. Since for different relaying strategies the relation between $\mathbf{P}$ and $\mathbf{R}$ differs, we denote it as $\mathbf{R} \in \mathrm{C}(\mathbf{P})$. This expression represents the achievable rate region, i.e. the set of all possible rates achievable at a given power level. For the detailed definition of the region reader is referred to [5].

Let $\mathbf{P1}$ denote the row sum of $\mathbf{P}$, hence $\mathbf{P1}$ is a $(K+1) \times 1$ vector which represents the total powers of the nodes across all the tones. Similarly, $\mathbf{R1}$ stands for the row sum of $\mathbf{R}$. Therefore $\mathbf{R1}$ is a $2K \times 1$ vector which represents the total data rates of each stream across all the tones. Hereafter, m th entry of $\mathbf{R1}$ is denoted as $R1_m$. Each node has its separate power constraint which is represented by a $(K+1) \times 1$ vector $\mathbf{p}$. Similarly, the traffic demand of each stream is denoted by a $2K \times 1$ vector $\mathbf{r}$.

Each data stream has an associated utility function [6], which represents user's satisfaction of the provided service. The choice of this function depends on the services used (voice, data, video, etc.). In this setting it is chosen to be an increasing and concave function of the data rate. This implies that the rate, initially, is more valuable than additional rates. The choice of the utility function is explained by a certain risk of having no rate at all. The objective becomes to maximize the sum utility in the network providing the fairness among all the users in the cell.

Let U_m be a utility function associated with data stream m. U_m is a function of $R1_m$, which should be lower-bounded by the traffic demand r_m of the stream. Many different choices of U_m are possible. Similarly to [3] for our simulations we define the utility function as

$$U_m(r_m) = \begin{cases} a(1 - e^{-br_m}), & r_m \geq 0 \\ -\infty, & r_m < 0 \end{cases}, \qquad (1)$$

where a and b are strictly positive real numbers; a represents the upper limit of the utility; b determines the rate $c = -\ln(0.1)/b$ at which the utility is equal to $0.9a$.

Optimization framework

In this section we aim to formulate an optimization problem in terms of utility maximization with respect to power and rate constraints and to solve it using corresponding optimization tools.

To provide fairness between users in terms of traffic distribution we use a concept of proportional fairness [7]. Here, the objective is to achieve a solution, such that each stream m has at least a certain minimal guaranteed rate which is not less than some percentage ω_m of the total rate of all the streams. This condition can be represented as a constraint

$$r_m / \sum_{j=1}^{M} r_j \geq \omega_m , \tag{2}$$

where r_m is the traffic demand of the stream m; ω_m is its minimal part of the sum of rates among all the streams. This small rate can be sufficient, for example, for a short message texting for users at the edge of a cell.

Note that the constraint (2) is not valid for iterative optimization algorithms since at the iteration l the sum $\sum_{j=1}^{M} r_j(l)$ is unknown. Therefore the solution will require an exhaustive search over all possible solutions that satisfy (2) and thus, will suffer from the computational complexity.

However, (2) can be modified into

$$r_m(l) \geq \omega_m \sum_{j=1}^{M} r_j(l-1) . \tag{3}$$

In this case, at iteration l we already know $\sum_{j=1}^{M} r_j(l-1)$ and can consider it as a new constraint to the main optimization problem, which becomes

$$\max_{\mathbf{P},\mathbf{R}} \quad \sum_{m=1}^{M} U_m \tag{4}$$

$$\text{s.t.} \quad \mathbf{P1} \leq \mathbf{p}, \quad \mathbf{R1} \geq \mathbf{r}, \quad \mathbf{R} \in \mathrm{C}(\mathbf{P}) \tag{5}$$

$$\mathbf{r} \geq \boldsymbol{\omega} \sum_{j=1}^{M} r_j(l-1) \tag{6}$$

The optimization problem can be read in the following way: maximize the objective sum utility function $\sum_{m=1}^{M} U_m$ over allocation of $\mathbf{P}$ and $\mathbf{R}$ (4) subject to the set of constraints (5), (6). The meaning of the first constraint $\mathbf{P1} \leq \mathbf{p}$ in (5) is that sum of allocated transmitted powers at each user's terminal should be smaller than its power constraints. Similarly, sum rate of each stream should be greater than its traffic demand, i.e. $\mathbf{R1} \geq \mathbf{r}$. Here we use an element-wise comparison of two vectors. The third constraint means that the rates of all streams should belong to the achievable rate region, which is determined by the particular relaying strategy used by a node. The last constraint (6) means that the minimum rate of each stream should be greater than a certain threshold computed at the previous iteration and based on the proportional fairness concept.

The optimal solution of this problem is a mixed integer programming problem and hence it needs an exhaustive search over all possible resource allocations, relays and relaying strategies. According to [8] the problem (4) has zero duality gap when the number of OFDM tones is large. It means that the optimal solution of (4) is a convex function of $\mathbf{p}$ and, therefore, we can use the convex optimization means to solve it.

To solve the problem (4) we use a dual decomposition technique from [9]. First, we relax the constraints $\mathbf{R1} \geq \mathbf{r}$ and $\mathbf{r}(l) \geq \boldsymbol{\omega} \sum_{j=1}^{M} r_j(l-1)$, and write the Lagrangian

$$L = \sum_{m=1}^{M} U_m + \boldsymbol{\lambda}^T(\mathbf{R1} - \mathbf{r}) + \boldsymbol{\xi}^T[\mathbf{r}(l) - \boldsymbol{\omega} \sum_{j=1}^{M} r_j(l-1)],$$

where $\boldsymbol{\lambda}$ and $\boldsymbol{\xi}$ are vectors of Lagrange multipliers (dual variables). They represent prices which users have to pay for violation the corresponding rate constraints.

Then we regroup the variables in the Lagrangian and write the corresponding dual function which according to [3] can be decoupled into two dual functions corresponding to application layer and physical layer subproblems. First subproblem is the rate adaptation problem in the application layer. Its dual function is given by

$$\begin{aligned} g_{appl} &= \max_{\mathbf{r}} \sum_{m=1}^{M} \{U_m - \lambda_m r_m + \\ &+ \xi_m [r_m - \omega_m \sum_{j=1}^{M} r_j(l-1)]\} = \\ &= \max_{\mathbf{r}} \sum_{m=1}^{M} [U_m + (\xi_m - \lambda_m) r_m + \\ &+ \xi_m \omega_m \sum_{j=1}^{M} r_j(l-1)] . \end{aligned} \tag{7}$$

The second subproblem is the joint relay-strategy and relay-node selection, power and bandwidth allocation in the physical layer with the corresponding dual function

$$g_{phy} = \max_{\mathbf{P},\mathbf{R}} \sum_{m=1}^{M} \lambda_m \sum_{n=1}^{N} R_{mn} .$$

where R_{mn} denotes entry (m, n) of matrix $\mathbf{R}$.

The optimization subproblem is thus

$$\min_{\lambda} g_{phy}$$
$$\text{s.t.}\quad \mathbf{P1} \le \mathbf{p},\quad \mathbf{R} \in \mathrm{C}(\mathbf{P})\,. \tag{8}$$

Subproblem (7) can be solved by the maximization of each of the summation elements separately. Since $(U_m - \lambda_m r_m)$ is a concave function in r, we can find the optimal traffic demand by making first derivative of (1) equal to zero:

$$dU_m / dr_m + \xi_m - \lambda_m = 0\,.$$

Since $\sum_{j=1}^{M} r_j (l-1) = \text{const}$ at the iteration l, it is not taken into account in the derivation. Thus, the optimal traffic demand can be determined by the following expression:

$$r_m{}^* = -\frac{1}{b} \ln \frac{\lambda_m - \xi_m}{ab}\,. \tag{9}$$

To solve (8) we must relax the power constraint $\mathbf{P1} \le \mathbf{p}$. The Lagrangian expression for this step becomes

$$Q = \sum_{m=1}^{M} \lambda_m \sum_{n=1}^{N} R_{mn} + \boldsymbol{\mu}^{\mathbf{T}} (\mathbf{p} - \mathbf{P1}) =$$
$$= \sum_{m=1}^{M} \lambda_m \sum_{n=1}^{N} R_{mn} + \sum_{k=1}^{K+1} \mu_k (p_k - P_{kn})\,,$$

where $\boldsymbol{\mu}$ is a vector dual variable for physical layer. It represents the price user has to pay for violation of his power constraint.

Now physical-layer dual function can be again decoupled to N per-tone subproblems

$$\max_{\bar{\mathbf{P}}_n, \bar{\mathbf{R}}_n} \quad \left(\sum_{m=1}^{M} \lambda_m R_{mn} - \sum_{k=1}^{K+1} \mu_k P_{kn}\right)$$
$$\text{s.t.}\quad \bar{\mathbf{R}}_n \in \mathrm{C}(\bar{\mathbf{P}}_n)\,, \tag{10}$$

where $\bar{\mathbf{P}}_n$ and $\bar{\mathbf{R}}_n$ denote vectors formed by nth columns of matrices $\mathbf{P}$ and $\mathbf{R}$ respectively.

The transmission can take either one or two time-slots depending on which the transmission strategy is used. Direct-channel (DC) strategy realizes transmission through source-destination link (σ — δ). When the relay is used, either decode-and-forward or amplify-and-forward strategies can be performed. In this case the transmission is realized through source-relay-destination link (σ — ρ — δ) and takes two time-slots, as shown in Fig. 3. It should be noted that either user's terminal or base transceiver station can be a source or a destination however the base transceiver station can never be the relay.

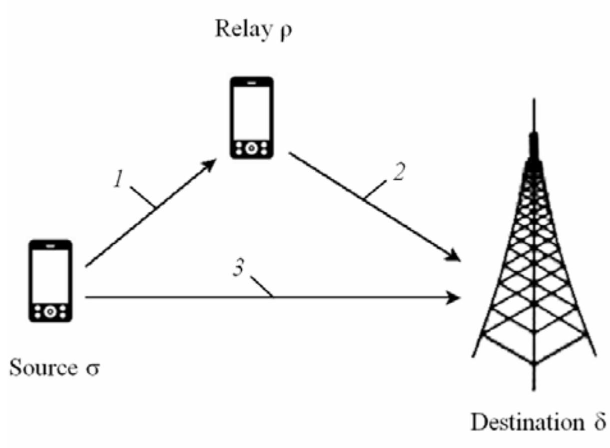

Fig. 3. Transmission from source σ to destination δ using relay ρ: (*1*) source-relay link with channel gain $h_{\sigma\rho}$; (*2*) relay-destination link with channel gain $h_{\rho\delta}$; (*3*) source-destination link with channel gain $h_{\sigma\delta}$.

Since in each tone there is only one active stream we can modify (10) and formulate the per-tone optimization subproblem:

$$\max \quad [\lambda_m R_{mn} - (\mu_\sigma P_{\sigma n} + \mu_\rho P_{\rho n})]$$
$$\text{s.t.}\quad \bar{\mathbf{R}}_n \in \mathrm{C}(\bar{\mathbf{P}}_n)\,, \tag{11}$$

where μ_σ and μ_ρ denote corresponding dual variables for source σ and relay ρ.

To find the optimal solution of this subproblem we realize an exhaustive search over all possible streams m, choices of the relay node, all possible R_{mn}, $P_{\sigma n}$ and $P_{\rho n}$.

Now we can derive the relations between the rate and the power at the node for a discrete bit-loading. When using direct-channel transmission strategy we have

$$P_{\sigma n} = (2^{R_{mn}} - 1) \frac{N_0 W}{|h_{\sigma\delta}|^2}\,,$$

where N_0 is the spectral density of noise in the channel; W is the bandwidth of one frequency tone; $h_{\sigma\delta}$ is the channel gain of the source-destination link.

We can obtain the formulae for the direct-channel strategy according to [3]. It has to use two time-slots for the transmission. Also, from [3] it is known that this strategy is feasible only for the case where $|h_{\sigma\rho}| \ge |h_{\rho\delta}|$, i.e. the source-relay link is stronger than the relay-destination link.

Denote an objective function of expression (10) as $f \overset{\text{def}}{=} \lambda_m R_{mn} - (\mu_\sigma P_{\sigma n} + \mu_\rho P_{\rho n})$. Then DC strategy is feasible only if $\partial f / \partial P_{\sigma n} \le 0$. The power expanded at the source node can be expressed as

$$P_{\sigma n} = (2^{R_{mn}} - 1)\frac{N_0 W}{|h_{\sigma\rho}|^2}.$$

It can be shown that the minimum of $P_{\rho n}$ is defined by the following expression:

$$P_{\rho n} = \frac{(2^{R_{mn}} - 1)N_0 W - P_{\sigma n}|h_{\sigma\delta}|^2}{|h_{\rho\delta}|^2}.$$

Similarly, amplify-and-forward strategy uses two times-slots for the transmission. In this case the power expanded by the relay node is given by

$$P_{\rho n} = \frac{(c_1 P_{\sigma n} + c_2)(c_3 P_{\sigma n} + c_4)}{(c_5 P_{\sigma n} + c_6)},$$

where $c_1 = |h_{\sigma\delta}|^2$; $c_2 = -(2^{R_{mn}} - 1)N_0 W$; $c_3 = |h_{\sigma\rho}|^2$; $c_4 = N_0 W$; $c_5 = -|h_{\rho\delta}|^2 (c_1 + c_3)$; $c_6 = c_2 / |h_{\rho\delta}|^2$. Detailed explanations of these values and their estimations are presented in [3].

To solve the optimization problem, we use the subgradient method developed by Shor [10]. It is suitable for convex optimization problems and convergent even when applied to non-differentiable objective functions.

According to [11], the subgradient for the dual function $g(\boldsymbol{\lambda})$ should be selected on the basis of the constraint of the primal problem. For example, the update of $\boldsymbol{\lambda}$ at iteration $l+1$ should be realized as

$$\boldsymbol{\lambda}(l+1) = [\boldsymbol{\lambda}(l) + \nu(l)(\mathbf{r}^* - \mathbf{R1}^*)]^+,$$

where $\nu(l)$ is the appropriately chosen step size; $\mathbf{r}^*$ is the vector of optimal traffic demand found in (9); $\mathbf{R1}^*$ is the vector of the optimal sum rate obtained by solution of the physical layer subproblem (8); $[x]^+ = x$ if $x \geq 0$ and $[x]^+ = 0$ otherwise.

According to [11], the subgradient search direction suggests that the mth component of $\boldsymbol{\lambda}$ should increase if the mth component of $\mathbf{R1}^*$ exceeds r_m^*, and decrease otherwise. The subgradient method is guaranteed to converge under diminishing step size rule [12].

Similarly, the updates for $\boldsymbol{\xi}$ and $\boldsymbol{\mu}$ can be written as

$$\boldsymbol{\xi}(l+1) = [\boldsymbol{\xi}(l) + \theta(l)(\boldsymbol{\omega}\sum_{j=1}^{M} r_j(l-1) - \mathbf{r}^*)]^+$$

$$\boldsymbol{\mu}(l+1) = [\boldsymbol{\mu}(l) + \varepsilon(l)(\mathbf{P1}^* - \mathbf{p})]^+.$$

Here $\theta(l)$ and $\varepsilon(l)$ are appropriately chosen step sizes corresponding to $\boldsymbol{\mu}$ and $\boldsymbol{\xi}$ respectively; $\mathbf{r}^*$ is the vector of optimal rate allocation which should be not smaller than a vector of certain percentage $\boldsymbol{\omega}$ of the total rate of all the streams $\sum_{j=1}^{M} r_j(l-1)$; $\mathbf{P1}^*$ is the vector of row sum of powers corresponding to the optimal power allocation which should be not higher than the vector of power constraints $\mathbf{p}$.

Simulation results

In order to illustrate the performance of the joint optimization algorithm developed in previous sections we simulate a cellular network with $K = 2$ nodes and a BTS shown in Fig. 2. Multiple access is organized via OFDM with $N = 50$ frequency tones, each tone with a bandwidth of $W = 375$ kHz. We use exponential utility function (1) with parameters $a = 10$ and $b = 1.84 \cdot 10^{-8}$ for downlink, $a = 1$ and $b = 1.84 \cdot 10^{-7}$ for uplink to realize the preference to the downlink. Each node has a power constraint of 23 dBm and the BTS has a constraint of 30 dBm. The channel can be decomposed into a small-scale Rayleigh fading and a large-scale path loss component with loss exponent of 3.7. We also assume the spectral density of the channel noise $N_0 = -174$ dBm/Hz.

At first we carry out the simulation of a wireless cellular network with users' terminals situated at the same distance from the BTS. Fig. 4 illustrates the results of simulation which show that the system searches for the optimal traffic demand $\mathbf{r}$ which tends to the achievable rate $\mathbf{R1}$ for the stream. In this case all the users are treated equally well.

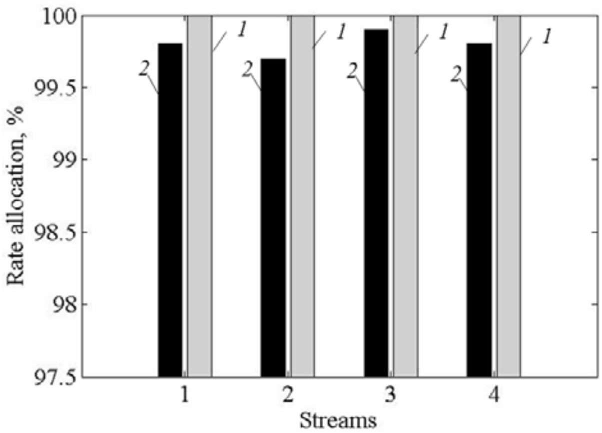

Fig. 4. Rate allocation for equidistant case nodes from BTS: (*1*) achievable rate; (*2*) allocated rate.

Our next simulation captures the opposite situation when node 1 is now closer to the BTS. In this case no fairness is provided, and the usage of node 2 decreases to zero. The simulation results depicted in Fig. 5 show that streams 2 and 4 have no rate at all. However, when introducing the additional constraint (3), we observe that system provides a minimal rate of several percents of the achievable at node 2 (streams 2 and 4), which is enough to start up some slow applications, e.g. e-mail service, sms, etc.

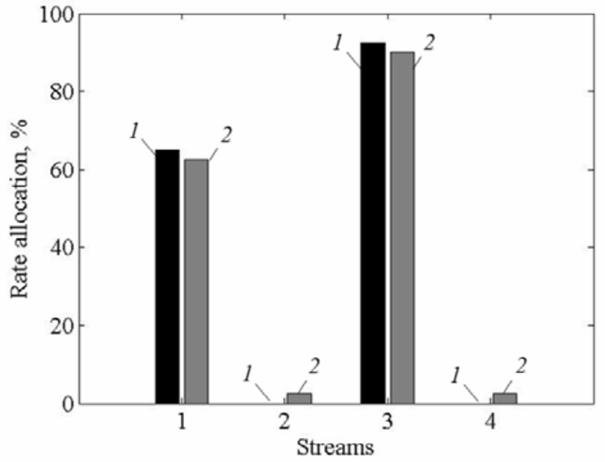

Fig. 5. Rate allocation for the case, where node 1 is closer to BTS than node 2: (*1*) unfair case; (*2*) fair case.

The results illustrate that the proposed algorithm is able to provide fair distribution of the quality of service among the users in a cellular network in terms of the minimal guaranteed rate.

Conclusions

In this work, cooperative wireless cellular networks and their performance optimization were studied. Cooperative communication provides the diversity for one-antenna cellular mobile stations and improves the achievable rate of the whole system. Using cooperation one can take advantage of unutilized power of the terminals which are temporarily situated close to the BTS.

To provide the cooperation into cellular networks, an optimization of the best relay node, the best relay strategy, and the best power, bandwidth and rate allocation has been utilized. Dual decomposition techniques for optimization of the performance have been used as well.

The existing algorithm of joint optimization of relay strategies and resource allocation is found out to be sensitive to users' positions. When one of the users is close to BTS and another one is quite far from it, the former will achieve the entire available rate while the latter can achieve no rate at all.

When introduced fairness constraint the algorithm is improved in terms of guaranteed bit-rate for all the users in an OFDMA-based network. Even the users which are situated far from BTS are provided with a certain rate which is higher than some fixed percentage of the total available bit-rate in the network. This leads to fair quality of service distribution among nodes.

By utilizing such cross-layer approach we are able to take into account resource allocation for users, cooperative relaying and proportional fairness and provide significant improvements of system performance.

The results can be applied to radio resource management and path selection in realistic wireless cellular networks in different scenarios. It may be useful for cognitive radio networks as well.

Possible directions of further work may include stochastic optimization of the average system performance, multi-hop relaying and routing in the network, multi-cell and topology based analysis, etc.

References

1. Nosratinia A., Hunter T. E., Hedayat A. Cooperative communication in wireless networks // Communications Magazine, IEEE. — 2004. — Vol. 42, issue 10. — P. 74—80.

2. Laneman J. N., Wornell G. W., Tse D. N. C. An efficient protocol for realizing cooperative diversity in wireless networks // Proceedings of IEEE International Symposium on Information Theory, Washington, USA, June 2001. — P. 294.

3. Ng T. C. Y, Yu W. Joint optimization of relay strategies and resource allocations in a cooperative cellular network // IEEE Journal on Selected Areas in Communications. — 2007. — Vol. 25, N. 2. — P. 328—339.

4. Fazel K., Fettweis G. P. Multi-carrier spread spectrum and related topics (3rd ed.). — Kluwer Academic Publisher, 2002. — 376 p.

5. Cover T. M., Thomas J. A. Elements of information theory (2nd ed.). — Wiley Interscience, 2006. — 748 p.

6. Saraydar C. U., Mandayam N. B., Goodman D. J. Efficient power control via pricing in wireless data networks // IEEE Transactions on Communications. — 2002. — Vol. 50, N. 2 — P. 291—303.

7. Lin X., Shroff N. B., Srikant R. A tutorial on cross-layer optimization in wireless networks // IEEE Journal on Selected Areas in Communications. — 2006. — Vol. 24, N. 8. — P. 1452—1463.

8. Yu W., Lui R. Dual methods for non-convex spectrum optimization of multicarrier systems // IEEE Transactions on Communications. — 2006. — Vol. 54, N. 7. — P. 1310—1322.

9. Palomar D., Chiang M. A tutorial to decompositon methods for network utility maximization // IEEE Journal on Selected Areas in Communications. — 2006. — Vol. 24, N. 8. — P. 1439—1450.

10. Shor N. Z. Minimization methods for non-differentiable functions // Springer Series in Computational Mathematics. — Springer-Verlag, 1985. — 162 p.

11. Boyd S. P., Vandenberghe L. Convex optimization. — Cambridge University Press, 2004. — 716 p.

12. Bertsekas D. P. Nonlinear Programming (2nd ed.). — Athena Scientific. — 802 p.

Received in final form October 29, 2010