EUROPEAN TRANSACTIONS ON TELECOMMUNICATIONS
Eur. Trans. Telecomms. 2011; **22**:296–314
Published online 15 June 2011 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/ett.1480

RESEARCH ARTICLE

Cross-layer optimisation for uplink transmission in OFDMA cellular networks with fixed relays

Lei You[1,2], Ping Wu[1]*, Mei Song[2], Junde Song[2] and Yong Zhang[2]

[1] Signals and Systems, Department of Engineering Sciences, Uppsala University, Uppsala, Sweden
[2] PCN and CAD Center, Beijing University of Posts and Telecommunications, Beijing, China

ABSTRACT

In this paper, we consider a cross-layer design aimed to enhance performance for uplink transmission in an orthogonal frequency division multiple-access (OFDMA)-based cellular network with fixed relay stations. Because mobile stations (MSs) spend most of the power on the uplink transmission, power efficiency resource allocation becomes very important to MSs. We develop a cross-layer optimisation framework for two types of uplink flows (inelastic and elastic flows) that have different quality-of-service requirements. For inelastic flows with fixed-rate requirement, we formulate the cross-layer optimisation problem as the minimisation of the sum transmission power of MSs under the constraints of flow conservation law, subcarrier assignment, relaying path selection and power allocation. For elastic flows with flexible-service-rate requirement, we consider the cross-layer trade-off between uplink service rate and power consumption of MSs and pose the optimisation problem as the maximisation of a linear combination of utility (of service rates) and power consumption (of MSs). Different trade-offs can be achieved by varying the weighting parameters. Dual decomposition and subgradient methods are used to solve the problems optimally with reduced computational complexity. The simulation results show that, through the proposed cross-layer resource optimisation framework and algorithms, significant benefits of deployment of multiple fixed relays in an OFDMA cellular network can be fully obtained such as reduction in power consumption, increase in service rate and energy savings in the uplink transmission of MSs.

KEY WORDS
cross-layer optimisation; OFDMA cellular networks; relays

***Correspondence**
P. Wu, Signals and Systems, Department of Engineering Sciences, Uppsala University, Uppsala, Sweden.
E-mail: Ping.Wu@angstrom.uu.se

Received 7 January 2010; Revised 15 February 2011; Accepted 14 March 2011

1. INTRODUCTION

Relay-based deployment has been viewed as one of the most promising architectures for next-generation cellular networks [1] because it can reduce subscribers' power consumption and deployment cost of the infrastructure, expand coverage of cells and also enhance system capacity and throughput in cellular networks. Therefore, the Institute of Electrical and Electronics Engineers (IEEE) 802.16 Working Group has created the Relay Task Group 802.16j [2] to add a relay functionality to the IEEE standard 802.16 and to develop an appropriate procedure for relaying operation.

In this paper, relay-based orthogonal frequency division multiple-access (OFDMA) cellular networks are considered. OFDMA has the capability of exploiting frequency-selectivity-enabled multi-user diversity by adaptive resource allocation. A deep faded subcarrier for one node may be favoured by another node. Therefore, the multi-user diversity may be exploited in a multi-user orthogonal frequency division multiplexing (OFDM) system if a subcarrier assignment for each user and power allocation for each subcarrier are appropriately adapted to the channel condition. Extensive research has been carried out to investigate resource allocation in traditional cellular networks without relays [3–7].

In order to attain the best performance, the optimal operation of a relay-enhanced OFDMA cellular network should be employed. This, however, is not an easy task as it involves so many different combinations of power and subcarrier allocations and path selections, especially when an adaptive modulation and coding is used, and quality-of-service (QoS) requirements of different flows should be considered. In addition, because a mobile station (MS) is a power-limited and energy-limited device, power efficiency resource management and rate control strategies for the uplink transmission are important to increase the lifetime of MSs. Although extensive research on relay-based OFDMA cellular networks has been performed, the problems related to all the above aspects are not optimally solved yet. For example, in [8], subcarrier and power allocations to maximise system capacity in OFDMA relay cellular networks are considered, and a suboptimal approach dividing the problem into two heuristic steps is adopted. In [9], the optimal source, relay and subcarrier allocation problem with the fairness constraint on relays is solved using a graph theoretical approach. However, it is assumed that fixed power is allocated to each subcarrier, which cannot yield optimal power control.

Recently, a dual decomposition method was proven to be a computationally efficient method to obtain optimal solution in the resource allocation of multicarrier systems [10, 11]. Related work using such a method in an OFDMA cellular system includes weighted sum rate maximisation and weighted sum power minimisation for downlink in traditional cellular networks [6], utility maximisation in user cooperation cellular networks [12] and sum rate maximisation for downlink in a relay-based OFDMA system [13]. In [12], optimal resource allocation and relay strategy (amplify-and-forward and decode-and-forward) selection for a user cooperation cellular network, in which subscribers can cooperatively forward data for each other while using the same OFDMA subcarrier as a source node, are presented under a utility maximisation framework [14, 15]. In [13], a similar problem for joint subcarrier and power allocation as that in [8] is formulated and solved by making continuous relaxation and using a dual decomposition method. However, their objective is to maximise the sum rate for the downlink, and they do not consider the QoS requirements of flows from upper layer, which would result in unfairness in resource allocation. In addition, because MSs spend little power on the transmission in the downlink case (a very small amount of power consumption for reception and signal processing), power efficiency of MSs in this case is not an issue considered.

In [16], the average achievable rate and the average power consumption on the uplink transmission of cooperative OFDMA cellular networks with relay nodes, which have little concern about their power consumption, are analytically examined. Although it focused on a cooperative relaying scheme, the analytical results showed that the benefits in increasing average transmission rate and reducing average power consumption of MSs from the deployment of multiple relay nodes are significant. However, it did not address how to get those benefits and how to achieve the best trade-off between those benefits through intelligent and optimal resource allocation.

In this paper, we develop a cross-layer resource optimisation framework for the power efficiency of MSs and flows with different QoS requirements in the uplink transmission in OFDMA cellular networks with dedicated relays, which are assumed to be fixed and have unlimited energy but have maximum transmission power limit. First, we consider inelastic flows with a specific rate requirement (e.g. voice over Internet Protocol services) and minimise the sum power consumption of MSs by optimally assigning subcarriers on direct and relaying uplinks and allocating power of RSs and MSs to subcarriers. Then, we consider elastic flows with a flexible service rate (e.g. best-effort and non-real-time service) and investigate the cross-layer trade-off between maximising the sum utility of uplink service rate of elastic flows and minimising the power consumption of MSs by fully utilising the relay nodes and the resource available. Thanks to the time-sharing property of multicarrier systems analysed and shown in [11], the problem has zero duality gap, which enables us to solve it almost optimally in its dual domain using a dual decomposition approach and a subgradient iteration algorithm with reduced computational complexity.

It should be pointed out that our work and the existing literature, although both are concerned with resource allocation of relay-based OFDMA cellular networks, deal with different scenarios and use different objectives in optimisation and thus develop different algorithms and present the results in different ways. Specifically, in the framework of our work, we minimise sum power consumption for uplink inelastic flows, whereas the existing works consider maximising the sum rate. Thus, the rates of flows are fixed in our model whereas total power is given in other authors' model. Moreover, we consider the trade-off between power consumption and utility of the services. As far as we know, no research works have considered this trade-off in their algorithms when dealing with resource allocation of OFDMA cellular networks with fixed relays; and the existing methods cannot be directly used to deal with the trade-off between rate and power. It is a consensus that both elastic and inelastic flows will co-exist in future wireless communication networks. Thus, the consideration of the two types of flows in network optimisation is very important and necessary.

The remainder of this paper is organised as follows. The system model considered is described in Section 2. Then, the resource optimisation and solution for inelastic flows and power minimisation are discussed in Section 3, and resource optimisation and solution for the trade-off between uplink service rate of elastic flows and power consumption of MSs are presented in Section 4. After simulation results are discussed in Section 5, the paper is concluded in Section 6.

2. SYSTEM MODEL

In this section, we will present a system model for uplink transmission in OFDMA cellular networks with two-hop fixed relays.

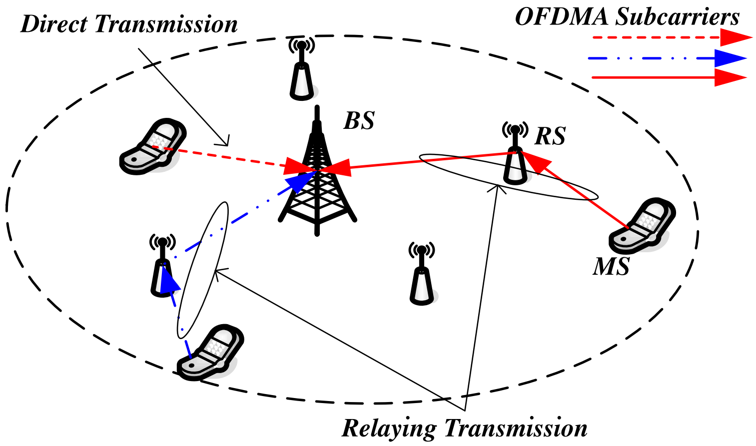

Figure 1. Uplink transmission model in the OFDMA cellular network with two-hop fixed relays. BS, base station; RS, relay station; MS, mobile station.

2.1. Network model and uplink frame structure

We consider the uplink transmission in an OFDMA relay-enhanced cellular network as shown in Figure 1. In each cell, there are a base station (BS) at the centre and K fixed relay stations (RSs) evenly located around the BS. In an uplink transmission, an MS can send signals either directly to the BS (called direct transmission, DT) or indirectly to the BS with two hops with the help of one of the K RSs (named relaying transmission, RT). In this case, there are $K+1$ possible paths for an MS to communicate with the BS for an uplink transmission. OFDMA subcarriers can be allocated separately to links between those nodes. The network model described here can be used to model the uplink transmission of an IEEE 802.16j relay-based network [2].*

We assume a half-duplex operation of RSs on each subcarrier because of radio limitation. To guarantee proper transmission, we design a special uplink transmission frame with two subframes as shown in Figure 2. Each uplink frame includes two subframes. In the first subframe, MSs transmit data to RSs or directly to BS. In the second subframe, RSs transmit the data they received from MSs in the first subframe to BS, although it is also possible for MSs to transmit directly to BS. Each subframe may use several OFDMA subcarriers in the frequency domain. To avoid interference, we impose that an OFDMA subcarrier can only be assigned to one of the uplink links MS–RS, MS–BS and RS–BS in any uplink subframe. We assume that the lengths of the two subframes are the same. Such an uplink frame structure is slightly different from the one proposed in the IEEE 802.16j MMR network [2], in which all MSs keep silent and only RSs are allowed to transmit in the second subframe. Our uplink frame structure can enable resource allocation to be more flexible.

We assume that wireless channels between nodes in the cellular are frequency-selective fading channels. OFDM technology divides the whole channel into many subcarriers so that each subcarrier experiences frequency-flat fading. We assume a slow-fading environment so that the channel remains unchanged during the resource allocation period. Full channel state information is known to the BS, which makes allocation decision in a centralised fashion and informs all RSs and MSs of the results of the resource allocation through a certain reliable control channel. A similar network model is also considered in [8] and [13]; however, it is focused on the downlink, whereas we consider the uplink case.

*In 802.16j, one cannot freely assign power to a single physical subcarrier. Subcarriers are grouped into subchannels. Thus, one cannot have direct access to the physical subcarriers but to the logical subchannels. This is helpful for the MAC layer to allocate resources to the subchannels. The assumption of subcarrier-wise resource allocation does not affect the optimisation algorithm proposed in this paper applying to 802.16 systems. In the case of subchannel-wise resource allocation, we can calculate the capacity of each subchannel using the average subchannel gain and the corresponding capacity of subchannels or links just as in the subcarrier case given in Sections 2.2 and 2.3 and then use the cross-layer algorithms in Sections 3 and 4 to allocate subchannels and power allocation.

2.2. Power allocation and transmission rate on subcarriers in physical layer

Assume that there are K RSs, labelled $\{1, ..., k, ..., K\}$, and M MSs randomly distributed in the cell, labelled $\{1, ..., m, ..., M\}$. The overall bandwidth B is divided into N OFDM subcarriers, labelled $\{1, ..., n, ..., N\}$. The channel coefficients of subcarrier n on the links MSm-to-BS, RSk-to-BS and MSm-to-RSk are $\gamma^n_{m,\mathrm{BS}}$, $\gamma^n_{k,\mathrm{BS}}$ and $\gamma^n_{m,k}$, respectively, the magnitudes of which follow a Rayleigh distribution with parameters $\sigma^2 = Cd^{-\alpha}$, where $\alpha \in [2, 6]$ is the path loss exponent of the channel, C is a constant and

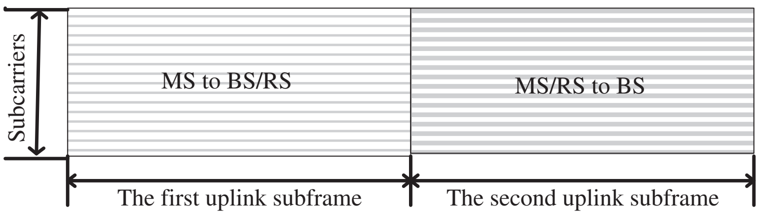

Figure 2. Frame structure for uplink transmission.

d indicates the distance between transmitting and receiving nodes. Consequently, the channel power gains $\left|\gamma_{m,\text{BS}}^{n}\right|^2$, $\left|\gamma_{k,\text{BS}}^{n}\right|^2$ and $\left|\gamma_{m,k}^{n}\right|^2$ follow the exponential distribution with mean $2\sigma^2$.

Let $p_m^{1,n}$ denote the power that MS m spends on subcarrier n during the first subframe. Let $p_m^{2,n}$ and $p_{k,\text{BS}}^{n}$ denote the power that MS m and RS k spend on subcarrier n during the second subframe, respectively. Note that an RS only spends power during the second subframe.

With the above parameters defined, we have the following transmission rate formulas for the subcarrier n:

(1) On link MSm-to-BS in the first uplink subframe

$$R_{m,\text{BS}}^{1,n} = W \log_2(1 + p_m^{1,n} \left|\gamma_{m,\text{BS}}^{n}\right|^2 / \Gamma W N_0)$$

(2) On link MSm-to-RS k in the first uplink subframe

$$R_{m,k}^{1,n} = W \log_2(1 + p_m^{1,n} \left|\gamma_{m,k}^{n}\right|^2 / \Gamma W N_0)$$

(3) On link MSm-to-BS in the second uplink subframe

$$R_{m,\text{BS}}^{2,n} = W \log_2(1 + p_m^{2,n} \left|\gamma_{m,\text{BS}}^{n}\right|^2 / \Gamma W N_0)$$

(4) On link RSk-to-BS in the second uplink subframe

$$R_{k,\text{BS}}^{2,n} = W \log_2(1 + p_{k,\text{BS}}^{n} \left|\gamma_{k,\text{BS}}^{n}\right|^2 / \Gamma W N_0)$$

where $W = B/N$ is the bandwidth of each subcarrier, Γ is the signal-to-noise ratio gap related to a targeted bit error rate and N_0 is the additive white Gaussian noise power spectral density, which is assumed to be the same for all the receiver nodes and subcarriers. In this case, given channel gains in the current frame, different power allocations may result in different transmission rates for subcarriers.

2.3. Subcarrier allocation and link layer rate

For the subcarrier allocation, we introduce binary indicators $d_{m,k}^{1,n}$, $d_{\text{MS}m}^{2,n}$ and $d_{\text{RS}k}^{2,n}$, which are explained below. Let $d_{m,k}^{1,n} = 1$ represent that subcarrier n is allocated to the link MSm-to-RSk or MSm-to-BS (when $k = 0$) in the first subframe, and $d_{m,k}^{1,n} = 0$ otherwise; let $d_{\text{MS}m}^{2,n} = 1$ represent that subcarrier n is allocated to the link MSm-to-BS in the second subframe, and $d_{\text{MS}m}^{2,n} = 0$ otherwise; let $d_{\text{RS}k}^{2,n} = 1$ represent that subcarrier n is allocated to the link RSk-to-BS in the second subframe, and $d_{\text{RS}k}^{2,n} = 0$ otherwise. Those binary indicators must satisfy

$$\sum_{k=0}^{K} \sum_{m=1}^{M} d_{m,k}^{1,n} = 1,\ d_{m,k}^{1,n} \in \{0,1\}, \forall n = 1, ..., n, ..., N \tag{1}$$

$$\sum_{k=1}^{K} d_{\text{RS}k}^{2,n} + \sum_{m=1}^{M} d_{\text{MS}m}^{2,n} = 1, \quad d_{\text{RS}k}^{2,n},\ d_{\text{MS}m}^{2,n} \in \{0,1\}, \forall n = 1, ..., n, ..., N \tag{2}$$

where Equation (1) means that for the first uplink subframe any subcarrier can only be assigned to one of the links MS–RS and MS–BS, whereas Equation (2) states that for the second uplink subframe any subcarrier can only be assigned to one of the links RS–BS and MS–BS.

With the above assumptions and conditions, we can formulate the aggregate rates on links MSm-to-BS and MSm-to-RSk in the first subframe and MSm-to-BS and RSk-to-BS in the second subframe, respectively, as follows:

$$T_{m,\text{BS}}^{1} = \sum_{n=1}^{N} d_{m,0}^{1,n} R_{m,\text{BS}}^{1,n}, \forall m = 1, \ldots, M$$

$$T_{m,k}^{1} = \sum_{n=1}^{N} d_{m,k}^{1,n} R_{m,k}^{1,n}, \forall k = 1, \ldots, K; \forall m = 1, \ldots, M$$

$$T_{m,\text{BS}}^{2} = \sum_{n=1}^{N} d_{\text{MS}m}^{2,n} R_{m,\text{BS}}^{2,n}, \forall m = 1, \ldots, M$$

$$T_{k,\text{BS}}^{2} = \sum_{n=1}^{N} d_{\text{RS}k}^{2,n} R_{k,\text{BS}}^{2,n}, \forall k = 1, \ldots, K$$

Thus, different subcarrier allocation policies result in different link layer rates. For convenience, let $\boldsymbol{d}$ denote the vector of the binary indicators for a specific subcarrier allocation policy, which is ordered as follows:

$$\boldsymbol{d} = \Big[d_{1,1}^{1,1}, \ldots, d_{1,1}^{1,n}, d_{2,1}^{1,1}, \ldots, d_{m,1}^{1,n}, d_{m,2}^{1,1}, \ldots, d_{m,k}^{1,n}, d_{\text{MS}1}^{2,1}, \ldots, d_{\text{MS}1}^{2,n}, d_{\text{MS}2}^{2,1}, \ldots, d_{\text{MS}m}^{2,n}, d_{\text{RS}1}^{2,1}, \ldots, d_{\text{RS}1}^{2,n}, d_{\text{RS}2}^{2,1}, \ldots, d_{\text{RS}k}^{2,n}\Big]$$

2.4. Flow conservation constraints for MSs and RSs

Here, we consider the cross-layer optimisation for two types of uplink flows from MSs to BS: inelastic and elastic flows, respectively. For inelastic flow, for example voice services, fixed service rate is usually required. Here, we assume that we have a reliable coding and a perfect admission control. Thus, we can ignore other requirements such as bit error rate and delay. For elastic flows such as non-real-time and best-effort services, which have no specific service rate requirements, some rate control schemes (e.g. TCP) are usually used to avoid network congestion and attain fairness.

We denote S_m (bits/frame) as the total service rate of the uplink (inelastic or elastic) flows from MSm to BS. For each MSm, the allocated aggregate uplink transmission rate in the two uplink subframes must be greater than or equal to S_m, and thus we have the following constraints:

$$S_m \leq T^1_{m,\mathrm{BS}} + \sum_{k=1}^{K} T^1_{m,k} \; + \; T^2_{m,\mathrm{BS}}, \; \forall m = 1, ..., M \quad (3)$$

For any RSk, the aggregate rate received in the first subframe must be less than or equal to the uplink rate of the link between the RSk and the BS in the second subframe. Thus, we have

$$\sum_{m=1}^{M} T^1_{m,k} \; \leq \; T^2_{k,\mathrm{BS}}, \; \forall k = 1, ..., K \quad (4)$$

3. RESOURCE OPTIMISATION FOR INELASTIC FLOWS AND POWER EFFICIENCY OF MOBILE STATIONS

In this section, we consider uplink inelastic flows with a fixed required service rate.

In the cellular networks deployed with special fixed RSs for the purpose of performance enhancement, each RS has a maximum power limitation although we have no concern about the energy expenditure of RSs. Letting $P_k^{\max}$ be the maximum power of RSk, we have the following constraints:

$$\sum_{n=1}^{N} p^n_{k,\mathrm{BS}} \; \leq \; P_k^{\max}, \;\; \forall k = 1, ..., K \quad (5)$$

However, for MSs, we need to concern more about their power and energy consumption. It is well known that transmission power should increase exponentially with distance to the receiver so that similar transmission rates could be attained. The deployment of relay nodes can significantly reduce transmission power of the MSs far away from the BS in the cell for the same uplink rate requirement [1]. So that power efficiency transmission of inelastic flows under the network model assumption described in Section 2.1 can be achieved, the choice of the MS to relay or not should depend not only on its location but also on the service rate required by its inelastic flow. This will be shown in a simple example in the following section.

3.1. A simple scenario with a single carrier

We consider a scenario in which a cell with a radius of 2000 m consists of one BS, one RS and one MS as shown in Figure 3, and the distance between the RS and the BS is 1400 m. We assume that only one carrier is available, the channel gain only includes a large scale path loss component with path loss exponent of 4, the bandwidth of the channel is $W = 100$ kHz, $\Gamma = 1$, and $N_0 = -174$ dBm.

From the network model described in Section 2.1, the MS can use DT and transmit data directly to BS in both uplink subframes or use RT and transmit data to the RS in the first subframe, and then the RS forwards the data to the BS in the second subframe. For the RT case, we assume that the RS has sufficient power to transmit to the BS all the data coming from the MS. Thus, the MS spends its power in both subframes for DT, whereas it just spends power in the first subframe for RT.

We calculate the power consumption of the MS using the relations given in Section 2.2. The results are shown in Figure 4 when the required data rate is 0.5 Mbit/s and Figure 5 when the required data rate is 1 Mbit/s. From Figures 4 and 5, we can see the following:

(1) The deployment of RSs can significantly reduce the level of uplink transmission power of the MS when it is far away from the BS, especially when it is located near the boundary of the cell;
(2) So that as much power as possible is saved, how an MS chooses a transmission scheme (DT or RT) should depend not only on its position in the cell but also on the service rate required by the inelastic flow. For example, when the MS is located at a distance between about 800–950 m to the BS, it should choose RT scheme when the required service rate is 0.5 Mbit/s (see Figure 4), but DT when the required service rate is 1 Mbit/s (see Figure 5).

This simple example shows that whether an MS chooses relaying or not depends on its location and the service rate required by its inelastic flow. In a more realistic scenario involving multicarriers with frequency-selective fading

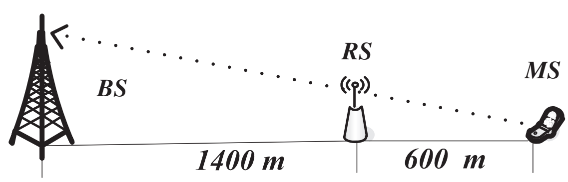

Figure 3. A simple scenario.

Eur. Trans. Telecomms. **22**:296–314 (2011) © 2011 John Wiley & Sons, Ltd.
DOI: 10.1002/ett

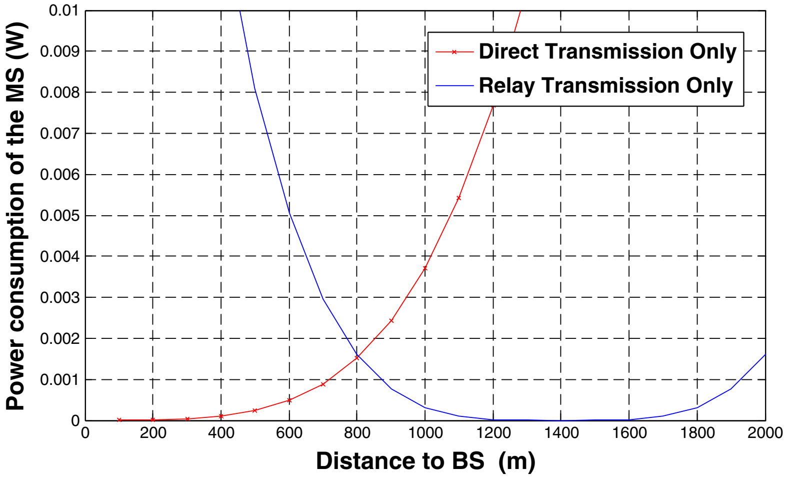

Figure 4. Power consumption of the MS at different locations and with different uplink transmission schemes (DT or RT) when the required service rate of the flow in the MS is 0.5 Mbit/s.

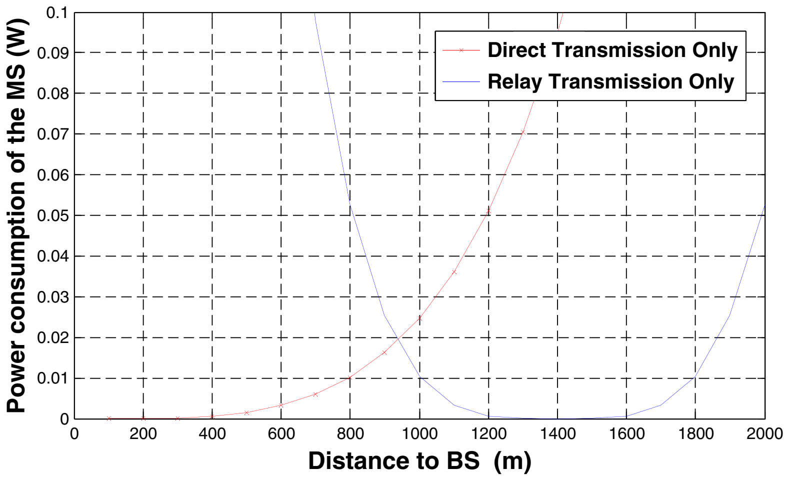

Figure 5. Power consumption of the MS at different locations and with different uplink transmission schemes (DT or RT) when the required service rate of the flow in the MS is 1 Mbit/s. (Note that the vertical scale is ten times of that in Figure 4).

channels, multiple RSs and MSs, the optimal operation of the whole network to minimise power consumption of all the MSs will become much more complicated because it also depends on the channel gain of subcarriers between different nodes, power allocation as well as different rate requirements from all MSs. In addition, when the rate requirement is high, an RS may reach its maximum power, which will limit its capacity for relaying. In the following subsection, we will tackle such a cross-layer resource allocation problem using nonlinear optimisation techniques.

3.2. Primal problem formulation, dual problem and subgradient method

As in the previous analysis, our objective is to find the optimal resource allocation solution to minimise the sum of power consumption of all the MSs in the two uplink subframes while satisfying the rate requirements from all the inelastic flows under the constraints described in the system model. The optimisation problem can be formulated as follows:

$$\underset{\mathbf{P}_{\mathrm{MS}},\mathbf{P}_{\mathrm{RS}},\mathbf{d}}{\text{Minimise}} \sum_{m=1}^{M} \left(\sum_{t=1}^{2} \sum_{n=1}^{N} p_m^{t,n} \right) \quad \text{(P1)}$$
Subject to Equations (1)–(5)

The minimisation in P1 is taken with respect to $\mathbf{P}_{\mathrm{MS}}$, $\mathbf{P}_{\mathrm{RS}}$ and $\mathbf{d}$ where $\mathbf{P}_{\mathrm{MS}}$ and $\mathbf{P}_{\mathrm{RS}}$ are the vectors of power allocation for all MSs and RSs, respectively. Note that the service rates $S_m, \forall m$ in constraint (3) of P1 are known and fixed for all MSs. The optimisation problem P1 is an NP (non-deterministic polynomial-time)-hard combination optimisation problem with non-linear constraints [17], and finding its optimal solution involves an exhaustive search over all possible transmission schemes, RS selection, subcarrier assignment policies and power allocations in the two uplink subframes, which is difficult to determine within a designated time, especially when the number of subcarriers is large because the dimension of the set of potential subcarrier assignment policies **d** increases exponentially with the number of subcarriers [6, 17].

It can be observed that Equations (3)–(5) are the coupling constraints of problem P1 between two subframes and among different subcarriers. By relaxing these coupling constraints using the Lagrange multiplier technique, we can decouple the problem into several subproblems that can be solved with low computational complexity given the Lagrange multipliers [10, 11, 18], and the optimal solution of the primal problem can be recovered optimally using a gradient/subgradient method in its dual domain if a strong duality holds [19].

However, because of binary variables in constraints (1) and (2), P1 is a mixed integer programming problem and not a typical convex programming problem. Thus, a strong duality may not hold. We can relax this integer constraint to a continuous one (i.e. let $d_{m,k}^{1,n}$, $d_{\mathrm{MS}m}^{2,n}$ and $d_{\mathrm{RS}k}^{2,n} \in [0,1]$), which corresponds to permitting the time sharing of subcarrier allocation policies, and can change the problem to a convex one so that a strong duality holds. This method needs to implement the resulting solution in multiple frames as in the TDMA (Time Division Multiple Access)-based stationary networks [5, 20, 21], which is impractical in the case of mobile cellular communication where the OFDMA channel varies from frame to frame [5].

Recently, it was discovered that the time-sharing condition under which the duality gap is zero is always satisfied in OFDM systems in the limit as the number of subcarriers goes to infinity as analysed and proven in [11]. The reason is, roughly speaking, that in practical OFDM systems with a large number of subcarriers, channel conditions in adjacent subcarriers are often similar. Then, the time sharing of each subcarrier may be approximately implemented with frequency sharing of these adjacent subcarriers [6, 11, 12]. Thanks to this result, we argue that the duality gap of P1 is approximately zero and that the dual method can still be used to solve P1 optimally.

By introducing three vectors of Lagrange multipliers, $\boldsymbol{\lambda} = [\lambda_1, ..., \lambda_M]^T$, $\boldsymbol{\mu} = [\mu_1, ..., \mu_k]^T$ and $\boldsymbol{\epsilon} = [\varepsilon_1, ..., \varepsilon_k]^T$, to relax coupling constraints (3), (4) and (5), we can write the corresponding partial Lagrangian as follows:

$$
\begin{aligned}
&L_1(\mathbf{P}_{\mathrm{MS}}, \mathbf{P}_{\mathrm{RS}}, \mathbf{d}; \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\epsilon})\\
&= \sum_{m=1}^{M}\left(\sum_{t=1}^{2}\sum_{n=1}^{N} p_m^{t,n}\right)\\
&+ \sum_{m=1}^{M}\lambda_m\left(S_m - T_{m,\mathrm{BS}}^1 - \sum_{k=1}^{K} T_{m,k}^1 - T_{m,\mathrm{BS}}^2\right)\\
&+ \sum_{k=1}^{K}\mu_k\left(\sum_{k=1}^{M} T_{m,k}^1 - T_{k,\mathrm{BS}}^2\right)\\
&+ \sum_{k=1}^{K}\varepsilon_k\left(\sum_{n=1}^{N} p_{k,\mathrm{BS}}^n - P_k^{\max}\right)
\end{aligned}
\tag{6}
$$

where $\mathbf{P}_{\mathrm{MS}}$ and $\mathbf{P}_{\mathrm{RS}}$ are the vectors of power allocation for all MSs and RSs, respectively. Then, with this Lagrangian, we can define the dual objective function as

$$
D(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\epsilon}) = \begin{cases} \min\limits_{\mathbf{P}_{\mathrm{MS}}, \mathbf{P}_{\mathrm{RS}}, \mathbf{d}} L_1(\mathbf{P}_{\mathrm{MS}}, \mathbf{P}_{\mathrm{RS}}, \mathbf{d}; \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\epsilon}) \\ s.t. \;\; (1)\;(2) \end{cases} \tag{7}
$$

Thus, the dual problem can be given as:

$$
\begin{aligned}
&\max \; D(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\epsilon})\\
&s.t.\; \boldsymbol{\lambda} \geq 0, \; \boldsymbol{\mu} \geq 0, \; \boldsymbol{\epsilon} \geq 0
\end{aligned}
\tag{D1}
$$

where Lagrange multipliers $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$ and $\boldsymbol{\epsilon}$ become the dual variables in dual problem D1. Because of the zero duality gap, that is, a strong duality holds, the solution of primal problem P1 can be recovered by solving its dual problem D1. The most important advantage of solving the primal problem in its dual domain is the decomposability of the dual function, with which we can decouple the coupling constraints, decompose the problem into several subproblems and solve them separately with low complexity [18]. As shown in the next subsection, we will show that D1 can be divided into separate per-subcarrier subproblems for each uplink subframe, which can significantly reduce the complexity of the problem.

Because the dual objective function $D(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\epsilon})$ is not a differentiable function, we can solve the dual problem using the subgradient method [22, 23]. To find the subgradient using the definition in [22, 23], we first convert D1 into the following equivalent convex optimisation problem with convex objective function.

$$
\begin{aligned}
&\min \; -D(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\epsilon})\\
&s.t.\; \boldsymbol{\lambda} \geq 0, \; \boldsymbol{\mu} \geq 0, \; \boldsymbol{\epsilon} \geq 0
\end{aligned}
\tag{D1$'$}
$$

Then, we can find a subgradient of the convex objective function in D1$'$ by definition with the following Lemma.

Lemma 1. *Considering the convex optimisation problem D1$'$ and assuming that $T_{m,\mathrm{BS}}^{1\;*}$, $T_{m,k}^{1\;*}$, $T_{m,\mathrm{BS}}^{2\;*}$, $T_{k,\mathrm{BS}}^{2\;*}$ and $p_{k,\mathrm{BS}}^{n\;*}$ are the optimal solution of minimisation in the function (7) for given $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$ and $\boldsymbol{\epsilon}$, then*

$$
g_m(\lambda_m) = T_{m,\mathrm{BS}}^{1\;*} + \sum_{k=1}^{K} T_{m,k}^{1\;*} + T_{m,\mathrm{BS}}^{2\;*} - S_m,
$$

$$
h_k(\mu_k) = T_{k,\mathrm{BS}}^{2\;*} - \sum_{k=1}^{M} T_{m,k}^{1\;*} \;\text{ and}
$$

$$f_k(\varepsilon_k) = P_k^{\max} - \sum_{n=1}^{N} p_{k,\mathrm{BS}}^{n\ *}$$

are the subgradients of $-D(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\epsilon})$ *at* λ_m, μ_k *and* ε_k*,respectively.*

The proof of Lemma 1 is presented in the Appendix at the end of the paper.

A subgradient update algorithm [23] for solving D1$'$ (and also D1) is stated as follows.

Given the optimal solution $T_{m,\mathrm{BS}}^{1\ *}$, $T_{m,k}^{1\ *}$, $T_{m,\mathrm{BS}}^{2\ *}$, $T_{k,\mathrm{BS}}^{2\ *}$ and $p_{k,\mathrm{BS}}^{n\ *}$ in the current iteration i, the algorithm updates dual variable in the following manner:

$$\begin{cases} \lambda_m(i+1) = \max(0,\ \lambda_m(i) - \sigma(i) g_m(\lambda_m(i))\ ),\forall m \\ \mu_k(i+1) = \ \max(0,\ \mu_k(i) - \phi(i) h_k(\mu_k(i))\ )\ ,\forall k \\ \varepsilon_k(i+1) = \ \max(0,\ \varepsilon_k(i) - \theta(i) f_k(\varepsilon_k(i))\ )\ ,\forall k \end{cases} \tag{8}$$

where $\sigma(i)$, $\phi(i)$ and $\theta(i)$ are the step sizes for the update of Lagrange multipliers $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$ and $\boldsymbol{\epsilon}$, respectively, in iteration i, until the algorithm converges.

According to [23], subgradient method is guaranteed to converge to the optimum if step sizes $\sigma(i)$, $\phi(i)$ and $\theta(i)$ are designed appropriately as follows:

Theorem 1. *Dual variables* $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$ *and* $\boldsymbol{\epsilon}$ *converge to the optimal dual solutions if the positive scalar step sizes* $\sigma(i)$, $\phi(i)$ *and* $\theta(i)$ *are chosen such that*

$$\lim_{i\to\infty} \sigma(i) = 0,\ \sum_{i=1}^{\infty} \sigma(i) = \infty;$$
$$\lim_{i\to\infty} \phi(i) = 0,\ \sum_{i=1}^{\infty} \phi(i) = \infty;$$
$$\lim_{i\to\infty} \theta(i) = 0,\ \sum_{i=1}^{\infty} \theta(i) = \infty$$

Remarks. Because a strong duality holds, the corresponding primal variables ($\boldsymbol{P}_{\mathrm{MS}}^*$, $\mathbf{P}_{\mathrm{RS}}^*$, $\mathbf{d}^*$) are the globally optimal variables of primal problem P1 for optimal dual variables ($\boldsymbol{\lambda}^*$, $\boldsymbol{\mu}^*$, $\boldsymbol{\epsilon}^*$).

3.3. Dual decomposition and subproblems solution

Each step of subgradient iteration algorithm requires the optimal solution of minimisation in the dual objective function (7) which can be divided into subproblems (9) and (10):

$$D(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\epsilon}) = D_1(\boldsymbol{\lambda},\boldsymbol{\mu}) + D_2(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\epsilon}) - \sum_{k=1}^{K} \varepsilon_k P_k^{\max}$$

where

$$D_1(\boldsymbol{\lambda},\boldsymbol{\mu}) = \min_{\mathbf{P}_{\mathrm{MS}}^1,\ \sum_{k=0}^{K}\sum_{m=1}^{M} d_{m,k}^{1,n}=1} \times \left(\sum_{m=1}^{M}\sum_{n=1}^{N} p_m^{1,n} - \sum_{m=1}^{M} \lambda_m \left(T_{m,\mathrm{BS}}^1 + \sum_{k=1}^{K} T_{m,k}^1 \right) + \sum_{k=1}^{K} \mu_k \sum_{m=1}^{M} T_{m,k}^1 \right) \tag{9}$$

$$D_2(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\epsilon}) = \min_{\mathbf{P}_{\mathrm{MS}}^2,\mathbf{P}_{\mathrm{RS}},\ \sum_{k=1}^{K} d_k^{2,n} + \sum_{m=1}^{M} d_m^{2,n}=1} \times \left(\sum_{m=1}^{M}\sum_{n=1}^{N} p_m^{2,n} - \sum_{m=1}^{M} \lambda_m T_{m,\mathrm{BS}}^2 - \sum_{k=1}^{K} \mu_k T_{k,\mathrm{BS}}^2 + \sum_{k=1}^{K} \varepsilon_k \sum_{n=1}^{N} p_{k,\mathrm{BS}}^n \right) \tag{10}$$

Given the dual variables, $D_1(\boldsymbol{\lambda},\boldsymbol{\mu})$ and $D_2(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\epsilon})$ are to determine resource allocation in the first and second uplink subframes, respectively.

$D_1(\boldsymbol{\lambda},\boldsymbol{\mu})$ can be further written as

$$D_1(\boldsymbol{\lambda},\boldsymbol{\mu}) = \sum_{n=1}^{N} \left(\min_{P_m^{1,n},\ \sum_{k=0}^{K}\sum_{m=1}^{M} d_{m,k}^{1,n}=1} \sum_{m=1}^{M}\sum_{k=0}^{K} \times \left[p_m^{1,n} - (\lambda_m - \mu_k) d_{m,k}^{1,n} R_{m,k}^{1,n} \right] \right) = \sum_{n=1}^{N} \left(D_1^n(\lambda_m,\mu_k) \right)$$

where $\mu_0 = 0$ and $R_{m,0}^{1,n} = R_{m,\mathrm{BS}}^{1,n}$, and can then be decomposed to N subproblems as follows,

$$D_1^n(\lambda_m,\mu_k) = \min_{P_m^{1,n},\ \sum_{k=0}^{K}\sum_{m=1}^{M} d_{m,k}^{1,n}=1} \sum_{m=1}^{M}\sum_{k=0}^{K} d_{m,k}^{1,n} \times \left[p_m^{1,n} - (\lambda_m - \mu_k) R_{m,k}^{1,n} \right],\quad \forall 1,\dots,n,\dots,N \tag{11}$$

Each subproblem in Equation (11) is to determine the assignment of one subcarrier to one of the links between MSs and RSs/BS and also the amount of power of the MSs spent on it. The constraints in those N subproblems are independent, and thus these subproblems can be solved separately. Because constraint $\sum_{k=0}^{K}\sum_{m=1}^{M} d_{m,k}^{1,n} = 1$ and binary variables $d_{m,k}^{1,n} \in \{0,1\}$ in each subproblem in Equation (11), the optimal solution of $D_1^n(\lambda_m, \mu_k)$ is that subcarrier n is allocated exclusively to the node pair (m^*, k^*) such that

$$[m^*, k^*] = \arg\min_{m,k}\ p_m^{1,n} - (\lambda_m - \mu_k) R_{m,k}^{1,n}$$
$$= \arg\min_{m,k}\left[p_m^{1,n} - (\lambda_m - \mu_k) R_{m,k}^{1,n}\right]\Bigg|_{p_m^{1,n} = \max\left(0,\ \frac{\lambda_m - \mu_k}{\log 2} - \Gamma W N_0 / \left|\gamma_{m,k}^n\right|^2\right)} \quad (12)$$

and power allocated to subcarrier n is

$$p_{m^*}^{1,n} = \max\left(0, \frac{\lambda_{m^*} - \mu_{k^*}}{\log 2} - \Gamma W N_0 / \left|\gamma_{m^*,k^*}^n\right|^2\right) \quad (13)$$

Then, we set $d_{m^*,k^*}^{1,n} = 1$ and $d_{m,k}^{1,n} = 0, \forall m \neq m^*, k \neq k^*$. In Equations (12) and (13), $k^* = 0$ means that subcarrier n is allocated to the link between MS m^* and BS in the first uplink frame. To solve Equation (12), we need $M^*(K+1) - 1$ comparisons.

Similarly, $D_2(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\varepsilon})$ can be further written as

$$D_2(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\varepsilon}) = \sum_{n=1}^{N} \min_{\substack{P_m^{2,n}, P_{k,\mathrm{BS}}^{2,n}, \\ \sum_{k=1}^{K} d_{\mathrm{RS}K}^{2,n} + \sum_{m=1}^{M} d_{\mathrm{MS}m}^{2,n} = 1}}$$
$$\times \sum_{m=1}^{M}\sum_{k=0}^{K} d_{\mathrm{MS}m}^{2,n}\left[p_m^{2,n} - \lambda_m R_{m,\mathrm{BS}}^{2,n}\right]$$
$$+ d_{\mathrm{RS}K}^{2,n}\left[\varepsilon_k p_{k,\mathrm{BS}}^{1,n} - \mu_k R_{k,\mathrm{BS}}^{1,n}\right]$$
$$= \sum_{n=1}^{N} D_2^n(\lambda_m, \varepsilon_k, \mu_k)$$

and thus be further decomposed to N subproblems,

$$D_2^n(\lambda_m, \varepsilon_k, \mu_k) = \min_{\substack{P_m^{2,n}, P_{k,\mathrm{BS}}^{2,n}, \\ \sum_{k=1}^{K} d_{\mathrm{RS}K}^{2,n} + \sum_{m=1}^{M} d_{\mathrm{MS}m}^{2,n} = 1}}$$
$$\times \sum_{m=1}^{M}\sum_{k=0}^{K} d_{\mathrm{MS}m}^{2,n}\left[p_m^{2,n} - \lambda_m R_{m,\mathrm{BS}}^{2,n}\right]$$
$$+ d_{\mathrm{RS}K}^{2,n}\left[\varepsilon_k p_{k,\mathrm{BS}}^{1,n} - \mu_k R_{k,\mathrm{BS}}^{1,n}\right]$$
$$\forall 1, \ldots, n, \ldots, N \quad (14)$$

Each subproblem in Equation (20) is to determine the assignment of subcarrier n to one of the links MSs–BS and RSs–BS and the amount of power of MS or RS spent on the subcarrier. The second uplink frame is equivalent to the uplink transmission in a traditional multi-user OFDMA cellular network with $M + K$ users [24]. Similar to the case of Equation (11), those N subproblems can be solved separately. From constraint $\sum_{k=1}^{K} d_{\mathrm{RS}K}^{2,n} + \sum_{m=1}^{M} d_{\mathrm{MS}m}^{2,n} = 1$ and binaryvariables $d_{\mathrm{RS}K}^{2,n}, d_{\mathrm{MS}m}^{2,n} \epsilon \{0,1\}$ in each subproblem in Equation (20), it follows that each subcarrier should be allocated to either one of MSs or one of RSs. Wefirst find, for subcarrier n, the optimal MS m^* and the corresponding optimal power allocation, respectively, in the following fashion:

$$m^* = \arg\min_{m}\left[p_m^{2,n} - \lambda_m R_{m,\mathrm{BS}}^{2,n}\right]\Bigg|_{p_m^{2,n} = \max\left(0,\ \lambda_m / \log 2 - \Gamma W N_0 / \left|\gamma_{m,\mathrm{BS}}^n\right|^2\right)} \quad (15)$$

$$p_{m^*}^{2,n} = \max\left(0,\ \lambda_{m^*} / \log 2 - \Gamma W N_0 / \left|\gamma_{m^*,\mathrm{BS}}^n\right|^2\right) \quad (16)$$

and then find the optimal RS k^* and the corresponding optimal power allocation, respectively, in the following manner:

$$k^* = \arg\min_{k}\left[\varepsilon_k p_{k,\mathrm{BS}}^{1,n} - \mu_k R_{k,\mathrm{BS}}^{1,n}\right]\Bigg|_{p_k^{2,n} = \min\left(\max\left(0,\ \frac{\lambda_k}{\varepsilon_k \log 2} - \Gamma W N_0 / \left|\gamma_{k,\mathrm{BS}}^n\right|^2\right),\ P_k^{\max}\right)} \quad (17)$$

Eur. Trans. Telecomms. **22**:296–314 (2011) © 2011 John Wiley & Sons, Ltd.
DOI: 10.1002/ett

$$p_{k^*}^{2,n} = \min\left(\max\left(0,\ \lambda_{k^*}/(\varepsilon_{k^*}\log 2) - \Gamma W N_0 / \left|\gamma_{k^*,\mathrm{BS}}^{n}\right|^2 \right),\ P_{k^*}^{\max} \right) \quad (18)$$

Then, from MSm^* and RSk^*, we determine the best one to which the subcarrier is allocated in the following way:

(1) If $(p_{m^*}^{2,n} - \lambda_{m^*} R_{m^*,\mathrm{BS}}^{2,n}) < (\varepsilon_{k^*} p_{k^*,\mathrm{BS}}^{1,n} - \mu_{k^*} R_{k^*,\mathrm{BS}}^{1,n})$, the subcarrier n is assigned to MSm^* with power allocation according to Equation (22), and $d_{\mathrm{MS}m^*}^{2,n} = 1$ and $d_{\mathrm{RS}K}^{2,n} = 0, \forall k, d_{\mathrm{MS}m}^{3,n} = 0, \forall m \neq m^*$ are set.

(2) If $(p_{m^*}^{2,n} - \lambda_{m^*} R_{m^*,\mathrm{BS}}^{2,n}) > (\varepsilon_{k^*} p_{k^*,\mathrm{BS}}^{1,n} - \mu_{k^*} R_{k^*,\mathrm{BS}}^{1,n})$, the subcarrier n is allocated to RSk^* with power allocation according to Equation (25), and $d_{\mathrm{RS}k^*}^{2,n} = 1$ and $d_{\mathrm{MS}m}^{2,n} = 0, \forall m, d_{\mathrm{RS}k}^{2,n} = 0, \forall k \neq k^*$ are set.

When the optimal subcarrier assignment policies and power allocation are determined, the values of $T_{m,\mathrm{BS}}^{1\ *}$, $T_{m,k}^{1\ *}$, $T_{m,\mathrm{BS}}^{2\ *}$ and $T_{k,\mathrm{BS}}^{2\ *}$ can be calculated according to the formulas for link layer rate given in Section 2.3.

3.4. Summary of the algorithm

We summarise the complete procedure of the whole algorithm in Algorithm 1 as follows.

Algorithm 1: Cross-layer resource optimisation for inelastic flows and power efficiency of MSs

(1) The BS collects service rate request from all MSs and channel state information through uplink control channels at the beginning of the frame. Then the BS initialises the dual variables $\lambda(0)$, $\mu(0)$ and $\varepsilon(0)$.
(2) Given $\lambda(t)$, $\mu(t)$ and $\varepsilon(t)$ in iteration t, the BS solves 2N per-subcarrier subproblems in Equations (11) and (20) to obtain the optimal subcarrier assignment and the power allocation in the first and second uplink subframes, respectively.
(3) The BS calculates $T_{m,BS}^{1\ *}$, $T_{m,k}^{1\ *}$, $T_{m,BS}^{2\ *}$ and $T_{k,BS}^{2\ *}$ based on the results obtained in step 2, calculates subgradients according to Lemma 1 and then updates the dual variables using Equation (8).
(4) Return to step 2 until the algorithm converges.
(5) The BS broadcasts the resulting subcarrier assignment policies $\mathbf{d}^*$ and power allocation $\mathbf{P}_{\mathrm{MS}}^*$ and $\mathbf{P}_{\mathrm{RS}}^*$ to all the MSs and RSs through downlink control channels.

4. CROSS-LAYER TRADE-OFF BETWEEN SERVICE RATE OF ELASTIC FLOWS AND POWER EFFICIENCY OF MOBILE STATIONS

In this section, we consider elastic flows in the uplink, which can have a flexible service rate. A proper rate control scheme is essential for elastic flows to avoid congestion and fairly utilise the available resource. Network utility maximisation (NUM) has been developed in [15, 25] to formulate joint resource allocation and rate control as an optimisation problem. By solving the NUM problem, we develop the algorithms to allocate the resource (bandwidth and power) available in Physical Layer (PHY) to achieve optimal rate control for each flow. A typical example for a multihop Code Division Multiple Access (CDMA) network is shown in [25]. Through the joint optimization of rate control and power allocation under the NUM framework, transport and physical layers are perfectly balanced, that is, the resource (power) in physical layer is fully utilised whereas the sum utility of source rate is maximised.

In cellular networks where MSs are power-limited and energy-limited devices, power efficiency rate control and resource allocation strategies become important issues. Because the service rate of elastic flow is flexible, we can get another balance between power efficiency of MS and service rate of elastic flow. We first get the fundamental trade-off between rate and energy consumption from Shannon formula for channel capacity. We assume a basic stationary channel with gain γ, bandwidth W, noise power spectral density N_0 and capacity gap Γ. The capacity of the channel with transmission power P is

$$C = W^* \log(1 + P\,|\gamma|^2 / \Gamma W N_0)$$

and the energy needed for transmitting one bit is

$$E_{\mathrm{bit}} = P/C = [(W/P)^* \log(1 + (P/W)\,|\gamma|^2 / \Gamma N_0)]^{-1}$$

It is easy to find that $E_{\mathrm{bit}} \to \ln(2)(\Gamma N_0/|\gamma|^2)$ as $W/P \to \infty$. This indicates that one can reduce energy per bit by reducing the power (which will lower the transmission rate) for a given bandwidth. Accordingly, for elastic flows from MSs, one can save energy for transmitting the same amount of data by reducing its service rate (through using lower transmission power level).

The above fact motivates us to develop another cross-layer optimisation framework and algorithm to balance the service rate of elastic flows and uplink transmission power of MSs. However, to get the best trade-off in OFDMA cellular networks with fixed RSs, we need to take into account the optimal uplink service rate, the full utilisation of the power of fixed RSs (which are not energy limited), proper selection of transmission path and assignment of OFDMA subcarriers for relaying links and direct links. To this purpose, we will pose a cross-layer trade-off optimisation problem by incorporating the utility of service rate of elastic flows as well as the power consumption of MSs into the objective of the problem and then solve it using the dual decomposition method.

4.1. Problem for trade-off between service rate of elastic flows and power efficiency of mobile stations

We assume that the utility function associated with elastic flows in MSm is $U_m(S_m)$, where S_m is the total service rate of all elastic flows from MSm to the BS. $U_m(S_m)$ is usually assumed to be a concave, non-decreasing and continuously differentiable function of S_m. There are many such utility functions available for different optimisation objectives. A class of utility functions for fair resource allocation [26] is

$$U^{\alpha}(z) = \begin{cases} (z)^{1-\alpha}/1-\alpha, & \alpha > 0 \text{ and } \alpha \neq 1 \\ \log(z), & \alpha = 1 \end{cases}$$

which is called α-fairness utility. α is a positive constant and represents the level of fairness. For example, maximising total utility corresponds to the proportional fairness when $\alpha = 1$ and the max–min fairness as $\alpha \to \infty$ [26].

An ideal objective is to achieve simultaneously both maximisation of each utility $U_m(S_m)$ and minimisation of the power consumption of each MS. The problem in this case can be formulated as a class of vector (multicriterion) optimisation problem as follows [19]:

$$\text{Maximise } \left[\begin{array}{l} U_1(S_1), \ldots, U_\mathrm{m}(S_m), \ldots, U_\mathrm{M}(S_M), \\ (-\sum_{t=1}^{2}\sum_{n=1}^{N} p_1^{t,n}), \ldots, (-\sum_{t=1}^{2}\sum_{n=1}^{N} p_m^{t,n}), \ldots, \\ (-\sum_{t=1}^{2}\sum_{n=1}^{N} p_M^{t,n}) \end{array} \right]$$

$$\text{Subject to Equations (1)–(5) and } S \geq 0$$

where S is the vector of service rate S_m , $\forall m$.

We cannot obtain the 'optimal solution' for this optimisation problem because all the objectives are not possible to satisfy at the same time. However, we can find a trade-off between the service rate of elastic flows and the power consumption of MSs by using the 'scalarisation technique' in [19] and introducing $2M$ parameters to get a linear combination of those objectives. The optimisation problem in this case becomes

$$\text{Maximise} \sum_{m=1}^{M} \left[\alpha_m U_\mathrm{m}(S_m) - \beta_m \sum_{t=1}^{2}\sum_{n=1}^{N} p_m^{t,n} \right] \quad \text{(P2)}$$

$$\text{Subject to Equations (1)–(5) and } S \geq 0$$

where α_m and β_m are the weighting parameters associated with MSm to determine the trade-off between the service rate and the power assumption. α_m can be viewed as the reward earned by the utility $U_\mathrm{m}(S_m)$, whereas β_m can be viewed as the price paid to the power consumed by MSm. Different trade-offs can be obtained by varying those parameters. Different from P1, S_m , $\forall m$ in P2 are decision variables that need to be optimised for fixed trade-off parameters α_m and β_m, $\forall m$.

4.2. Solution via dual problem

Following the same argument in Section 3 and the assumption of concavity of utility function, the dual gap of problem P2 is also zero. Therefore, strong duality holds, and we can solve it by dual decomposition and subgradient methods. The corresponding partial Lagrangian of P2 can be written as follows:

$$\begin{aligned} & L_2(\mathbf{S}, \mathbf{P}_{\mathrm{MS}}, \mathbf{P}_{\mathrm{RS}}, \mathbf{d}; \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\epsilon}) \\ &= \sum_{m=1}^{M} \left[\alpha_m U_\mathrm{m}(S_m) - \beta_m \sum_{t=1}^{2}\sum_{n=1}^{N} p_m^{t,n} \right] \\ &\quad - \sum_{m=1}^{M} \lambda_m \left(S_m - T_{m,\mathrm{BS}}^1 - \sum_{k=1}^{K} T_{m,k}^1 - T_{m,\mathrm{BS}}^2 \right) \\ &\quad - \sum_{k=1}^{K} \mu_k \left(\sum_{m=1}^{M} T_{m,k}^1 - T_{k,\mathrm{BS}}^2 \right) \\ &\quad - \sum_{k=1}^{K} \varepsilon_k \left(\sum_{n=1}^{N} p_{k,\mathrm{BS}}^n - p_k^{\max} \right) \\ &= \left\{ \sum_{m=1}^{M} \alpha_m U_\mathrm{m}(S_m) - \lambda_m S_m \right\} \\ &\quad + \sum_{n=1}^{N} \left\{ \sum_{m=1}^{M}\sum_{k=0}^{K} d_{m,k}^{1,n} \left[(\lambda_m - \mu_k) R_{m,k}^{1,n} - \beta_m p_m^{1,n} \right] \right\} \\ &\quad + \sum_{n=1}^{N} \left\{ \sum_{m=1}^{M}\sum_{k=0}^{K} d_{\mathrm{MS}m}^{2,n} \left[\lambda_m R_{m,\mathrm{BS}}^{2,n} - \beta_m p_m^{2,n} \right] \right. \\ &\quad \left. + d_{\mathrm{RS}K}^{2,n} \left[\mu_k R_{k,\mathrm{BS}}^{1,n} - \varepsilon_k p_{k,\mathrm{BS}}^{1,n} \right] \right\} \end{aligned} \tag{19}$$

The dual objective function is given by

$$D_t(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\epsilon}) = \begin{cases} \max\limits_{\mathbf{S}, \mathbf{P}_{\mathrm{MS}}, \mathbf{P}_{\mathrm{RS}}, \mathbf{d};} L_2(\mathbf{S}, \mathbf{P}_{\mathrm{MS}}, \mathbf{P}_{\mathrm{RS}}, \mathbf{d}; \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\epsilon}) \\ s.t. \quad (1)\ (2) \end{cases} \tag{20}$$

and the corresponding dual problem is

$$\begin{array}{l} \text{Minimise } D_t(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\epsilon}) \\ \text{Subject to } s.t.\ \boldsymbol{\lambda} \geq 0,\ \boldsymbol{\mu} \geq 0,\ \boldsymbol{\epsilon} \geq 0 \end{array} \tag{D2}$$

With Equation (19), the dual objective function can be decomposed into $M + 2N$ subproblems as follows:

(1) Service rate control

$$D_3^m(\lambda_m) = \max_{S_m} \alpha_m \mathrm{U_m}(S_m) - \lambda_m S_m, \quad \forall m = 1, \ldots, M \tag{21}$$

Eur. Trans. Telecomms. **22**:296–314 (2011) © 2011 John Wiley & Sons, Ltd.
DOI: 10.1002/ett

(2) Subcarrier assignment and power allocation in the first uplink subframe

$$D_1^n(\lambda_m, \mu_k) = \max_{P_m^{1,n},\ \sum_{k=0}^{K}\sum_{m=1}^{M} d_{m,k}^{1,n}=1} \sum_{m=1}^{M}\sum_{k=0}^{K} \times d_{m,k}^{1,n}\left[(\lambda_m - \mu_k)R_{m,k}^{1,n} - \beta_m p_m^{1,n}\right] \quad \forall n = 1,\dots N \tag{22}$$

(3) Subcarrier assignment and power allocation in the second uplink subframe

$$D_2^n(\lambda_m, \varepsilon_k, \mu_k) = \max_{\substack{P_m^{2,n}, P_{k,\mathrm{BS}}^{2,n},\\ \sum_{k=1}^{K} d_{\mathrm{RS}K}^{2,n} + \sum_{m=1}^{M} d_{\mathrm{MS}m}^{2,n}=1}} \sum_{m=1}^{M}\sum_{k=0}^{K} \times d_{\mathrm{MS}m}^{2,n}\left[\lambda_m R_{m,\mathrm{BS}}^{2,n} - \beta_m p_m^{2,n}\right] + d_{\mathrm{RS}k}^{2,n}\left[\mu_k R_{k,\mathrm{BS}}^{1,n} - \varepsilon_k P_{k,\mathrm{BS}}^{1,n}\right] \quad \forall n = 1,\dots N \tag{23}$$

Given the dual variable λ_m, we can get the optimal service rate for each MS in Equation (21) as follows:

$$S_m = U'^{-1}_m(\lambda_m/\beta_m), \ \forall m = 1,\dots,M \tag{24}$$

where $U'^{-1}_m(\cdot)$ is the inverse function of the derivative of utility function $U(\cdot)$.

The solutions to Equations (22) and (23) are similar to the solutions to Equations (10) and (14) (except weighting parameters α_m and β_m).

Similarly, a subgradient method can be used to solve dual problem D2, which has a convex objective function (20). For the subgradient at given dual variables, we have the following Lemma.

Lemma 2. *Considering the convex optimisation problem D2 and assuming that S_m^*, $T_{m,\mathrm{BS}}^{1\ *}$, $T_{m,k}^{1\ *}$, $T_{m,\mathrm{BS}}^{2\ *}$, $T_{k,\mathrm{BS}}^{2\ *}$ and $p_{k,\mathrm{BS}}^{n\ *}$ are the optimal solution of maximisation in the dual objective function (20), which can be obtained by solving Equations (21), (22) and (23), then*

$$g_{t,m}(\lambda_m) = T_{m,\mathrm{BS}}^{1\ *} + \sum_{k=1}^{K} T_{m,k}^{1\ *} + T_{m,\mathrm{BS}}^{2\ *} - S_m^*,$$

$$h_{t,k}(\mu_k) = T_{k,\mathrm{BS}}^{2\ *} - \sum_{k=1}^{M} T_{m,k}^{1\ *} \quad \text{and}$$

$$f_{t,k}(\varepsilon_k) = P_k^{\max} - \sum_{n=1}^{N} p_{k,\mathrm{BS}}^{n\ *}$$

are the subgradients of $D_t(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\epsilon})$ at λ_m, μ_k and ε_k, respectively.

The proof of Lemma 2 is similar to that of Lemma 1, which is presented in the Appendix.

Given the optimal solution $S_m^*, T_{m,\mathrm{BS}}^{1\ *}, T_{m,k}^{1\ *}, T_{m,\mathrm{BS}}^{2\ *}$, $T_{k,\mathrm{BS}}^{2\ *}$ and $p_{k,\mathrm{BS}}^{n\ *}$ in the current iteration i, the dual variables for D2 are updated in the following fashion:

$$\begin{cases} \lambda_m(i+1) = \max(0,\ \lambda_m(i) - \sigma(i) g_{t,m}(\lambda_m(i))\,), \forall m \\ \mu_k(i+1) = \max(0,\ \mu_k(i) - \phi(i) h_{t,k}(\mu_k(i))\,)\,, \forall k \\ \varepsilon_k(i+1) = \max(0,\ \varepsilon_k(i) - \theta(i) f_{t,k}(\varepsilon_k(i))\,)\,, \forall k \end{cases} \tag{25}$$

where $\sigma(i)$, $\phi(i)$ and $\theta(i)$ are the step sizes for λ_m, μ_k and ε_k, respectively, in iteration i.

The dual variables converge to the optimum if step sizes $\sigma(i)$, $\phi(i)$ and $\theta(i)$ are designed appropriately according to Theorem 1. Because strong duality holds, the corresponding primal variables $(\mathbf{S}^*, \mathbf{P}_{\mathrm{MS}}^*, \mathbf{P}_{\mathrm{RS}}^*, \mathbf{d}^*)$ are globally optimal variables of primal problem P2 for optimal dual variables $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*, \boldsymbol{\epsilon}^*)$.

4.3. Summary of the algorithm

We summarise the complete procedure of the whole algorithm in Algorithm 2 as follows.

Algorithm 2: Resource optimisation for the cross-layer trade-off between service rate of elastic flows and power efficiency of mobile stations

(1) The BS collects the trade-off parameters α_m and β_m from all MSs and channel state information through uplink control channels at the beginning of the frame. Then, the BS initialises the dual variables with $\lambda(0)$, $\mu(0)$ and $\varepsilon(0)$.
(2) Given $\lambda(t)$ in iteration t, the BS solves M service rate control subproblems in Equation (21) using Equation (24) to find $S_m^*, \forall m = 1,\dots M$.
(3) Given $\lambda(t)$, $\mu(t)$ and $\varepsilon(t)$ in iteration t, the BS solves 2N per-subcarrier subproblems in Equations (22) and (23) to obtain the optimal subcarrier assignment policies and the power allocation in the first and second uplink subframes, respectively.
(4) The BS calculates the $T_{m,BS}^{1\ *}$, $T_{m,k}^{1\ *}$, $T_{m,BS}^{2\ *}$ and $T_{k,BS}^{2\ *}$ with the solution obtained in steps 2 and 3, calculates subgradients according to Lemma 2 and then updates the dual variables using Equation (25).
(5) Return to step 2 until the algorithm converges.
(6) The BS broadcasts the resulting service rates S_m^*, subcarrier assignment policies d^* and power allocation $\mathrm{P}_{\mathrm{MS}}^*$ and $\mathrm{P}_{\mathrm{RS}}^*$ to all the MSs and RSs through downlink control channels.

4.4. Complexity of the algorithm

In the preceding subsections, we have solved optimisation problem P2 for the trade-off between the service rate of elastic flows and the power efficiency of MSs using subgradient method in its dual domain. The problem is decomposed into $M + 2N$ subproblems, that is, M service rate control subproblems in (21), N subcarrier assignment and power allocation subproblems in the first uplink subframe in Equation (22) and N subcarrier assignment and power allocation subproblems in the second uplink subframe in Equation (23). The complexity of solving per-subcarrier subproblems in the first and second uplink subframes is $O((K+1)M)$ and $O(K+M)$, respectively. The complexity of solving M service rate control subproblems is $O(M)$. Thus, the complexity of each iteration is $O((KM+K+M)N+M)$, which is linear in K, M and N, respectively. The complexity of subgradient method is polynomial in the number of dual variables (which is $2K + M$ for the D1′). So the computational complexity of the whole algorithm is linear in the number of the subcarriers N, which is significantly lower than employing the exhaustive search solution to the master primal problem P2 because the number of subcarrier assignment policies (d) increases exponentially with N.

The similar results of complexity to Algorithm 1 in Section 3.4 can be obtained. The only difference is that the subproblems (21) are not present in Algorithm 1. Thus, the complexity of computation for each iteration is $O((KM + K + M)N)$.

5. SIMULATION AND RESULTS

To show the performance of an OFDMA cellular network with fixed relays in the cross-layer optimisation framework and algorithms proposed in Sections 3 and 4, we conducted and presented a few simulations here. In the simulations, we consider a wireless OFDMA cellular network with a coverage of a 2-km radius. The MSs are assumed to have low mobility. The distance between the BS and each RS is about three-fifths of the cell radius. The locations of MSs are randomly generated and evenly distributed over the cell. However, from our simulation experience, too large channel gain would significantly reduce the rate of convergence. Thus, we impose some additional limits on the locations of MSs as follows:

- The distance between any MS and any RS is not less than 300 m;
- The distance between any MS and the BS is not less than 500 m.

Such restriction is reasonable because we care more about the power consumption of MSs far away from any fixed node (RSs or BS).

We model instantaneous channel gain of each subcarrier in a frame as the multiplication of a deterministic path loss with path loss exponent of 4 and a random Rayleigh fading component. We assume that the Rayleigh fading component is independent and identically distributed among all OFDMA subcarriers. The other parameters and their values used in the simulation are shown in Table I. The simulations are made using MATLAB. The results in the following simulations are obtained from the average values of 1000 simulation trials.

Table I. Simulation parameters.

Simulation parameter	Value
Total bandwidth	10 MHz
Noise power spectral density	−174 dBm/Hz
Number of subcarriers	64
Bandwidth of subcarrier	156.25 kHz
Number of MSs	24
Number of RSs	3, 6
P_k^{max}	36 dBm
Cell radius	2 km
Γ	1
C	1
α	4

A comparison with existing works is a good way to see the benefits of our method and how well our algorithm performs compared with the state-of-the-art algorithms. However, as it is pointed out in the Introduction, the scenario and objectives in optimisation that we consider in this paper are all different from those in the existing works. This makes it difficult to find a reasonable and fair way for comparison so that in this paper we do not make comparison. Nevertheless, our algorithm is developed based on dual decomposition and subgradient so that it is optimal in the resource allocation for the scenario and objectives in our optimisation model. The optimality of our algorithm is shown in Theorem 1.

5.1. Inelastic flow with fixed service rate

In the first simulation, we assume that there are only inelastic flows with fixed service rate requirement, which is the same for all MSs. The average transmission power (mW) consumed per MSs versus the required service rate (Mbps) of the inelastic flow in each MS is shown in Figure 6 for the scenarios where the numbers of RSs are three and six, respectively. Minimal total power assumption is obtained using Algorithm 1. The figure shows that the required average power per MS increases with the required service rate. The power consumed by the MS for the six-RS case is much less than for the three-RS case. This performance gain is about 32–58% when the number of RSs increases from three to six. This demonstrates that significant reduction of power consumption of MSs in OFMDA cellular network because of the deployment of more RSs can be fully obtained through our cross-layer optimisation Algorithm 1.

Figure 7 shows the total throughput via RSs versus the total required service rate of inelastic flows from all the MSs (the required service rate for inelastic flows in all MSs is the same). The figure reveals that with the increase of total required service rate, the throughput via RSs increases first for the lower total required service rate until it reaches the maximum and then decreases for the

Eur. Trans. Telecomms. **22**:296–314 (2011) © 2011 John Wiley & Sons, Ltd.
DOI: 10.1002/ett

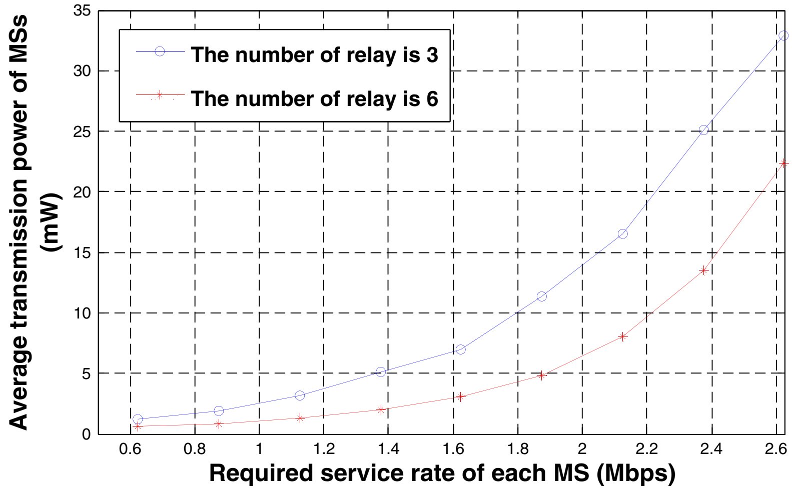

Figure 6. Required service rate versus average power consumption per mobile station.

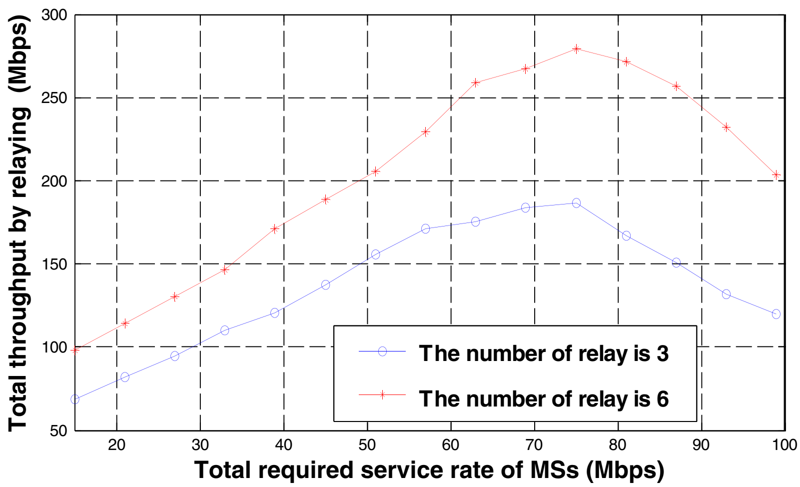

Figure 7. Total required service rate versus total throughput via RSs.

larger total required service rate. The reasons for this are the following:

- When the required service rate lies in the lower value range, it is efficient for the MSs that are near to RSs to transmit data by *relaying*. When the required service rate continues to increase, the RSs need to increase their capacity for relaying by utilising their maximum power and acquiring more subcarriers in the second subframe.
- When the required service rate lies in the higher value range, it is more efficient for some MSs to transmit data to the BS in both subframes (as discussed in the example in Section 3.1). Consequently, as the required service rate further increases, more subcarriers in the second subframe are allocated to MSs, which in turn reduces the capacity of relays.

5.2. Trade-off between service rate of elastic flows and power efficiency of mobile stations

In the second simulation, we show the resource allocation optimisation for the trade-off between the service rate of flows and the power efficiency of MSs in the case of only elastic flows present. We choose to use the logarithmic utility function $U_{\mathrm{m}}(S_m) = \log(S_m)$, $\forall m$, which provides proportional fairness for all elastic flows [26]. To easily illustrate the results, we further assume the weighting parameters $\alpha_m = \alpha$ and $\beta_m = \beta$, $\forall m$ in P2. Thus, the ratio β/α can be used to determine the trade-off between the service rate of flows and the power consumption of MSs. By varying β/α, we examine the change of the average service rate, total utility and average power spent per MS in the cell for two cases where the numbers of RSs are three and six, respectively. The results are obtained based on Algorithm 2 and shown in Figs. 8–10.

From those figures, we can see that the average service rate, the total utility and the average power consumption decrease with the increase of the value of β/α. As mentioned in Section 4, α_m can be viewed as the reward earned by the utility $U_{\mathrm{m}}(S_m)$, whereas β_m can be viewed as the price paid to the power expense of MSm. Thus, increasing β/α implies that the price of power of MSs becomes higher, and this will require the MSs to reduce power expenditure. This indicates that, for elastic flows, we can reduce transmission power of MSs by increasing the ratio of the weighting parameter β/α (Figure 10), which, however, will result in lower average service rate of elastic flows (Figure 8). But the benefit is that the decrease of the energy required to send one bit data for MSs is decreased (Figure 11). This means that an MS, which is an energy-limited device, can send more data for elastic flows for given fixed total energy in its battery (but with low rate). For each given (β/α), the trade-off is optimal because of the cross-layer optimisation Algorithm 2 in Section 3.

Another observation from the results shown in Figs. 8–10 is that more relays deployed can achieve better trade-off. In other words, the average service rate and total utility in the case of six RSs are larger than in the case of three RSs, whereas the average power consumed by MSs is less. But the performance gain in power consumption, which is about 22–25% (Figure 10), is smaller than the case of inelastic flows (Figure 6). This is because we have

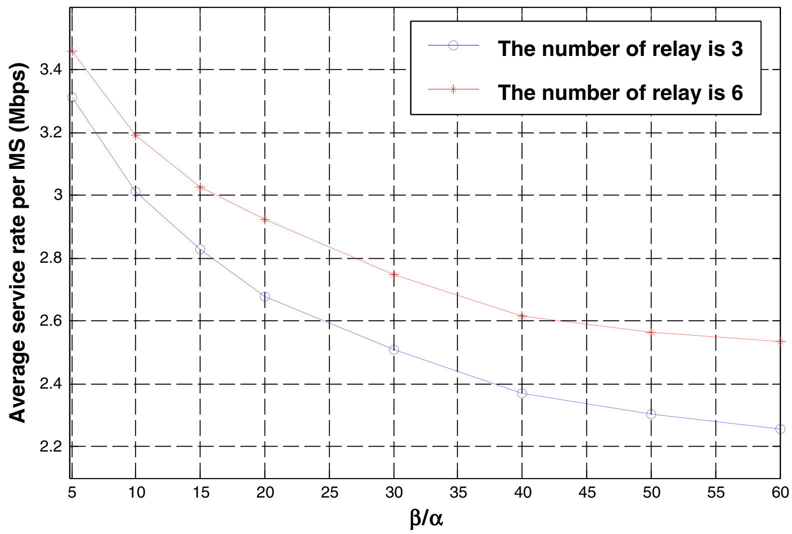

Figure 8. Average service rate versus parameter ratio (β/α).

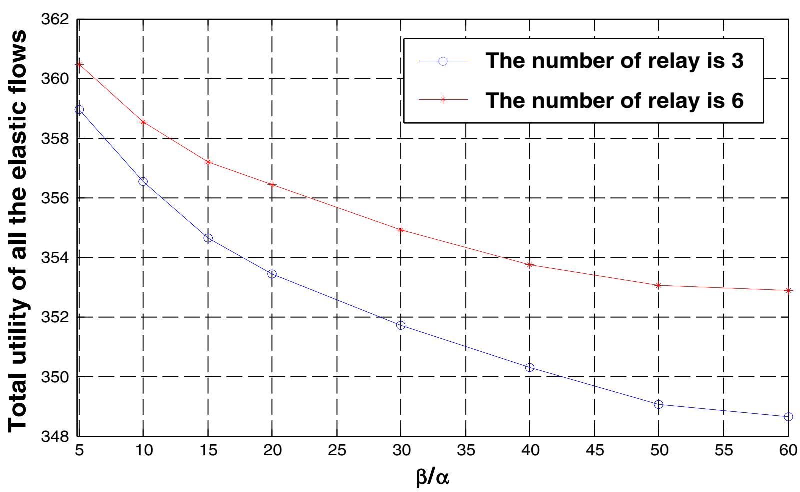

Figure 9. Total utility versus parameter ratio (β/α).

Eur. Trans. Telecomms. **22**:296–314 (2011) © 2011 John Wiley & Sons, Ltd.
DOI: 10.1002/ett

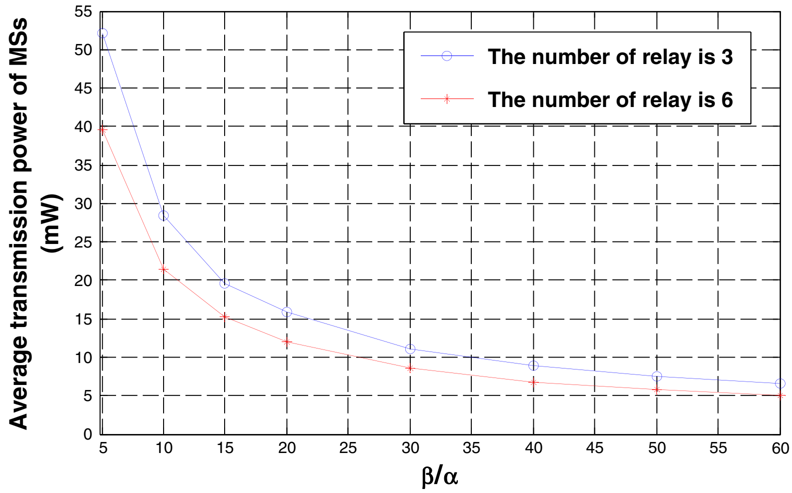

Figure 10. Average power consumption per MS versus parameter ratio (β/α).

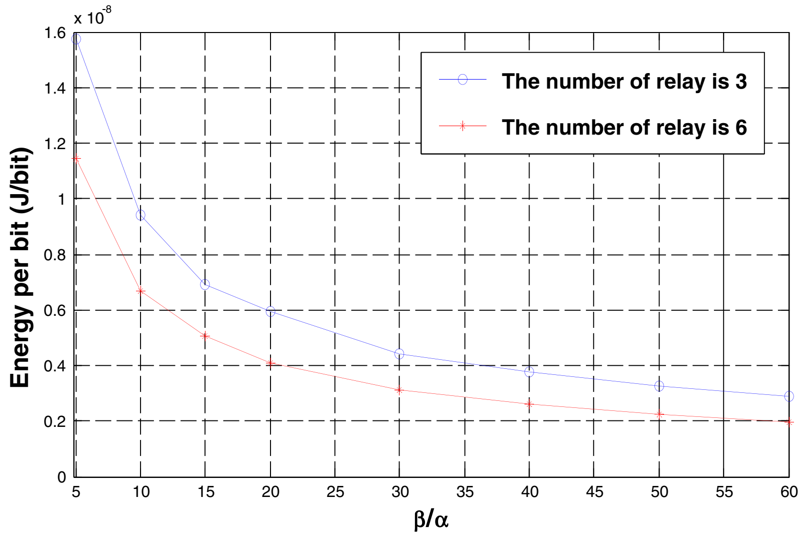

Figure 11. Energy needed for the transmission of one bit versus parameter ratio (β/α).

obtained higher gain in average service rate (Figure 8) for the six-RS case in the trade-off optimisation.

Comparing the results in Figures 6, 8 and 10, we can notice that it takes much lower average power in the trade-off optimisation of elastic flows to achieve the same average rate as in the optimisation of inelastic flow. This is because we impose the strict fairness (i.e. identical service rate) for all the inelastic flows. The MSs being located far away from any RS or BS have to spend quite much power to satisfy this rate requirement, which results in high average power consumption, whereas, in the trade-off optimisation of its elastic flow, such MSs may lower their service rate to reduce its transmission power, and thus the average power consumption is reduced.

6. CONCLUSION

In this paper, we have developed the cross-layer resource optimisation framework for both inelastic and elastic flows in the uplink transmission of an OFDMA cellular network with fixed RSs. The RSs are assumed to be dedicated devices to relay the data from MSs to a BS and have no energy limitation whereas the MSs are power-limited and energy-limited devices. We have formulated the cross-layer optimisation problem to minimise the sum power consumption of MSs in the case of inelastic flow and presented the cross-layer trade-off between the service rate and power consumption of MSs in the case of elastic flows. Dual decomposition and subgradient update methods were employed to obtain the optimal solution with reduced computational complexity, and simulations were conducted using MATLAB. Simulation results have shown that through the proposed cross-layer resource optimisation framework and algorithms, the benefit of deployment of multiple RSs in the uplink transmission of OFDMA cellular network in significantly reducing power consumption in the inelastic flow case can be fully obtained, and trade-off between benefits in increasing service rate and saving energy can also be achieved.

APPENDIX A: PROOF OF LEMMA 1

Proof: Definition of subgradient [22, 23] given a convex function $f : R^n \to R$, a vector $h \in R^n$ is a subgradient of f at the point $v \in R^n$ if $f(u) \geq f(v) + (u - v)^T h$, $\forall u \in R^n$.

Consider objective function $-D(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\epsilon})$ in D1$'$ at two different points (λ, μ, ϵ) and$(\lambda', \mu, \epsilon)$, where $\lambda = (\lambda_1, \ldots, \lambda_m, \ldots, \lambda_M)$ and $\lambda' = (\lambda_1, \ldots, \lambda'_m, \ldots, \lambda_M)$. We have

$$-D(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\epsilon}) = \begin{cases} -\min\limits_{\mathbf{P}_{\mathrm{MS}}, \mathbf{P}_{\mathrm{RS}}, \mathbf{d}} L_1(\mathbf{P}_{\mathrm{MS}}, \mathbf{P}_{\mathrm{RS}}, \mathbf{d}; \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\epsilon}) \\ s.t. \quad (1)\ (2) \end{cases} \tag{26}$$

$$-D(\boldsymbol{\lambda}', \boldsymbol{\mu}, \boldsymbol{\epsilon}) = \begin{cases} -\min\limits_{\mathbf{P}_{\mathrm{MS}}, \mathbf{P}_{\mathrm{RS}}, \mathbf{d}} L_1(\mathbf{P}_{\mathrm{MS}}, \mathbf{P}_{\mathrm{RS}}, \mathbf{d}; \boldsymbol{\lambda}', \boldsymbol{\mu}, \boldsymbol{\epsilon}) \\ s.t. \quad (1)\ (2) \end{cases} \tag{27}$$

Letting the optimal values of $\mathbf{P}_{\mathrm{MS}}, \mathbf{P}_{\mathrm{RS}}$ and $\mathbf{d}$ in Equations (26) and (27) be $\mathbf{P}^*_{\mathrm{MS}}, \mathbf{P}^*_{\mathrm{RS}}, \mathbf{d}^*$ and $\mathbf{P}'^*_{\mathrm{MS}}, \mathbf{P}'^*_{\mathrm{RS}}, \mathbf{d}'^*$, respectively, we can find the subgradient of $-D(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\epsilon})$ at λ_m in the following manner:

$$\begin{aligned}
&\left[-D(\boldsymbol{\lambda}', \boldsymbol{\mu}, \boldsymbol{\epsilon})\right] - \left[-D(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\epsilon})\right] \\
&= -L_1(\mathbf{P}'^*_{\mathrm{MS}}, \mathbf{P}'^*_{\mathrm{RS}}, \mathbf{d}'^*; \boldsymbol{\lambda}', \boldsymbol{\mu}, \boldsymbol{\epsilon}) \\
&\quad + L_1(\mathbf{P}^*_{\mathrm{MS}}, \mathbf{P}^*_{\mathrm{RS}}, \mathbf{d}^*; \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\epsilon}) \\
&\geq -L_1(\mathbf{P}^*_{\mathrm{MS}}, \mathbf{P}^*_{\mathrm{RS}}, \mathbf{d}^*; \boldsymbol{\lambda}', \boldsymbol{\mu}, \boldsymbol{\epsilon}) \\
&\quad + L_1(\mathbf{P}^*_{\mathrm{MS}}, \mathbf{P}^*_{\mathrm{RS}}, \mathbf{d}^*; \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\epsilon}) \\
&= -\lambda'_m \left(S_m - T^{1\ *}_{m,\mathrm{BS}} - \sum_{k=1}^{K} T^{1\ *}_{m,k} - T^{2\ *}_{m,\mathrm{BS}} \right) \\
&\quad + \lambda_m \left(S_m - T^{1\ *}_{m,\mathrm{BS}} - \sum_{k=1}^{K} T^{1\ *}_{m,k} - T^{2\ *}_{m,\mathrm{BS}} \right) \\
&= (\lambda'_m - \lambda_m) \left(T^{1\ *}_{m,\mathrm{BS}} + \sum_{k=1}^{K} T^{1\ *}_{m,k} + T^{2\ *}_{m,\mathrm{BS}} - S_m \right)
\end{aligned} \tag{28}$$

$T^{1\ *}_{m,\mathrm{BS}}, T^{1\ *}_{m,k}$ and $T^{2\ *}_{m,\mathrm{BS}}$ are the link layer rate of links MSm–BS and MSm–RSk in the first uplink subframe and MSm–BS in the second uplink subframe, respectively, when the dual variables are $(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\epsilon})$. The inequality in Equation (28) holds because of the definition of dual function in Equation (7). The second equality holds because of the definition of Lagrange in Equation (6). Thus, we get

$$\begin{aligned}
-D(\boldsymbol{\lambda}', \boldsymbol{\mu}, \boldsymbol{\epsilon}) \geq -D(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\epsilon}) + (\lambda'_m - \lambda_m)(T^{1\ *}_{m,\mathrm{BS}} \\
+ \sum_{k=1}^{K} T^{1\ *}_{m,k} + T^{2\ *}_{m,\mathrm{BS}} - S_m)
\end{aligned} \tag{29}$$

By the definition of subgradient, the subgradient of $-D(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\epsilon})$ at the point λ_m is

$$g_m(\lambda_m) = T^{1\ *}_{m,\mathrm{BS}} + \sum_{k=1}^{K} T^{1\ *}_{m,k} + T^{2\ *}_{m,\mathrm{BS}} - S_m$$

Similarly, we can also get the subgradients of $-D(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\epsilon})$ at μ_k and ε_k, which are , respectively, of the following form

$$h_k(\mu_k) = T^{2\ *}_{k,\mathrm{BS}} - \sum_{k=1}^{M} T^{1\ *}_{m,k} \ ,$$

$$f_k(\varepsilon_k) = P_k^{\max} - \sum_{n=1}^{N} p^{n\ *}_{k,\mathrm{BS}} \ .$$

ACKNOWLEDGEMENTS

This work is supported in part by VINNOVA Sweden through grant 2008-00954, the Chinese Scholarship Council (CSC) and the Ministry of Science and Technology China through the National International Science and Technology Cooperation Project (no. 2008DFA12090).

REFERENCES

1. Pabst R, Walke HB, Schultz CD, *et al*. Relay-based deployment concepts for wireless and mobile broadband radio. *IEEE Communication Magazine* 2004; **42**(9): 80–89.
2. IEEE 802.16’s Relay Task Group. http://ieee802.org/16/relay/index.htm.
3. Rhee W, Cioffi JM. Increasing in capacity of multiuser OFDM system using dynamic subchannel allocation. *Proceedings of Vehicular Technology Conference*, VTC2000-Spring, vol. 2, 15–18 May, 2000; pp. 1085–1089.
4. Jang J, Lee KB. Transmit power adaptation for multiuser OFDM systems. *IEEE Journal on Selected Areas in Communications* 2003; **21**:171–178.
5. Cheong YW, Cheng RS, Lataief KB, Murch RD. Multiuser OFDM with adaptive subcarrier, bit, and power allocation. *IEEE Journal on Selected Areas in Communications* 1999; **17**(10):1747–1758.

6. Seong K, Mohseni M, Cioffi JM. Optimal resource allocation for OFDMA downlink systems. *Proceedings of IEEE International Symposium on Information Theory 2006*, 9 –14 July 2006; pp. 1394 –1398.
7. Huang J, Subramanian V, Berry R, Agrawal R. Joint Scheduling and Resource Allocation in Uplink OFDM Systems. *Proceedings of Asilomar Conference*, Pacific Grove, CA, Nov. 2007.
8. Huang L, Rong M, Wang L, *et al*. Resource allocation for OFDMA based relay enhanced cellular networks. *Proceedings of Vehicular Technology Conference*, VTC 2007, April 2007; pp. 3160–3164.
9. Li G, Liu H. Resource allocation for OFDMA relay networks with fairness constraints. *IEEE Journal on Selected Areas in Communications* 2006; **24**(11):2061–2069.
10. Cendrillon R, Yu W, Moonen M, Verlinden J, Bostoen T. Optimal multiuser spectrum balancing for digital subscriber lines. *IEEE Transactions on Communications* May 2006; **54**(5):922–933.
11. Yu W, Lui R. Dual methods for nonconvex spectrum optimization of multicarrier systems. *IEEE Transactions on Communications* 2006; **54**(7):1310 –1322.
12. Ng T, Yu W. Joint optimization of relay strategies and resource allocations in cooperative cellular networks. *IEEE Journal on Selected Areas in Communications* 2007; **25**(2):328–339.
13. Nam W, Chang W, Chung SY, Lee YH. Transmit optimization for relay-based cellular OFDMA systems. *Proceedings of IEEE ICC'07*, June 2007; pp. 5714–5719.
14. Kelly FP, Maulloo A, Tan D. Rate control for communication networks: shadow prices, proportional fairness and stability. *Journal of the Operations Research Society* 1998; **49**(3):237–252.
15. Chiang M, Low SH, Calderbank AR, Doyle JC. Layering as optimization decomposition: a mathematical theory of network architectures. *Proceedings of the IEEE* 2007; **95**(1):255–312.
16. Lee K-D, Leung VCM. Evaluations of achievable rate and power consumption in cooperative cellular networks with two classes of nodes. *IEEE Transactions on Vehicular Technology* 2008; **57**(2):1166 –1175.
17. Hoo LMC, Halder B, Tellado J, Cioffi JM. Multiuser transmit optimization for multicarrier broadcast channels: asymptotic FDMA capacity region and algorithms. *IEEE Transactions on Communications* 2004; **52**(6): 922–930.
18. Palomar DP, Chiang M. A tutorial on decomposition methods for network utility maximization. *IEEE Journal on Selected Areas in Communications* 2006; **24**(8): 1439 –1451.
19. Boyd S, Vandenberghe L. *Convex Optimization*. Cambridge University Press: London, 2004.
20. Karakayali K, Kang JH, Kodialam M, *et al*. Cross-layer optimization for OFDMA-based wireless mesh backhaul networks. *Proceedings of IEEE Wireless Communications and Networking Conference 2007*, 11–15 March 2007; pp. 276–281.
21. Soldati P, Johansson M. Network-wide resource optimization of wireless OFDMA mesh networks with multiple radios. *Proceedings of IEEE International Conference on Communications, 2007 (ICC '07)*, 24–28 June 2007; pp. 4979–4984.
22. Bertsekas DP. *Nonlinear Programming,* 2nd edn. Athena Scientific: Belmont, 1999.
23. Boyd S, Xiao L, Mutapcic A. Subgradient methods. Lecture notes of EE392o, 2003–2004. Stanford University, Autumn Quarter.
24. Kim K, Han Y, Kim S-L. Joint subcarrier and power allocation in uplink OFDMA systems. *IEEE Communications Letters* 2005; **9**(6):526–528.
25. Chiang M. Balancing transport and physical layers in wireless multihop networks: Jointly optimal congestion control and power control. *IEEE Journal on Selected Areas in Communications* 2005; **23**(1):104–116.
26. Mo J, Walrand J. Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on Networking* 2000; **8**(5):556–567.

AUTHORS' BIOGRAPHIES

Lei You received the Ph.D degree from Beijing University of Posts and Telecommunications, Beijing, China, in 2010. During 2008.9-2009.8, He was a visiting PhD student with Uppsala University, Uppsala, Sweden. He is currently a lecture of the school of Electronic Information Engineering at Tianjin University. His research interests include wireless multi-hop networks and cross-layer optimization.

Ping Wu received BS degree in electrical and electronic engineering in Nanjing Aeronautical & Astronomical University, China, MS and Ph D degrees in electrical and electronic engineering in Xi'an Jiaotong University, China, in 1982, 1987 and 1991, respectively. After receiving PhD degree, he worked as a postdoctor in Fraunhofer Institute for Biomedical Engineering, St. Ingbert, Germany for two years (1992-1994). Since Oct. 1994 he has been with the Division of Signals and Systems in Department of Engineering Sciences, Uppsala University, Sweden, and is an associate professor. He has published more than 50 journal conference papers. His current research interest includes wireless mesh and sensor networks, OFDM and MIMO OFDM, cross-layer design and optimization, network coding, and signal processing with applications in telecommunications.

Mei Song received the B.S. and M.S. degrees in electrical engineering from Tianjin University, in 1983 and 1986, respectively, and received Ph.D. degree in electrical engineering from Beijing University of Posts and Telecommunications (BUPT). She is currently Professor in BUPT, and also the Vice Dean of Electronic Engineering Department. Prof. Song has participated and completed dozens of research and development projects, including those of National Ministry of Education, National Nature Science Fund council, National 863 Program, Ministry of Information Industry Minister Fund Projects, Ministry of Information Industry, China Mobile, China Telecom, US Intel Corporation and US Ericsson Corporation. Until now, Prof. Song has already published five translations and monographs, more than one hundred papers independently or cooperatively, and applied 9 patents. Her research interests are integrated design technology, VLSI&CAD system, mobile communication system, computer networks, and related technological specifications.

Junde Song graduated from the Radio Engineering Department, Beijing University of Posts and Telecommunications (BUPT), and awarded the honorary Doctor of Moscow Electronic Engineering Academy in 1991. He is currently Professor in BUPT, and also the Chairman of Academic Degrees Committee of BUPT. He is the chairman of CMIS (China Mobile Internet Application Specification), and also the member of Science and Technology Committee of MII (Ministry of Information Industry). Prof. Song has visited 32 foreign countries including Sweden, USA, Japan, Germany, England, Holland, Swiss, France, India, Canada and Thailand, etc. mainly for teaching and co-researching purpose. He has been awarded many prizes from Nation, Chinese Education Ministry, Electronic Ministry, Posts and Telecommunications Ministry (now Ministry of Information Industry) etc. He has published 14 books and more than two hundred papers. Now his research interests are mobile communication network management, network optimization and frequency planning, the 3rd generation mobile communication, software radio and coverage of wireless/mobile communication system.

Yong Zhang is an associate professor in the School of Electronic Engineering at the Beijing University of Posts and Telecommunications, Beijing, China. He holds a Ph.D degree in Beijing University of Posts and Telecommunications. Until now, He has already published three translations and monographs. He has published more than 30 journal conference papers. His research interests include cognitive radio, cognitive network, and self-organizing network.

Eur. Trans. Telecomms. **22**:296–314 (2011) © 2011 John Wiley & Sons, Ltd.
DOI: 10.1002/ett