Self-Organized Resource Allocation in LTE Systems with Weighted Proportional Fairness

I-Hong Hou, Chung Shue Chen

Self-Organized Resource Allocation in LTE Systems with Weighted Proportional Fairness

I-Hong Hou
Computer Engineering and System Group & Department of ECE
Texas A&M University
College Station, TX 77840, USA
Email: ihou@tamu.edu

Chung Shue Chen
Networks & Networking Domain
Alcatel-Lucent Bell Labs
Centre de Villarceaux, 91620 Nozay, France
Email: cs.chen@alcatel-lucent.com

***Abstract*—We consider the problem of LTE network self organization and optimization of resource allocation. One particular challenge for LTE systems is that, by applying OFDMA, a transmission may use multiple resource blocks scheduled over the frequency and time. There are three key components involved in the resource allocation and network optimization: resource block scheduling, power control, and client association. We propose a distributed protocol that aims to achieve weighted proportional fairness (WPF) among clients by jointly consider them. The cross-layer design includes: (i) an optimal online policy for resource block scheduling, (ii) a heuristic for transmit power control, and (iii) a selfish strategy for client association. The proposed scheme only requires limited local information exchange and thus can be easily implemented for large networks. Simulation results have shown its effectiveness in both the system throughput and user fairness.**

I. Introduction

With the foreseen exponentially increasing number of users and traffic in LTE/LTE-A and future 4G systems [1], existing deployment and practice of cellular radio networks that strongly rely on highly hierarchical architectures with centralized control and resource management becomes economically unsustainable [2]. Network self organization and optimization are among the key targets of future cellular networks so as to relax the heavy demand of human efforts in the network planning and optimization tasks and to reduce the system's capital and operational expenditure (CAPEX/OPEX) [3]. The next-generation mobile networks is expected to provide a full coverage of broadband wireless service and support fair and efficient radio resource utilization with a high degree of operation autonomy and system intelligence.

In this paper, we aim to design distributed algorithms of radio resource allocation and network optimization for large LTE networks. It is known that system-wide radio resource optimization is usually very challenging [4]. A joint optimization of user association, channel selection and power control is in general non-convex and hard to solve even if centralized algorithm is allowed [5]. Notice that in LTE, the channel selection involves multi-carrier assignment and scheduling, since a transmission can use multiple resource blocks in the frequency and time space. This yields extra complexity and difficulty. We assume that there is a lack of centralized control entity for global coordination. It is interesting to see how individual nodes can operate autonomously and support inter-cell interference coordination in a self-organized manner.

Here, we consider the problem of achieving weighted proportional fairness for LTE systems. We formulate this problem as an optimization problem and identify that there are three important components involved: (i) resource block scheduling, (ii) transmit power control, and (iii) client association. While this optimization problem is non-convex, making it difficult to obtain the optimum solution, we propose a distributed protocol that jointly considers the three components. This protocol applies an online algorithm that achieves the optimum solution for the resource block scheduling, given solutions to the other two. It then deploys a heuristic for the component of transmit power control. Finally, this protocol considers the selfish behavior of a client when it chooses the base station or access point to be associated with, while taking into account its influence to the resource block scheduling. The protocol only requires base stations and clients to exchange information locally and can achieve network weighted proportional fair resource allocation optimization in a self-organized manner. Simulation results have shown that the proposed protocol substantially outperforms today's default methods in both user throughput and bandwidth sharing fairness.

There has been some work on self-organized wireless systems. For example, in [6], a distributed algorithm for the self optimization of radio resources built on Gibbs's sampler was developed. However, it is focused on the network throughput optimization with potential delay fairness. A similar approach was proposed in [7] which aims to maximize the sum of arbitrary user throughput utility functions. The result is however based on the assumption that the system and channel gains are static and all fixed. A comprehensive literature review

This material is based upon work partially supported by NSF under Contracts CNS-1035378, CCF-0939370, CNS-1035340, and CNS-0905397, USARO under Contract Nos. W911NF-08-1-0238 and W-911-NF-0710287, and AFOSR under Contract FA9550-09-0121.

The work presented in this paper has been carried out at LINCS (www.lincs.fr). We would like to thank François Baccelli and Laurent Thomas for their valuable discussion and support.

of distributed power allocation, load balancing and network utility maximization can be found in [7]. From a methodological standpoint, the work which is most related to the proposed scheme here is that of [8], which provided a distributed solution of resource allocation for proportional fairness in the context of multi-band wireless systems. The work however only addresses time-division multiplexing and the interference model is restricted to collision-based random access channels. In contrast, the results established here are developed in the context of SINR-metric model with respect to LTE systems and have very different setup.

II. System Model and Problem Setup

Consider a reuse-1 radio system with several base stations and clients that operate and use LTE orthogonal frequency division multiple access (OFDMA) [9]. LTE divides frequency bandwidth into subcarriers, and time into frames, which are further divided into 20 time slots. The bandwidth of a subcarrier is 15 kHz while the duration of a time frame is 10 ms. In this paper, we consider LTE frequency division duplex (FDD) and the downlink transmission and resource scheduling. In LTE, a resource block consists of 12 consecutive subcarriers and one time slot of duration 0.5 ms. Under the OFDMA, each user can be allocated any number of resource blocks. However, within a cell a resource block cannot be allocated to more than one user. LTE can thus achieve both time-division multiplexing and frequency-division multiplexing.

Note that in each resource block, one can consider that the channel is usually flat over the subcarriers given that the channel coherence bandwidth is greater than 180 kHz [2, Ch.12]. It is time invariant in each time slot given that the channel coherence time is greater than 0.5 ms [2, Ch.23]. However, the channel or link gain of a user may change from one resource block to another in the frequency and time domain. We hereby define $\mathbb{M}$ to be the set of base stations, $\mathbb{I}$ to be the set of clients, and $\mathbb{Z} = \mathbb{F} \times \mathbb{Q}$ to be set of resource blocks, where each $f \in \mathbb{F}$ represents a collection of 12 consecutive subcarriers and each $q \in \mathbb{Q}$ represents a time slot. In the sequel, we use both $z \in \mathbb{Z}$ and (f, q) to denote a resource block for the notational convenience. Note that here we consider reuse-1 systems. However, the result could be extended to other systems.

We denote by $P_{m,z}$ the amount of transmission power that a base station $m \in \mathbb{M}$ assigns on resource block z. If base station m does not operate in resource block z, we have $P_{m,z} = 0$. Further, we assume that each base station has a fixed power budget W, and it is required that $\sum_f P_{m,(f,q)} \leq W$, which is also known as the per base station transmit power constraint. Propagation loss and path condition are captured by the channel or link gain. Let $G_{i,m,z}$ be the channel gain between base station m and client i on resource block z. To be more specific, when the base station m transmits with power $P_{m,z}$, the received power at client i on resource block z is $G_{i,m,z}P_{m,z}$. The received power, $G_{i,m,z}P_{m,z}$, of client i is considered to be its received signal strength if base station m is transmitting data to client i, and is considered to be interference, otherwise. Therefore, when base station m is transmitting data to client i on resource block z, the signal-to-interference-plus-noise ratio (SINR) of client i on z is expressible as:

$$\mathrm{SINR}_{i,z} = \frac{G_{i,m,z}P_{m,z}}{N_{i,z} + \sum_{l \neq m} G_{i,l,z}P_{l,z}},$$

where $N_{i,z}$ is the thermal noise experienced by client i on resource block z. The throughput of this transmission can then be described by the Shannon capacity as $B \log(1 + \mathrm{SINR}_{i,z})$, where B is the bandwidth of a resource block.

Each client i is associated with one base station $m(i) \in \mathbb{M}$. In each frame, base station m schedules one client that is associated with m in each of the resource blocks in the frame. Let $\phi_{i,m,z}$ be the proportion of frames that client i is scheduled in resource block z by base station m. $G_{i,m,z}$ does not vary over the time. We will then discuss in the following sections how to take channel time variation into account. The influence of channel fading is also demonstrated by simulations in Section VI.

Consider that $G_{i,m,z}$ does not vary over the time, the overall throughput of client i, which is the sum of its throughput over all the resource blocks, can hence be written as:

$$r_i := \sum_{z \in \mathbb{Z}} \phi_{i,m(i),z} B \log\Big(1 + \frac{G_{i,m,z}P_{m,z}}{N_{i,z} + \sum_{l \neq m} G_{i,l,z}P_{l,z}}\Big). \quad (1)$$

We aim to achieve weighted proportional fairness among all the clients. Let w_i be the priority weight of client i or user-dependent priority indicator. The weighted proportional fairness can be achieved by maximizing $\sum_{i \in \mathbb{I}} w_i \log r_i$. Therefore, we can formulate our problem as:

$$\begin{aligned}
\text{Max} \quad & \sum_{i \in \mathbb{I}} w_i \log r_i && (2)\\
\text{s.t.} \quad & \sum_{i: m(i) = m} \phi_{i,m(i),z} = 1, \quad \forall m \in \mathbb{M}, z \in \mathbb{Z}, && (3)\\
& \sum_{f \in \mathbb{F}} P_{m,(f,q)} \leq W, \quad \forall m \in \mathbb{M}, q \in \mathbb{Q}, && (4)\\
& m(i) \in \mathbb{M}, \quad \forall i \in \mathbb{I}, && (5)\\
& P_{m,z} \geq 0, \phi_{i,m(i),z} \geq 0, \quad \forall i, m, z. && (6)
\end{aligned}$$

This formulation shows that there are three important components involved. In each frame, a base station needs to decide which client should be scheduled in each resource block. This essentially determines the values of $\phi_{i,m(i),z}$. We call this component the *Scheduling Problem*. In each frame, a base station also needs to decide how much power it should allocate in each resource block, subject to the constraint on its power budget. This component is referred as the *Power Control Problem*. Besides, every client needs to choose a base station to be associated with. We denote this component as the *Client Association Problem*.

III. Online Algorithm for the Scheduling Problem

In this section, we study the Scheduling Problem, given solutions to the Power Control Problem and the Client Association Problem, i.e., values of $P_{m,z}$ and $m(i)$. We thus define

$$H_{i,m(i),z} := B \log(1 + \frac{G_{i,m(i),z} P_{m(i),z}}{N_{i,z} + \sum_{l \neq m(i)} G_{i,l,z} P_{l,z}}),$$

which is the throughput of client i on resource block z when it is scheduled by base station $m(i)$, to simplify the notations. The Scheduling Problem can then be written as:

$$\text{Max} \quad \sum_i w_i \log r_i = \sum_i w_i \log(\sum_z \phi_{i,m(i),z} H_{i,m(i),z}) \quad (7)$$

$$\text{s.t.} \quad \sum_{i:m(i)=m} \phi_{i,m(i),z} = 1, \forall m \in \mathbb{M}, z \in \mathbb{Z}, \quad (8)$$

$$\phi_{i,m(i),z} \geq 0, \forall i, z. \quad (9)$$

One can see that the above optimization problem is in fact convex and hence can be solved by standard techniques of convex optimization. To further simplify the computation overhead, we propose an online scheduling policy for the Scheduling Problem. Let $\phi_{i,m(i),z}[k]$ be the proportion of frames that base station $m(i)$ has scheduled client i for resource block z in the first $k-1$ frames. Similarly, let $r_i[k] = \sum_{z \in \mathbb{Z}} \phi_{i,m(i),z}[k] H_{i,m(i),z}$ be the average throughput of client i in the first k frames. We then have:

$$\begin{cases} \phi_{i,m(i),z}[k+1] = \frac{k-1}{k} \phi_{i,m(i),z}[k] + \frac{1}{k} \\ \text{ if client } i \text{ is scheduled for } z \text{ in } (k+1)\text{-th frame,} \\ \phi_{i,m(i),z}[k+1] = \frac{k-1}{k} \phi_{i,m(i),z}[k], \quad \text{otherwise.} \end{cases} \quad (10)$$

In our online scheduling policy, the base station schedules the client i that maximizes $w_i H_{i,m(i),z} / r_i[k]$. It solves the Scheduling Problem. The result is summarized below.

Theorem 1: Using the above scheduling policy, the value of $\liminf_{k \to \infty} \sum_{i \in \mathbb{I}} w_i \log r_i[k]$ achieves the maximum of the optimization problem (7)–(9).

Proof: We denote by ϕ the vector consisting $\{\phi_{i,m(i),z}\}$ and by $\phi[k]$ the vector consisting $\{\phi_{i,m(i),z}[k]\}$. Let

$$L(\phi) = \sum_{i \in \mathbb{I}} w_i \log(\sum_{z \in \mathbb{Z}} \phi_{i,m(i),z} H_{i,m(i),z}).$$

We first show that $\lim_{k \to \infty} L(\phi[k])$ exists, and then show that $\lim_{k \to \infty} L(\phi[k])$ achieves the maximum of the optimization problem (7)–(9).

Since $L(\phi)$ is infinitely differentiable, by Taylor's theorem, we have for any ϕ and $\Delta\phi$,

$$L(\phi + \Delta\phi) = L(\phi) + \nabla L(\phi) \Delta\phi + \pi(\phi, \Delta\phi),$$

where $|\pi(\phi, \Delta\phi)| < a |\Delta\phi|^2$, for some constant a. Let $\Delta\phi[k]$ be the vector consisting of $\{\phi_{i,m(i),z}[k+1] - \phi_{i,m(i),z}[k]\}$. By the design of our online scheduling policy, $\Delta\phi_{i,m(i),z}[k] := \phi_{i,m(i),z}[k+1] - \phi_{i,m(i),z}[k] = \frac{1}{k} - \frac{\phi_{i,m(i),z}[k]}{k}$ if client i has the largest $\frac{w_i H_{i,m(i),z}}{r_i[k]}$ among all clients j with $m(j) = m(i)$; and $\Delta\phi_{i,m(i),z}[k] = -\frac{\phi_{i,m(i),z}[k]}{k}$, otherwise. Since $|\Delta\phi_{i,m(i),z}[k]| < \frac{1}{k}$, for all i and z, we have:

$$\begin{aligned} & L(\phi[k+1]) = L(\phi[k] + \Delta\phi[k]) \\ \geq & L(\phi[k]) + \nabla L(\phi[k]) \Delta\phi[k] - \tfrac{a}{k^2} \\ = & L(\phi[k]) + \textstyle\sum_{m \in \mathbb{M}, z \in \mathbb{Z}} \sum_{i:m(i)=m} \frac{w_i H_{i,m,z}}{r_i[k]} \Delta\phi_{i,m,z}[k] - \frac{a}{k^2} \\ = & L(\phi[k]) + \tfrac{1}{k} \textstyle\sum_{m \in \mathbb{M}, z \in \mathbb{Z}} (\max_{i:m(i)=m} \frac{w_i H_{i,m,z}}{r_i[k]} \\ & - \textstyle\sum_{i:m(i)=m} \frac{w_i H_{i,m,z}}{r_i[k]} \phi_{i,m,z}[k]) - \frac{a}{k^2} \\ \geq & L(\phi[k]) - \frac{a}{k^2}, \end{aligned}$$

where the last inequality follows since $\sum_{i:m(i)=m} \phi_{i,m,z}[k] = 1$, for all m and z.

Let $\beta := \limsup_{k \to \infty} L(\phi[k])$. For any $\epsilon > 0$, there exists large enough K so that $L(\phi[K]) > \beta - \frac{\epsilon}{2}$ and $\sum_{k=K}^{\infty} \frac{a}{k^2} < \frac{\epsilon}{2}$. We now have that, for any $\hat{k} > K$, $L(\phi[\hat{k}]) \geq L(\phi[K]) - \sum_{k=K}^{\hat{k}} \frac{a}{k^2} > \beta - \epsilon$. Therefore, $L(\phi[k])$ converges to β, as $k \to \infty$.

Next, we show that $\lim_{k \to \infty} L(\phi[k])$ achieves the maximum of the optimization problem (7)–(9). Due to the constraint (8), ϕ is a solution to the optimization problem if and only if $\frac{\partial L}{\partial \phi_{i,m(i),z}} = \frac{w_i H_{i,m(i),z}}{r_i} = \max_{j:m(j)=m(i)} \frac{w_j H_{j,m(j),z}}{r_j}$, for all i, z such that $\phi_{i,m(i),z} > 0$, and ϕ satisfies constraints (8)–(9). Suppose $\lim_{k \to \infty} L(\phi[k])$ does not achieve the maximum of the optimization problem. Since $\phi[k]$ satisfies constraints (8)–(9), there exists $\delta > 0$, $\lambda > 0$, and positive integer K such that for all $k > K$, there exists some $i_k \in \mathbb{I}$ and $z_k \in \mathbb{Z}$ so that $\phi_{i_k,m(i_k),z_k}[k] > \delta$ and $\frac{w_{i_k} H_{i_k,m(i_k),z_k}}{r_{i_k}} < \max_{j:m(j)=m(i_k)} \frac{w_j H_{j,m(j),z_k}}{r_j} - \lambda$. Now we have:

$$\begin{aligned} & L(\phi[k+1]) - L(\phi[k]) \geq \nabla L(\phi[k]) \Delta\phi[k] - \tfrac{a}{k^2} \\ = & \textstyle\sum_{m \in \mathbb{M}, z \in \mathbb{Z}} \sum_{i:m(i)=m} \frac{w_i H_{i,m,z}}{r_i[k]} \Delta\phi_{i,m,z}[k] - \frac{a}{k^2} \\ = & \tfrac{\delta\lambda}{k} - \tfrac{a}{k^2} \geq \tfrac{\delta\lambda}{2k}, \end{aligned}$$

for large enough k. Since $\sum_{k=1}^{\infty} \frac{1}{k} = \infty$, $\lim_{k \to \infty} L(\phi[k]) = \infty$, which is a contradiction. Hence, $\lim_{k \to \infty} L(\phi[k])$ achieves the maximum of the optimization problem. □

Note that in the previous discussions, we have assumed that the channel gain, $G_{i,m,z}$, does not vary over time. In practice, however, channel gains fluctuate due to fading. To take fading into account, we let $\hat{G}_{i,m,z}$ be the instantaneous time-varying channel gain, and $\hat{H}_{i,m(i),z}$ be the instantaneous throughput that client i can get from resource block z if it is scheduled by AP $m(i)$. Our scheduling policy can then be easily modified such that the base station schedules the client with the largest $w_i \hat{H}_{i,m(i),z} / r_i[k]$ on resource block z. In Section VI, we show that this modification can further improve performance.

IV. A Heuristic for the Power Control Problem

In this section, we discuss the Power Control Problem, i.e., how the APs choose $P_{m,z}$ so as to solve (2)–(6).

Obviously, APs need to know the solution of the Scheduling Problem $\{\phi_{i,m,z}\}$ and the values of channel gains $\{G_{i,m,z}\}$, in order to choose suitable $P_{m,z}$. To reduce computation and communication overhead and maintain operational simplicity, the APs assume that for all clients i associated with AP m, the perceived thermal noises are all $N_{m,z}$, the channel gains between them and AP m are all $G_{m,m,z}$, and the channel gains between them and AP l ($\neq m$) are all $G_{m,l,z}$ on resource block z. In practice, $N_{m,z}$ and $G_{m,m,z}$ can be set to be the average value of the noise powers and the average value of the channel gains between m and its clients, respectively, and $G_{m,l,z}$ can be set to be the channel gain between AP m and AP l on resource block z. Under this assumption, one can see that the optimal solution to the Scheduling Problem (7)–(9) is given by:

$$\phi_{i,m(i),z} = \frac{w_i}{\sum_{j:m(j)=m(i)} w_j}.$$

Let $w^m := \sum_{i:m(i)=m} w_i$ be the sum weight of the clients that are associated with AP m. Under a fixed solution to the Client Association Problem, the Power Control Problem can be rewritten as:

$$\text{Max } \sum_m w^m \log \sum_z B \log\left(1 + \frac{G_{m,m,z}P_{m,z}}{N_{m,z}+\sum_{l\neq m} G_{m,l,z}P_{l,z}}\right) \quad (11)$$

$$\text{s.t. } \sum_{f\in\mathbb{F}} P_{m,(f,q)} \leq W, \forall m \in \mathbb{M}, q \in \mathbb{Q}, \quad (12)$$

$$P_{m,z} \geq 0, \forall i, m, z. \quad (13)$$

This problem is non-convex and it may be infeasible to find the global optimal solution. Instead, we propose a distributed heuristic that converges to a local optimum. We apply a gradient method for the Power Control Problem as follows. Let P be the vector consisting of $\{P_{m,z}\}$,

$$\text{SINR}_{m,z}(P) := \frac{G_{m,m,z}P_{m,z}}{N_{m,z} + \sum_{l\neq m} G_{m,l,z}P_{l,z}},$$

and

$$U(P) := \sum_{m\in\mathbb{M}} w^m \log[\sum_{z\in\mathbb{Z}} B\log(1 + \text{SINR}_{m,z}(P))].$$

We have:

$$\begin{aligned}\frac{\partial U(P)}{\partial P_{m,z}} =& \frac{w^m}{\sum_{y\in\mathbb{Z}}\log(1+\text{SINR}_{m,y}(P))} \times \frac{G_{m,m,z}}{N_{m,z}+\sum_l G_{m,l,z}P_{l,z}}\\ &+\sum_{o\neq m} \frac{w^o}{\sum_{y\in\mathbb{Z}}\log(1+\text{SINR}_{o,y}(P))} \times [\frac{G_{o,m,z}}{N_{o,z}+\sum_l G_{o,l,z}P_{l,z}}\\ &-\frac{G_{o,m,z}}{N_{o,z}+\sum_{l\neq o} G_{o,l,z}P_{l,z}}].\end{aligned}$$

Each AP updates its power periodically. When AP m updates its power, it sets its power on resource block (f,q) to be:

$$\begin{cases} [P_{m,(f,q)} + \alpha\frac{\partial U(P)}{\partial P_{m,(f,q)}}]^+ \\ \qquad \text{if } \sum_e [P_{m,(e,q)} + \alpha\frac{\partial U(P)}{\partial P_{m,(e,q)}}]^+ \leq W, \\ W\frac{[P_{m,(f,q)}+\alpha\frac{\partial U(P)}{\partial P_{m,(f,q)}}]^+}{\sum_e [P_{m,(e,q)}+\alpha\frac{\partial U(P)}{\partial P_{m,(e,q)}}]^+}, \quad \text{otherwise}, \end{cases} \quad (14)$$

where $x^+ := \max\{x, 0\}$ and α is a small constant. AP m needs to compute $\frac{\partial U(P)}{\partial P_{m,z}}$ to update its power on each resource block. The computation of $\frac{\partial U(P)}{\partial P_{m,z}}$ can be further simplified by setting $G_{o,m,z} = 0$ for all o such that $G_{o,m,z}$ is small and has little influence on the value of $\frac{\partial U(P)}{\partial P_{m,z}}$. Thus, to compute $\frac{\partial U(P)}{\partial P_{m,z}}$, AP m exchanges information with AP o that is physically close to it so as to know:

- the sum weight of clients associated with o, i.e., w^o,
- the channel gain $G_{o,m,z}$ from m to o,
- the sum of interference and noise $N_{o,z} + \sum_{l\neq o} G_{o,l,z}P_{l,z}$ at o,
- the received signal strength $G_{o,o,z}P_{o,z}$ at o, and
- the average total throughput $\sum_{y\in\mathbb{Z}} \log(1 + \text{SINR}_{o,y}(P))$ in the downlink of AP o,

for all o that $G_{o,m,z}$ is large. In LTE, the above information can be obtained through periodic channel quality indicator and reference signal reports.

This method is easy to implement and only requires limited information exchange between neighbor cells. We assume that the neighbor cell communication takes place between base stations and is supported by the wired backhaul network.

V. A Selfish Strategy for the Client Association Problem

In the following, we discuss how to solve the Client Association Problem, i.e., how each client i should choose a base station $m(i)$. Instead of aiming to achieve network bandwidth sharing fairness, we assume that each client i is selfish and would like to choose $m(i)$ that maximizes its own throughput, which is expressible as:

$$r_i = \sum_{z\in\mathbb{Z}} \phi_{i,m(i),z} H_{i,m(i),z}.$$

We make this assumption under three main reasons. First, this conforms to the selfish behaviors of clients. Secondly, in a dense network, the decision of $m(i)$ by one client i only has a limited and indirect impact on the overall fairness. Resource allocation and scheduling with inter-cell interference coordination will finally play a key role on this. Thirdly, heavily sacrificing the goal of maximizing throughput would highly penalize the network capacity.

To choose $m(i)$ that maximizes r_i, client i needs to know the values of $H_{i,m,z}$ and $\phi_{i,m,z}$, if it is associated with AP m. Client i assumes that the powers used by APs are not influenced much by its choice, which is true in a dense network. Thus, client i only needs to know its perceived SINR with each AP on each resource block to compute $H_{i,m,z}$. It remains for the client i to compute the value of $\phi_{i,m,z}$ if it is associated with AP m. We propose two different approaches to compute this value. In the first approach, which we call the *Exact Simulator* (ES), client i first obtains the values of w_j and $H_{j,m,z}$ for all clients j that are associated with m. Client i can then simulates the scheduling decisions of AP m by running

Algorithm 1 Approximate Estimator

1: Sort all resource blocks such that $\frac{\overline{H}_{m,1}}{H_{i,m,1}} \le \frac{\overline{H}_{m,2}}{H_{i,m,2}} \le \frac{\overline{H}_{m,3}}{H_{i,m,3}} \le \cdots$
2: $\phi_{i,m,z} \leftarrow 0, \forall z; r_i \leftarrow 0, \forall i; \overline{r}_m \leftarrow \sum_z \overline{H}_{m,z}, \forall m$
3: **for** $z = 1 \to |\mathbb{Z}|$ **do**
4:     **if** $\frac{w_i H_{i,m,z}}{r_i + H_{i,m,z}} > \frac{w^m_{-i}\overline{H}_{m,z}}{\overline{r}_m - \overline{H}_{m,z}}$ **then**
5:         $\phi_{i,m,z} \leftarrow 1$
6:         $r_i \leftarrow r_i + H_{i,m,z}$
7:         $\overline{r}_m \leftarrow \overline{r}_m - \overline{H}_{m,z}$
8:     **else if** $\frac{w_i H_{i,m,z}}{r_i} < \frac{w^m_{-i}\overline{H}_{m,z}}{\overline{r}_m}$ **then**
9:         Break
10:     **else**
11:         $\phi_{i,m,z} \leftarrow \frac{\overline{r}_m w_i H_{i,m,z} - r_i w^m_{-i}\overline{H}_{m,z}}{(w^m_{-i}+w_i)H_{i,m,z}\overline{H}_{m,z}}$
12:         $r_i \leftarrow r_i + \phi_{i,m,z}H_{i,m,z}$
13:         $\overline{r}_m \leftarrow \overline{r}_m - \phi_{i,m,z}\overline{H}_{m,z}$
14:         Break
15: **return** r_i

the online scheduling policy introduced in Section III, and obtains the value of $\phi_{i,m,z}$ on each resource block z. While this approach offers an accurate estimation on $\phi_{i,m,z}$, it requires high computation and communication overhead.

In the second approach, which we call the *Approximate Estimator* (AE), client i only obtains the values of $w^m_{-i} := \sum_{j:j\neq i, m(j)=m} w_j$, and

$$\overline{H}_{m,z} := \sum_{j:m(j)=m} \phi_{j,m,z} H_{j,m,z},$$

which is the average throughput of AP m on resource block z. Client i assumes that, when another client j is scheduled by AP m on resource block z, its throughput on z equals the average throughput $\overline{H}_{m,z}$. Client i can then estimate r_i by *Algorithm 1*. The complexity of *Algorithm 1* is $O(|\mathbb{Z}|)$, and therefore this approach is much more efficient than the Exact Simulator. Moreover, the following theorem suggests that the Approximate Simulator provides reasonably good estimates on the throughput of client i if it is associated with AP m.

Theorem 2: If for each client j other than i that is associated m, $H_{j,m,z} = \overline{H}_{m,z}$, then, under the online scheduling policy introduced in Section III, the throughput of client i equals the value of r_i obtained by Algorithm 1 when it is also associated with m.

Proof: Theorem 1 has shown that the online scheduling policy in Section III achieves the optimum solution to the Scheduling Problem. We claim that, by setting $\phi_{i,m,z}$ as that derived in Algorithm 1 and $\phi_{j,m,z} = \frac{w_j}{\sum_{k\neq i, m(k)=m} w_k}(1-\phi_{i,m,z})$, for all $j \neq i$, $m(j) = m$, the resulting r_i and $\{r_j | j \neq i, m(j) = m\}$ also achieve the optimum solution to the Scheduling Problem.

In the proof of Theorem 1, it has been shown that $\{\phi_{j,m,z} | m(j) = m\}$ maximizes $\sum_{j:m(j)=m} w_j \log r_j$ if and only if $\phi_{j,m,z} \ge 0$, for all j, z, $\sum_j \phi_{j,m,z} = 1$, for all m, z, and $\frac{\partial L}{\partial \phi_{j,m,z}} = \frac{w_j H_{j,m,z}}{r_j} = \max_{k:m(k)=m} \frac{w_k H_{k,m,z}}{r_k}$, for all j, z such that $\phi_{j,m,z} > 0$. By our settings of $\phi_{j,m,z}$, the first two conditions hold, and we only need to verify the last condition.

Sort all PRBs such that $\frac{\overline{H}_{m,1}}{H_{i,m,1}} \le \frac{\overline{H}_{m,2}}{H_{i,m,2}} \le \frac{\overline{H}_{m,3}}{H_{i,m,3}} \le \dots$, we consider two possible cases: there exists some z_0 such that $0 < \phi_{i,m,z_0} < 1$, and such z_0 does not exist, i.e., $\phi_{i,m,z} \in \{0, 1\}$ for all z. In the first case, we have that $\phi_{i,m,z} = 1$, for all $z < z_0$, and $\phi_{i,m,z} = 0$, for all $z > z_0$. By setting $\phi_{j,m,z} = \frac{w_j}{\sum_{k\neq i, m(k)=m} w_k}(1-\phi_{i,m,z})$, for all $j \neq i$, $m(j) = m$, we have $r_j = \frac{w_j}{\sum_{k:m(k)=m, k\neq i} w_k}\overline{r}_m = \frac{w_j}{w^m_{-i}}\overline{r}_m$. Let r^*_i and $\overline{r}^*_m$ be the values of r_i and $\overline{r}_m$ in the z_0^{th} iteration of the for loop in Algorithm 1. As $0 < \phi_{i,m,z_0} < 1$, lines 13–17 are executed in this iteration, and we have $r_i = r^*_i + \phi_{i,m,z_0} H_{i,m,z_0}$ and $\overline{r}_m = \overline{r}^*_m - \phi_{i,m,z_0}\overline{H}_{m,z_0}$. Moreover, in line 13, the value of ϕ_{i,m,z_0} is chosen so that $\frac{w_i H_{i,m,z_0}}{r^*_i + \phi_{i,m,z_0} H_{i,m,z_0}} = \frac{w^m_{-i}\overline{H}_{m,z_0}}{\overline{r}^*_m - \phi_{i,m,z_0}\overline{H}_{m,z_0}}$. Therefore, $\frac{w_i H_{i,m,z_0}}{r_i} = \frac{w^m_{-i}\overline{H}_{m,z_0}}{\overline{r}_m} = \frac{w_j \overline{H}_{m,z_0}}{r_j}$, for all j such that $m(j) = m$ and $j \neq i$. Thus, the last condition holds for PRB z_0. For any PRB $z < z_0$, $\frac{\overline{H}_{m,z}}{H_{i,m,z}} \le \frac{\overline{H}_{m,z_0}}{H_{i,m,z_0}}$. We then have $(\frac{w_i H_{i,m,z}}{r_i})/(\frac{w_j \overline{H}_{m,z}}{r_j}) \ge (\frac{w_i H_{i,m,z_0}}{r_i})/(\frac{w_j \overline{H}_{m,z_0}}{r_j}) = 1$, and hence $\frac{w_i H_{i,m,z}}{r_i} \ge \frac{w_j \overline{H}_{m,z}}{r_j}$, for all j such that $m(j) = m$ and $j \neq i$. As we set $\phi_{j,m,z} = 0$ for all $j \neq i$, the last condition holds for all $z < z_0$. Similarly, for any PRB $z > z_0$, $\frac{w_i H_{i,m,z}}{r_i} \le \frac{w_j \overline{H}_{m,z}}{r_j}$, for all j such that $m(j) = m$ and $j \neq i$. As we set $\phi_{i,m,z} = 0$, the last condition also holds for all $z > z_0$. In sum, the last condition holds for the case that there exists some z_0 such that $0 < \phi_{i,m,z_0} < 1$.

Next consider the case that $\phi_{i,m,z} \in \{0, 1\}$ for all z. Let z_1 be the smallest integer so that $\phi_{i,m,z_1} = 0$. In the z_1^{th} iteration of the for loop in Algorithm 1, steps 10–11 are executed, and we have $\frac{w_i H_{i,m,z}}{r_i} < \frac{w^m_{-i}\overline{H}_{m,z}}{\overline{r}_m} = \frac{w_j \overline{H}_{m,z}}{r_j}$, for all j such that $m(j) = m$ and $j \neq i$. The last condition then holds for PRB z_1. A similar argument as in the previous paragraph shows that the last condition holds for all z. □

VI. Simulation Results

In the following, we present the simulation results. Consider a reuse-1 LTE-FDD system with 9 APs that share a bandwidth of 10 MHz, which can accommodate 600 subcarriers [10] and there are $\frac{600}{12} \times 20 = 1000$ resource blocks. We consider that the 9 APs are placed as a 3×3 grid in symmetry. The distance between two adjacent APs is 500 meters. The four sectors have various user densities: there are 25 clients uniformly distributed in the sector $[0, 500] \times [0, 500]$, 16 clients uniformly distributed in each of the sectors $[0, 500] \times [500, 1000]$ and $[500, 1000] \times [0, 500]$, and 9 clients uniformly distributed in the sector $[500, 1000] \times [500, 1000]$. Channel gains are derived from the following equation:

$$PL(d) = 128.1 + 37.6 \log_{10}(d) + X + Y, \qquad (15)$$

where $PL(d)$ is the channel gain in dB and d is distance in km. X and Y represent shadowing and fast fading, respectively. X is the log-normal shadowing with mean 0 and standard deviation 8 dB. Since X is a slow fading, we consider that it is time invariant. However, it will vary in frequency, in every 180 kHz. On the other hand, Y represents Rayleigh fast fading with a Doppler of 5 Hz. It also varies in frequency. We assume that the power budget of each AP is 1 W. The thermal noise is randomly generated in $[3.5, 4.5] \times 10^{-15}$ W. For simplicity, we set the priority weight $w_i = 1$, for all i.

We have implemented the proposed online scheduling policy, power control, and Approximate Estimator. We consider both the cases where APs use instant knowledge of channel gains for the Scheduling Problem and where APs only have knowledge of long-term average channel gains. These two cases are referred as *Fast Feedback* (FF) and *Slow Feedback* (SF), respectively. We compare our mechanisms against a *Default* mechanism which employs round-robin for the Scheduling Problem, uses the same amount of power on all resource blocks (i.e., no or very limited power control as in today's LTE downlink), and associates each client to the closest AP, respectively [2, Ch.12].

Fig. 1 and Fig. 2 show the simulation results, including both the total throughput and the cumulative distribution function (CDF) of throughput of each client, under the three considered mechanisms. Both FF and SF achieve much better total throughput than Default does. In addition, both FF and SF achieve better fairness. The proportions of clients that have throughputs at least 3 Mbps under FF and SF are 82% and 64%, respectively. On the other hand, by employing Default, only 15% of clients have throughputs at least 3 Mbps. It is observed that the performance of FF is better than SF, showing that the performance of the system can be further improved by incorporating instantaneous knowledge of channel gains.

VII. Conclusion

In this paper, we developed a cross-layer protocol for self-organizing LTE systems. This protocol aims to achieve weighted proportional fairness and network optimization by jointly addressing the resource block scheduling, power control, and client association. We proposed an optimal on-line resource block scheduling algorithm, a heuristic for power control, and a selfish strategy for client association, which take into account the different components and their interplay. The resulting protocol only requires local and limited information exchange. It can be easily implemented and applied to LTE/LTE-A. Simulation results have shown that the proposed protocol achieved much better performance than current mechanisms in both the system throughput and user fairness.

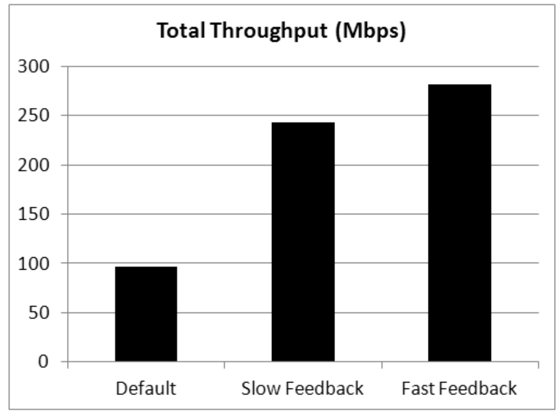

Fig. 1: System throughput under different mechanisms.

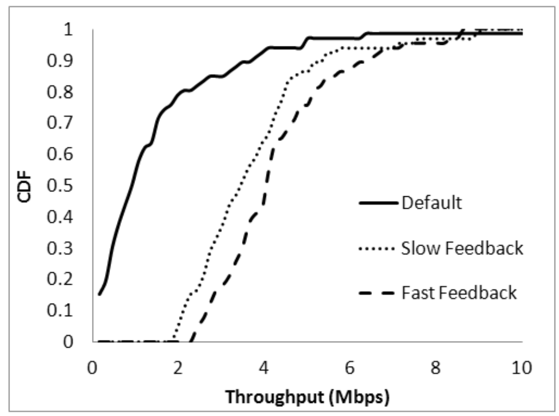

Fig. 2: The CDF of user throughput in the system.

References

[1] 3rd Generation Partnership Project. 3GPP LTE-Advanced. http://www.3gpp.org/lte-advanced.

[2] S. Sesia, I. Toufik, and M. Baker, *LTE - The UMTS Long Term Evolution: From Theory to Practice*, 2nd ed. John Wiley & Son, 2011.

[3] Next Generation Mobile Networks Group (NGMN). http://www.ngmn.org.

[4] C. S. Chen, K. W. Shum, and C. W. Sung, "Round-robin power control for the weighted sum rate maximisation of wireless networks over multiple interfering links," *European Trans. on Telecommunications*, pp. 1–13, 2011.

[5] Z.-Q. Luo and S. Zhang, "Dynamic spectrum management: complexity and duality," *IEEE J. Sel. Topics in Signal Processing*, vol. 2, no. 1, pp. 57–73, Feb. 2008.

[6] C. S. Chen and F. Baccelli, "Self-optimization in mobile cellular networks: Power control and user association," in *IEEE ICC*, May 2010.

[7] S. Borst, M. Markakis, and I. Saniee, "Distributed power allocation and user assignment in OFDMA cellular networks," in *Allerton Conference on Communication, Control, and Computing*, Sep. 2011.

[8] I.-H. Hou and P. Gupta, "Distributed resource allocation for proportional fairness in multi-band wireless systems," in *IEEE ISIT*, Jul. 2011.

[9] 3GPP TS 36.211, "Evolved universal terrestrial radio access: Physical channels and modulation," Tech. Spec. v10.2.0, Jun. 2011.

[10] 3GPP TS 36.101, "Evolved universal terrestrial radio access: User equipment radio transmission and reception," Tech. Spec. v10.3.0, Jun. 2011.