

Dynamic Resource Management for LTE-Based Hybrid Access Femtocell Systems

Ying Loong Lee, *Student Member, IEEE*, Jonathan Loo and Teong Chee Chuah

Abstract—Hybrid access femtocells for Long Term Evolution (LTE)-based cellular networks provide a tradeoff between closed and open access femtocells whereby all subscribers are granted access albeit with priority given to closed access subscribers. Due to the need to accommodate both closed and open access subscribers, quality of service (QoS) provisioning for LTE-based hybrid access femtocells has become more challenging. This paper addresses this issue and proposes a new dynamic resource management scheme for such hybrid architectures. In particular, the proposed scheme first classifies and performs lexicographic admission control on the incoming traffic data flows using an optimal greedy algorithm. A sub-optimal delay-bounded packet scheduling algorithm and a dual decomposition-based power allocation algorithm are developed to solve the non-convex maximization problem such that the weighted sum rate of each femtocell is maximized, subject to bounded packet delays and power constraints. Simulation results show that the proposed scheme can significantly outperform existing schemes in terms of QoS, throughput and fairness.

Index Terms—LTE, Hybrid Access Femtocell, Packet Scheduling, Admission Control, QoS, Power Allocation, Fairness.

I. INTRODUCTION

FEMTOCELLS have been regarded as a promising technology for Long Term Evolution (LTE)-based cellular networks. This technology mitigates indoor coverage holes and provides traffic offloading from evolved NodeBs (eNBs). Also, femtocells shorten the distance between user equipment (UE) and base stations, thus providing more reliable data transmission links and better quality of service (QoS) provisioning. In addition, femtocells are user-deployed and of low-cost. Intrigued by these features of femtocells, intensive research has now been conducted to investigate the application of femtocell technology into LTE systems under various paradigms such as coexistence between femtocell and wireless fidelity (WiFi) networks [2] and self-organizing disaster-resilient femtocell networks [3].

Femtocell base stations, also known as Home evolved NodeBs (HeNBs), can operate in closed and open modes [4], [5]. The closed access mode allows only the UE registered under the closed subscriber group (CSG) [6] to access the

HeNB whereas the open access mode allows all UEs to access the HeNB. However, the closed access mode may result in inferior system-wide performance [7], [8] whereas the open access mode may lead to inferior performance for the HeNB owners [9]. Thus, the so-called hybrid access mode [4], [5] has been introduced as a tradeoff between the two modes whereby all UEs can still access the HeNB albeit with priority given to the CSG subscribers. Since the hybrid access mode deals with both CSG and non-CSG subscribers, resource allocation and QoS provisioning has become more challenging.

Numerous resource allocation studies for femtocell networks have been conducted. A survey of resource allocation techniques for LTE-based femtocell networks in [10] showed that most of these techniques employed round robin and proportional fair scheduling methods to allocate resources among UEs (which are only suitable for closed or open access femtocells) while QoS provisioning was mostly neglected. On the other hand, several recent studies in [11]–[17] have further investigated into resource allocation and QoS provisioning for femtocell networks. In [11], a hybrid access femtocell resource management scheme was developed to only admit CSG and non-CSG subscribers based on their QoS requirements and to allocate resources to these subscribers based on opportunistic proportional fair scheduling. A resource allocation scheme was proposed in [12] for hybrid access femtocell networks which prioritizes CSG subscribers by assigning them larger weighting coefficients compared to non-CSG subscribers when maximizing the total weighted sum rate. In [13], a utility-based scheduling and admission control algorithm was proposed to prioritize CSG subscribers with real-time traffic in hybrid access femtocell networks. In [14], a resource allocation scheme was designed to maximize the number of admitted CSG subscribers is maximized while serving all non-CSG subscribers in the hybrid access femtocell network. QoS provisioning for co-channel femtocells with heterogeneous services was investigated in [15] in which the network throughput was maximized under delay-sensitive user and interference constraints. In [16], cooperative bargaining game theory was used to design a resource allocation scheme for cognitive femtocell networks that mitigates interference and guarantees fairness under imperfect channel estimation conditions, while accounting for minimum outage probability and rate constraints. Resource allocation for cognitive femtocell networks with imperfect spectrum sensing was investigated in [17] to maximize the network throughput under QoS, fairness and interference constraints.

The resource allocation schemes in [11]–[17] were mainly designed for femtocells in generic orthogonal frequency divi-

Manuscript received ...

A preliminary version of this work appeared in the proceedings of the IEEE 26th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Hong Kong, China, Aug.-Sep. 2015 [1].

Y. L. Lee and T. C. Chuah are with the Faculty of Engineering, Multimedia University, Cyberjaya, 63100 Selangor, Malaysia (e-mail: lee.ying.loong12@student.mmu.edu.my, tcchuah@mmu.edu.my).

J. Loo is with the School of Science and Technology, Middlesex University, London NW4 4BT, UK. (e-mail: J.Loo@mdx.ac.uk).

This work is supported in part by the Ministry of Higher Education Malaysia under the Fundamental Research Grant Scheme MMUE/140082.

sion multiple access (OFDMA) networks with little consideration of the LTE system architecture. In particular, they do not take into account the LTE protocols in their scheduling design in which admission control and resource allocation should be performed over radio bearers, i.e., data flows. Also, the studies in [11], [12], [16] and [17] did not differentiate real-time and non-real-time traffic classes and did not address packet loss and delay issues. Moreover, hybrid access femtocells were not studied in [15]–[17]. Therefore, to bridge the aforementioned gaps, we are motivated to investigate resource allocation and QoS provisioning for LTE-based hybrid access femtocell systems which account for various traffic classes.

In this paper, a new downlink resource management scheme which dynamically prioritizes CSG subscribers and guarantees QoS provisioning for LTE-based hybrid access femtocell systems is proposed. The proposed scheme first classifies and differentiates the incoming real-time and non-real-time flows of CSG and non-CSG subscribers. Then, admission control is performed over these data flows with priority given to data flows of CSG subscribers. Thereafter packet scheduling and power allocation are performed over the admitted data flows in such a way that bounded packet delays, high throughput and high degrees of fairness are guaranteed. The contributions of this paper are summarized as follows:

- 1) A dynamic resource management framework encompassing traffic classification, admission control, packet scheduling and power allocation is designed. In this framework, incoming data flows are classified into real-time and non-real-time flows as well as data flows of CSG and non-CSG subscribers. Based on this classification, the incoming data flows are admitted with priority given to real-time and non-real-time flows of CSG subscribers followed by those of non-CSG subscribers. Packet scheduling is then performed by allocating resources to the admitted data flows to ensure bounded packet delays and a high degree of fairness. Power allocation is further performed after packet scheduling by setting the transmission power level on each resource allocated to the data flows to enhance the achievable throughput and fairness.
- 2) The admission control problem is formulated as a lexicographic optimization problem whereby real-time and non-real-time flows of CSG subscribers are first prioritized followed by those of non-CSG subscribers under constrained resource availability. To solve this problem, an optimal greedy admission control algorithm is developed.
- 3) The packet scheduling and power allocation problem of each femtocell is formulated as a non-convex optimization problem that maximizes the weighted sum rate, subject to delay and power constraints. The problem is solved in two steps. In the first step, packet scheduling is performed assuming power allocation has been performed. A suboptimal delay-bounded packet scheduling algorithm which prioritizes real-time flows is developed. In the second step, power allocation is performed on the resources allocated to each data flow during the first

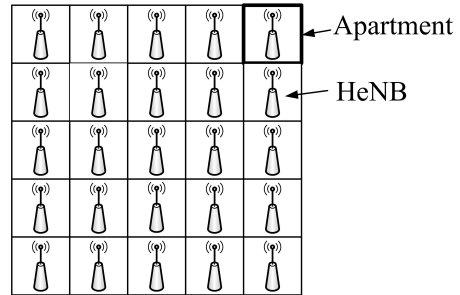


Fig. 1. Femtocell Network Model.

step by using a dual decomposition method for further enhancing throughput and fairness.

The remainder of this paper is organized as follows: Section II describes the system model. Section III presents the proposed resource management scheme. Performance results are presented and discussed in Section IV. Finally, Section V concludes the paper.

II. SYSTEM MODEL

We consider the LTE-based indoor femtocell network model shown in Fig. 1, which is established over an indoor building. The indoor building consists of a number of apartments each accommodating one femtocell. Let \mathcal{H} denote the set of HeNBs in the femtocell network. In LTE-based cellular systems, the channel bandwidth is divided into a number of physical resource blocks (PRBs) [18] each spanning a duration of 0.5 ms and utilizes a 180 kHz sub-channel bandwidth. As packet scheduling in LTE-based cellular systems is performed every transmission time interval (TTI) of 1 ms, the smallest resource unit that can be allocated is a PRB pair (PRBP) spanning a 1 ms duration. Let \mathcal{K}_h denote the set of PRBPs of HeNB $h \in \mathcal{H}$. Without loss of generality, we assume that each PRBP experiences slow and flat fading, and the femtocell network is perfectly synchronized. Furthermore, we assume that a set of PRBPs has been allocated to each femtocell and co-channel interference between femtocells has been mitigated. This assumption can be made valid by using the frequency reuse scheme in [19], which divides the system bandwidth into several frequency bands and allocates these bands to femtocells in such a way that no two adjacent femtocells share the same frequency band; or by using the resource allocation scheme in [20], which allocates resources to femtocells based on the resource demands and interference conditions of the femtocells.

Hybrid access femtocells can choose to reject the data flows of non-CSG subscribers to satisfy the QoS requirements of CSG subscribers [4]. Also, real-time data flows should be prioritized over non-real-time data flows as the former requires a guaranteed bit rate (GBR) to be achieved. Let $\mathcal{C}_{h,1}, \mathcal{C}_{h,2}, \mathcal{C}_{h,3}$ and $\mathcal{C}_{h,4}$ denote the sets of GBR flows of CSG subscribers, non-GBR flows of CSG subscribers, GBR flows of non-CSG subscribers and non-GBR flows of non-CSG subscribers of HeNB h , respectively. These sets of data flows follow the same order in decreasing admission priority. Also, let $\mathcal{C}_{h,1}^a, \mathcal{C}_{h,2}^a, \mathcal{C}_{h,3}^a$ and $\mathcal{C}_{h,4}^a$ denote the set of data flows admitted

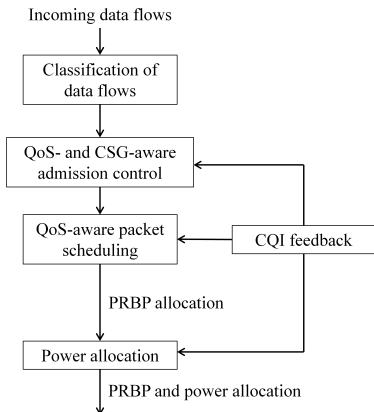


Fig. 2. Proposed Resource Management Framework.

from $\mathcal{C}_{h,1}$, $\mathcal{C}_{h,2}$, $\mathcal{C}_{h,3}$ and $\mathcal{C}_{h,4}$, respectively. Additionally, let $\mathcal{C}_h = \bigcup_{i=1}^4 \mathcal{C}_{h,i}$ and $\mathcal{C}_h^a = \bigcup_{i=1}^4 \mathcal{C}_{h,i}^a$. Furthermore, we define a_c as the admission indicator of data flow c whereby $a_c = 1$ if data flow c is admitted; otherwise, $a_c = 0$.

The achievable rate of data flow c on PRBP k at the t -th TTI can be expressed using Shannon's capacity formula as

$$r_{ck}(t) = B \log_2(1 + \Gamma_{ck}(t)), \quad (1)$$

where B is the bandwidth of a PRBP and $\Gamma_{ck}(t)$ is the signal-to-interference-plus-noise ratio (SINR) of PRBP k for data flow c at the t -th TTI, which can be expressed as follows:

$$\Gamma_{ck}(t) = \frac{p_k(t)g_{ck}(t)}{I_{ck}(t) + N_0}, \quad (2)$$

where $p_k(t)$, $g_{ck}(t)$ and $I_{ck}(t)$ are the power level transmitted on PRBP k , the channel gain and the interference power level experienced by data flow c on PRBP k , respectively, at the t -th TTI, and N_0 is the additive white Gaussian noise (AWGN) power. It is noteworthy that the value of $\Gamma_{ck}(t)$ can be obtained from the link adaptation module of LTE-based cellular systems based on the channel quality indicator (CQI) feedback from the UE. In fact, the link adaptation module can directly estimate the achievable spectral efficiency, i.e., $\log_2(1 + \Gamma_{ck}(t))$ by mapping the CQI value to a proper modulation scheme [21]. Let $\omega_{ck}(t)$ denote the PRBP assignment indicator of PRBP k to data flow c at the t -th TTI whereby $\omega_{ck}(t) = 1$ if PRBP k is allocated to data flow c ; otherwise $\omega_{ck}(t) = 0$.

III. PROPOSED RESOURCE MANAGEMENT SCHEME

In this section, we propose a resource management scheme depicted in Fig. 2 in which the incoming data flows are first classified into sets $\mathcal{C}_{h,1}$, $\mathcal{C}_{h,2}$, $\mathcal{C}_{h,3}$ and $\mathcal{C}_{h,4}$. Thereafter, admission control is performed to admit data flows based on their resource demands and admission priority. The resource demands are estimated based on the wideband channel conditions obtained from the CQI feedback. Then, packet scheduling and power allocation are performed to assign PRBPs and transmission powers to the admitted data flows based on the channel conditions and queue information of the data flows.

A. Traffic Classification and Admission Control

For admission control, we aim to maximize the number of admissible data flows, subject to the availability of the PRBPs, according to their admission priority and resource demands. The data flows which cannot be satisfied will be rejected, starting from the least prioritized ones.

Let $f_1(\mathbf{a}) = \sum_{c \in \mathcal{C}_{h,1}} a_c$, $f_2(\mathbf{a}) = \sum_{c \in \mathcal{C}_{h,2}} a_c$, $f_3(\mathbf{a}) = \sum_{c \in \mathcal{C}_{h,3}} a_c$ and $f_4(\mathbf{a}) = \sum_{c \in \mathcal{C}_{h,4}} a_c$ where $\mathbf{a} = [a_1 \dots a_{|\mathcal{C}_h|}]$. The admission control problem of any HeNB h can be formulated as a lexicographic optimization problem as

$$\text{lex max}_{\mathbf{a}} \mathbf{F}(\mathbf{a}) = \{f_1(\mathbf{a}), f_2(\mathbf{a}), f_3(\mathbf{a}), f_4(\mathbf{a})\} \quad (3)$$

subject to:

$$\sum_{c \in \mathcal{C}_h} a_c D_c \leq |\mathcal{K}_h| \quad (3a)$$

$$a_c \in \{0, 1\} \quad \forall c \in \mathcal{C}_h, \quad (3b)$$

where D_c is the resource demand of data flow c . Note that $|\mathcal{X}|$ is the cardinality of set \mathcal{X} . Constraint (3a) ensures that the sum of the resource demands of the admitted data flows does not exceed the number of PRBPs available. Constraint (3b) enforces that a_c takes only binary values. The "lexmax" operator means that the objective functions in \mathbf{F} is maximized iteratively following the lexicographic order, i.e., f_1 is maximized in the first iteration followed by f_2 in the second iteration and so on, all are subject to constraints (3a) and (3b); the operator also imposes a constraint such that the solution found for maximizing f_i in iteration $i \neq 1$ retains the value of f_j^* for $j = 1, \dots, i-1$ and f_j^* is the maximum value of f_j obtained in previous iterations.

In fact, (3) is a multi-objective optimization problem and thus a Pareto optimal [22] solution should be sought. With regard to (3), Pareto optimality is defined as follows:

Definition 1. Let \mathcal{A} be the set of feasible solutions for (3) and $i \in \{1, 2, 3, 4\}$. A solution, $\mathbf{a}^* \in \mathcal{A}$ is Pareto optimal if there does not exist another solution, $\mathbf{a} \in \mathcal{A}$ such that $\mathbf{F}(\mathbf{a}) \geq \mathbf{F}(\mathbf{a}^*)$, and at least one $f_i(\mathbf{a}) > f_i(\mathbf{a}^*)$.

The optimal solution to a lexicographic optimization problem such as (3) is said to be lexicographically optimal [23]. We provide the following definition with regard to (3).

Definition 2. Let \mathcal{A} be the set of feasible solutions for (3) and $i \in \{1, 2, 3, 4\}$. A solution, $\mathbf{a}^* \in \mathcal{A}$ is lexicographically optimal if there does not exist another solution, $\mathbf{a} \in \mathcal{A}$ such that $f_i(\mathbf{a}) > f_i(\mathbf{a}^*)$ for some i , and $f_j(\mathbf{a}) = f_j(\mathbf{a}^*)$ for all $j \in \{1, \dots, i-1\}$ if $i \neq 1$.

The lexicographic optimal solution to (3) can be proven to be Pareto-optimal in the following lemma.

Lemma 1. The lexicographically optimal solution to (3) is also Pareto-optimal.

Proof. Refer to Appendix A. \square

From analyzing (3), we obtain the following proposition.

Proposition 1. The problem in (3) can be classified as a 0-1 multi-objective knapsack problem.

Proof. Refer to Appendix B. \square

Since any 0-1 multi-objective knapsack problem is generally NP-hard [24], Proposition 1 implies that (3) is also NP-hard. However, due to its lexicographic nature, (3) can be solved efficiently using greedy approaches. In this work, we propose a low-complexity greedy admission control algorithm for (3) (cf. Algorithm 1). In this algorithm, we first estimate the resource demand of all data flows:

$$D_c = \begin{cases} \left\lceil \frac{R_{c,\text{req}}}{B\eta_{c,\text{wb}}} \right\rceil & \text{if data flow } c \text{ is a GBR flow} \\ 1 & \text{otherwise,} \end{cases} \quad (4)$$

where $R_{c,\text{req}}$ is the target bit rate required by GBR flow c and $\eta_{c,\text{wb}}$ is the achievable wideband spectral efficiency for GBR flow c . Based on Shannon's capacity formula, $\eta_{c,\text{wb}}$ can be calculated as $\log_2(1 + \Gamma_{c,\text{wb}})$ where $\Gamma_{c,\text{wb}}$ is the wideband SINR experienced across all PRBPs by the UE associated with data flow c . In LTE-based cellular systems, $\Gamma_{c,\text{wb}}$ can be estimated and obtained from the CQI feedback from the UE. In fact, the value of $\eta_{c,\text{wb}}$ can also be estimated in the link adaptation module of LTE-based cellular systems by mapping the wideband CQI to a proper modulation scheme [21]. After the resource demand estimation, a_c are set to zero for all $c \in \mathcal{C}_h$ and $\mathcal{C}_{h,1}^a, \mathcal{C}_{h,2}^a, \mathcal{C}_{h,3}^a$ and $\mathcal{C}_{h,4}^a$ are initialized as empty sets. Then, the data flows in \mathcal{C}_h are sorted as $\mathcal{C}_{h,1}^s$ in an order where the first data flow has the smallest resource demand followed by those with larger resource demands using `SortAscendDemand(.)`. Subsequently, the admission of each data flow in set $\mathcal{C}_{h,1}^s$ is evaluated using constraint (3a), starting from the first element of $\mathcal{C}_{h,1}^s$. If a data flow is admitted and constraint (3a) holds, the data flow will be admitted into set $\mathcal{C}_{h,1}^a$. After all the data flows of $\mathcal{C}_{h,1}^s$ are evaluated, steps 2-8 are repeated for $\mathcal{C}_{h,2}$, followed by $\mathcal{C}_{h,3}$ and $\mathcal{C}_{h,4}$. Algorithm 1 is efficient for solving (3) as it only incurs an asymptotic complexity of $\mathcal{O}(|\mathcal{C}_h|)$.

Algorithm 1 Greedy admission control algorithm

- 1: Estimate D_c for all $c \in \mathcal{C}_h$ using (4); Initialize $a_c = 0$ for all $c \in \mathcal{C}_h$;
Initialize $\mathcal{C}_{h,1}^a = \emptyset, \mathcal{C}_{h,2}^a = \emptyset, \mathcal{C}_{h,3}^a = \emptyset$ and $\mathcal{C}_{h,4}^a = \emptyset$
 - 2: $\mathcal{C}_{h,1}^s = \text{SortAscendDemand}(\mathcal{C}_h)$
 - 3: **for all** $c \in \mathcal{C}_{h,1}^s$ **do**
 - 4: Set $a_c = 1$ and evaluate constraint (3a)
 - 5: **if** constraint (3a) holds **then**
 - 6: $\mathcal{C}_{h,1}^a = \mathcal{C}_{h,1}^a \cup c$
 - 7: **else**
 - 8: $a_c = 0$
 - 9: **end if**
 - 10: **end for**
 - 11: Repeat steps 2-8 for $\mathcal{C}_{h,2}$ followed by $\mathcal{C}_{h,3}$ and $\mathcal{C}_{h,4}$.
-

Next, we show that Algorithm 1 always produces a Pareto-optimal solution to (3).

Theorem 1. *Algorithm 1 always gives a Pareto optimal solution to (3).*

Proof. Refer to Appendix C. \square

B. Packet Scheduling and Power Allocation

For packet scheduling and power allocation, our objective is to maximize the weighted sum rate of all admitted data flows

while prioritizing those of the CSG subscribers and guaranteeing bounded packet delays for the GBR flows. Mathematically for HeNB h , this can be represented as follows

$$\max \sum_{c \in \mathcal{C}_h^a} w_c(t) R_c(t) \quad (5)$$

subject to:

$$d_c(t+1) \leq d_{c,\text{max}} \quad \forall c \in \mathcal{C}_{h,1}^a \cup \mathcal{C}_{h,3}^a \quad (5a)$$

$$\sum_{c \in \mathcal{C}_h^a} \omega_{ck}(t) \leq 1 \quad \forall k \in \mathcal{K}_h \quad (5b)$$

$$\sum_{k \in \mathcal{K}_h} p_k(t) \leq P_{\text{max},h} \quad \forall k \in \mathcal{K}_h \quad (5c)$$

$$\omega_{ck}(t) \in \{0, 1\} \quad \forall k \in \mathcal{K}_h, \forall c \in \mathcal{C}_h^a \quad (5d)$$

$$p_k(t) \geq 0 \quad \forall k \in \mathcal{K}_h, \quad (5e)$$

where $\boldsymbol{\omega} = [\omega_{11}(t) \dots \omega_{|C_h^a| |\mathcal{K}_h|}(t)]$, $\mathbf{p} = [p_1(t) \dots p_{|\mathcal{K}_h|}(t)]$, $R_c(t) = \omega_{ck}(t) r_{ck}(t)$, $d_{c,\text{max}}$ is the maximum delay allowable for data flow c ; $R_c(t)$, $w_c(t)$ and $d_c(t)$ are the achievable data rate, the weighting coefficient and head-of-line packet delay of data flow c at the t -th TTI respectively. Here, $R_c(t) = \sum_{k \in \mathcal{K}_h} \omega_{ck} r_{ck}(t)$ where $r_{ck}(t)$ is the data rate achievable by data flow c on PRBP k at the t -th TTI. Constraint (5a) ensures that the packet delay of all GBR flows is bounded within the next TTI. In constraint (5b), each PRBP is ensured to be only assigned to one data flow. Constraint (5c) prevents the total transmission power of HeNB h on all of its PRBPs from exceeding the maximum allowable transmission power, $P_{\text{max},h}$. Constraint (5d) ensures that $\omega_{ck}(t)$ takes only binary values. Constraint (5e) guarantees that the transmission power of HeNB h on each PRBP is nonnegative. It is noteworthy that the problem formulation in (5) differs from that in our previous work [1] where power allocation is additionally considered in the former to further optimize the achievable throughput and fairness.

In (5), the setting of the weighting coefficient, $w_c(t)$ is instrumental for achieving different notions of fairness [25]. For GBR flows, we follow the modified-largest weighted delay first (M-LWDF) rule in [26], [27] by setting $w_c(t) = \frac{\gamma_c d_c(t)}{\bar{R}_c(t)}$; where $\gamma_c = -\frac{\log \delta_c}{d_{c,\text{max}}}$ with δ_c being the maximum probability of $d_c(t)$ exceeding $d_{c,\text{max}}$ and $\bar{R}_c(t)$ is the average transmission rate which can be estimated as $\bar{R}_c(t) = 0.8\bar{R}_c(t-1) + 0.2R_c(t)$ [28]. This weight setting allows for achieving a good balance between spectral efficiency, QoS provisioning and fairness for GBR flows [21]. For non-GBR flows, we follow the proportional fairness (PF) rule [26] by setting $w_c(t) = \frac{1}{R_c(t)}$ to attain a good tradeoff between spectral efficiency and fairness for the non-GBR flows [21].

Obtaining the global optimal solution to the non-convex mixed-integer programming problem in (5) is difficult and computationally exhaustive. As such, we propose a sub-optimal method for solving (5). In this method, we first assume that power allocation has been performed across all PRBPs of each HeNB. Then, the problem in (5) can be rewritten as:

$$\max_{\boldsymbol{\omega}} \sum_{c \in \mathcal{C}_h^a} w_c(t) R_c(t) \quad (6)$$

subject to (5a), (5b) and (5d).

As we have set the weighting coefficients according to the M-LWDF rule, the packet scheduling problem in (6) can be solved using the M-LWDF scheduling rule [21], [29] whereby an M-LWDF metric is calculated for each data flow on each PRBP. Then, for each PRBP, the data flow with the highest M-LWDF metric will be allocated the PRBP. Here, the M-LWDF metric can be given as

$$m_{ck} = \begin{cases} \frac{\gamma_c d_c(t) r_{ck}(t)}{R_c(t)} & \text{if data flow } c \text{ is a GBR flow} \\ \frac{r_{ck}(t)}{R_c(t)} & \text{otherwise.} \end{cases} \quad (7)$$

In (7), $r_{ck}(t)$ can be estimated based on the CQI value from the corresponding UE in LTE-based cellular systems. After calculating the M-LWDF metrics for all data flows on each PRBP using (7), the data flow, e.g., data flow c which has the largest $m_{ck}(t)$, denoted by c^* , will be allocated PRBP k . This can be mathematically expressed as

$$\omega_{c^*k}(t) = \begin{cases} 1 & c^* = \arg \max_c m_{ck}(t) \\ 0 & \text{otherwise} \end{cases} \quad \forall k \in \mathcal{K}_h. \quad (8)$$

However, the M-LWDF scheduling rule in (8) does not guarantee the fulfillment of constraint (5a). As such, we propose a packet scheduling algorithm shown in Fig. 3, which prioritizes the urgent GBR flows with critical delay conditions. In this algorithm, an *urgent list*, which stores the urgent GBR flows, and a *normal list*, which stores all the non-GBR flows and non-urgent GBR flows are created. The proposed algorithm begins by scheduling over the data flows in the urgent list followed by those in the normal list if free PRBPs are available using the M-LWDF scheduling rule. This scheduling process is performed every one TTI. The urgent GBR flows are identified using the following inequality

$$d_c(t) > \alpha_c d_{c,\max}, \quad (9)$$

where α_c is a fractional value which falls within [0, 1] for GBR flow c .

The value of α_c can be fixed but different values of α_c may lead to different performance. For instance, if α_c is set too small, GBR flow c will always be scheduled and other data flows may suffer from resource starvation. If α_c is set too large, GBR flow c may suffer from a high packet loss rate as the probability of exceeding $d_{c,\max}$ is high. Also, since each GBR flow may experience a different delay, priority should be given to GBR flows with longer delays. To address these issues, we propose to dynamically adapt the value of α_c based on the queue information of GBR flow c . In our proposed packet scheduling algorithm, adaptation of α_c is performed every one LTE frame of 10 TTIs. Let $q_c(l)$ denote the queue length of the GBR flow c at the l -th frame. The following queue equation holds

$$q_c(l+1) - q_c(l) = u_c(l) + b_c(l), \quad (10)$$

where $u_c(l)$ and $b_c(l)$ are the amounts of data arrived at and transmitted from the queue of GBR flow c during the l -th frame respectively. Let T_f be the time duration of one LTE frame, the number of frames corresponds to the maximum delay allowable for GBR flow c can be calculated as $M_c =$

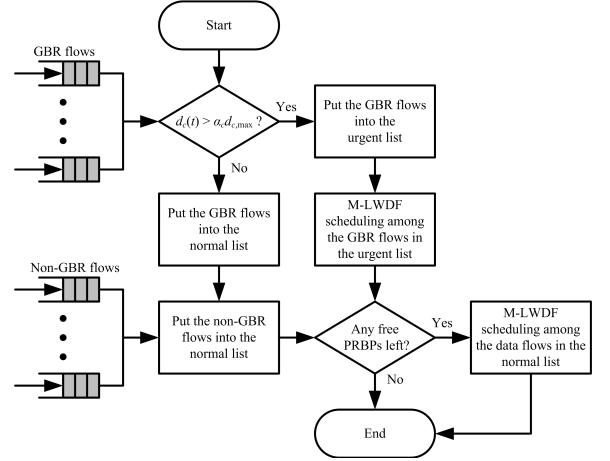


Fig. 3. Proposed scheduling algorithm.

$\frac{d_{c,\max}}{T_f}$. To guarantee bounded packet delays for GBR flow c , the following inequality must be satisfied:

$$\sum_{n=0}^{M_c-1} b_c(l+n) \geq q_c(l). \quad (11)$$

The inequality in (11) implies that the amount of data waiting in the queue at the l -th frame must be transmitted before the maximum allowable delay is exceeded, that is, when $(l + M_c)$ -th frame is reached. Using (10), (11) can be equivalently expressed as (See Appendix D for the detailed derivation):

$$\sum_{n=0}^{M_c-1} u_c(l - M_c + n) \geq q_c(l). \quad (12)$$

Using (10), the left-hand side of (12) can be derived as follows, which can be estimated in LTE-based cellular systems:

$$\sum_{n=0}^{M_c-1} u_c(l - M_c + n) = \sum_{n=0}^{M_c-1} b_c(l - M_c + n) - q_c(l - M_c) + q_c(l). \quad (13)$$

To ensure fulfillment of (12), α_c should be set smaller when the left-hand side of (12) almost equals its right-hand side, i.e., $\sum_{n=0}^{M_c-1} u_c(l - M_c + n) \approx q_c(l)$, and α_c should be set larger when the left-hand side of (12) is much larger than its right-hand side, i.e., $\sum_{n=0}^{M_c-1} u_c(l - M_c + n) \gg q_c(l)$. To do this, we first define the following:

$$y_c(l) = \frac{q_c(l)}{\sum_{n=0}^{M_c-1} u_c(l - M_c + n)}. \quad (14)$$

Then, we propose to adapt α_c as follows

$$\alpha_c = \begin{cases} 1 - y_c(l) & \text{if } y_c(l) < 1 \\ 0 & \text{otherwise.} \end{cases} \quad (15)$$

The setting in (15) allows each GBR flow c to be prioritized only if $y_c(l)$ is close to one, i.e., when $\sum_{n=0}^{M_c-1} u_c(l - M_c + n) \approx q_c(l)$; otherwise, the GBR flow c will not be prioritized, i.e., when $\sum_{n=0}^{M_c-1} u_c(l - M_c + n) \gg q_c(l)$. With this setting, bounded packet delays can be ensured for GBR flows while avoiding resource starvation in non-GBR flows.

With the proposed packet scheduling algorithm in Fig. 3, a sub-optimal PRBP allocation solution is obtained. It is noteworthy that the packet scheduling algorithm in Fig. 3 has an asymptotic complexity of $\mathcal{O}(|\mathcal{C}_h^a| |\mathcal{K}_h|)$ and the proposed adaptation strategy for α_c has an asymptotic complexity of $\mathcal{O}(M_{\max}(|\mathcal{C}_{h,1}^a| + |\mathcal{C}_{h,3}^a|))$ where $M_{\max} = \max\{M_c\}_{c=1}^{|\mathcal{C}_{h,1}^a| + |\mathcal{C}_{h,3}^a|}$.

After solving (6) for PRBP allocation, (5) is reduced to:

$$\max_{\mathbf{p}} \sum_{c \in \mathcal{C}_h^a} w_c(t) R_c(t) \quad (16)$$

subject to (5c) and (5e).

We employ the dual decomposition method in [30] to solve (16). Firstly, the Lagrangian of (16) can be written as follows:

$$\mathcal{L}(\mathbf{p}, \lambda) = \sum_{c \in \mathcal{C}_h^a} w_c(t) R_c(t) + \lambda \left(P_{\max, h} - \sum_{k \in \mathcal{K}_h} p_k \right), \quad (17)$$

where λ is the Lagrange multiplier corresponding to constraint (5c). Then, the Lagrange dual function of (16) can be expressed as $\mathbf{D}(\lambda) = \max_{\mathbf{p}} \mathcal{L}(\mathbf{p}, \lambda)$. Let (16) be the primal optimization problem, the dual optimization problem can be expressed as:

$$\min_{\lambda} \mathbf{D}(\lambda) \quad (18)$$

subject to $\lambda \geq 0$.

It is obvious that (18) is a convex problem which can be solved using convex optimization techniques. However, the solution to the dual problem generally serves as the upper bound of the solution to the primal problem in (16). In other words, there may exist a non-zero duality gap, i.e., the difference between the optimal values of (18) and (16). Nevertheless, if the duality gap is zero, the optimal solution to (18) will be the same optimal solution to (16). With regard to (16) and (18), it can be shown in the following theorem that the duality gap approaches zero if the number of PRBPs is sufficiently large.

Theorem 2. *The duality gap between (18) and (16) is nearly zero if the number of PRBPs is sufficiently large.*

Proof. Refer to Appendix E. \square

From Theorem 2, the optimal solution to (18) will approximate that to (16) if the number of PRBPs is sufficiently large.

With Theorem 2, we can solve for (16) by assuming that the number of PRBPs is sufficiently large. Firstly, we define $C(k)$ as the data flow that is allocated PRBP k , i.e., $C(k) = c \in \mathcal{C}_h$ whereby $w_{c_k}(t) = 1$. Using this notation, notation $w_{c_k}(t)$ in (17) can be removed and (17) can be rewritten by substituting (1) and (2) as

$$\begin{aligned} \mathcal{L}(\mathbf{p}, \lambda) &= \sum_{k \in \mathcal{K}_h} w_{C(k)}(t) B \log_2 \left(1 + \frac{p_k(t) g_{C(k)k}(t)}{I_{C(k)k}(t) + N_0} \right) \\ &+ \lambda \left(P_{\max, h} - \sum_{k \in \mathcal{K}_h} p_k(t) \right). \end{aligned} \quad (19)$$

Algorithm 2 Bisection method

```

1: Initialize  $\lambda_{\min} = 0$  and set  $\lambda_{\max}$  to an arbitrarily large value.
2: repeat
3:    $\lambda = \frac{\lambda_{\min} + \lambda_{\max}}{2}$ 
4:   for all  $k \in \mathcal{K}_h$  do
5:     Calculate  $p_k(t)$  using (20).
6:   end for
7:   if  $|P_{\max, h} - \sum_{k \in \mathcal{K}_h} p_k(t)| < \epsilon$  then
8:     break
9:   else
10:    if  $P_{\max, h} > \sum_{k \in \mathcal{K}_h} p_k(t)$  then
11:       $\lambda_{\max} = \lambda$ 
12:    else
13:       $\lambda_{\min} = \lambda$ 
14:    end if
15:  end if
16: until  $\lambda$  and  $\mathbf{p}$  converge.

```

By using the Karush-Kuhn-Tucker (KKT) conditions [30], the transmission power of the HeNB on each PRBP can be derived as:

$$p_k(t) = \left[\frac{w_{C(k)}(t) B}{\lambda \ln 2} - \frac{I_{C(k)k}(t) + N_0}{g_{C(k)k}(t)} \right]^+ \quad \forall k \in \mathcal{K}_h, \quad (20)$$

where $[x]^+$ is equivalent to $\max(0, x)$. In LTE-based cellular systems, the second term of the right-hand side of (20) can be estimated based on the CQI feedback from the corresponding UE and the power allocation information of the previous TTI. From the KKT conditions, the following must hold:

$$\lambda \left(P_{\max, h} - \sum_{k \in \mathcal{K}_h} p_k \right) = 0. \quad (21)$$

It is obvious from (20) that λ can only be non-zero for any feasible power allocation which satisfies constraint (5c), which implies that:

$$P_{\max, h} = \sum_{k \in \mathcal{K}_h} p_k. \quad (22)$$

As such, the solution to (18) can be obtained by solving (22) via a bisection method shown in Algorithm 2 and the power allocation solution for (16) can subsequently be obtained. It is noteworthy that (20) allows for increasing the transmission power on the resources allocated to the data flows which have low throughput achieved in the past, i.e., low long-term average rate. This is attributed to the setting of $w_{C(k)}(t)$ which makes (20) to be inversely proportional to the long-term average rate (See Section III-B). When the long-term average rate is low, its inverse will increase the value of (20) as well as the SINR, thereby increasing the achievable throughput and improving fairness. This also remedies the problem of data flows being constantly allocated PRBPs with poor channel quality. The asymptotic complexity of Algorithm 2 is $\mathcal{O}(N_b |\mathcal{K}_h|)$ where N_b is the number of iterations required by Algorithm 2 for the power allocation optimization to converge.

IV. PERFORMANCE EVALUATION

The proposed resource management scheme is evaluated using the LTE-Sim simulator [19], [31]. In the simulation setup, a two-dimensional building of the 5×5 apartment grid

TABLE I
LTE-BASED FEMTOCELL NETWORK SIMULATION SETTING

| Parameter | Setting |
|-----------------------------|-------------------------------------|
| Frame Structure | Frequency division duplexing |
| Bandwidth | 20 MHz (100 Sub-channels) |
| Simulation Duration | 10 000 ms |
| Traffic Model | Video, VoIP and best-effort |
| Maximum Delay for GBR Flows | 100 ms |
| HeNBs' Transmission Power | 20 dBm |
| Path Loss Model | Urban indoor propagation model [19] |
| Channel Fading Model | Rayleigh |
| Shadowing | Log-normal |
| Number of UEs per femtocell | 5 CSG and 5 non-CSG subscribers |
| UE speed | 3 km/h |

type [32], where each apartment has an area of $10 \times 10 \text{ m}^2$ and accommodates a femtocell, is considered¹. We employ the frequency reuse scheme in [19] with a frequency reuse of $\frac{1}{4}$ over the femtocell network where each femtocell is allocated a 5 MHz bandwidth. Each UE carries one video flow, one voice over Internet Protocol (VoIP) flow and one best-effort flow. The target bit rates of the video and VoIP flows are set to 128 kbps and 8 kbps respectively. For the best-effort flows, an infinite buffer model is used. For channel modeling, an urban indoor path loss model in [19]: $127 + 30 \log d$ in dB, where d is the distance between the HeNB and the FUE in km, is considered. Also, log-normal shadowing with zero mean and 8 dB standard deviation is implemented. We also consider time-varying flat Rayleigh fading in our simulation scenario. The time-varying flat Rayleigh fading is implemented based on the Jakes model [33], which accounts for the UE speed, the number of paths and the subchannel frequency. Here, we set the UE speed to 3 km/h and the number of paths is uniformly chosen from $\{6, 8, 10, 12\}$. The UE mobility is set based on a random direction mobility model whereby the UE travels with a constant speed and randomly chooses a travel direction when a boundary is reached. Other parameters used are shown in Table I.

For comparison, the PF, M-LWDF [29], exponential (EXP) scheduling rule [34], frame-level scheduler (FLS) [28] and the scheme proposed in [1] are used for benchmarking. For the proposed scheme and that in [1], we set $\delta_c = 0.005$ for the GBR flows of CSG subscribers and $\delta_c = 0.01$ for the GBR flows of non-CSG subscribers. For the M-LWDF scheme, $\delta_c = 0.005$ is set. In addition, $\epsilon = 10^{-10}$ is set in Algorithm 2 of the proposed scheme. The simulation results obtained are averaged over five runs where each run realizes different channel conditions.

A. Performance with Varying Activity Ratios

In this section, the activity ratio, r_a [32] is used as the varying parameter, which is defined as the probability that an HeNB is active. Here, we evaluate the proposed scheme with

¹Here, we only consider the hybrid access femtocell scenario because the main objective of this work is to investigate resource allocation schemes for hybrid access femtocell systems and their performance impact to the hybrid access femtocell users. Therefore, heterogeneous scenarios such as those where the femtocell apartment is covered by a macrocell are out of the scope of this work.

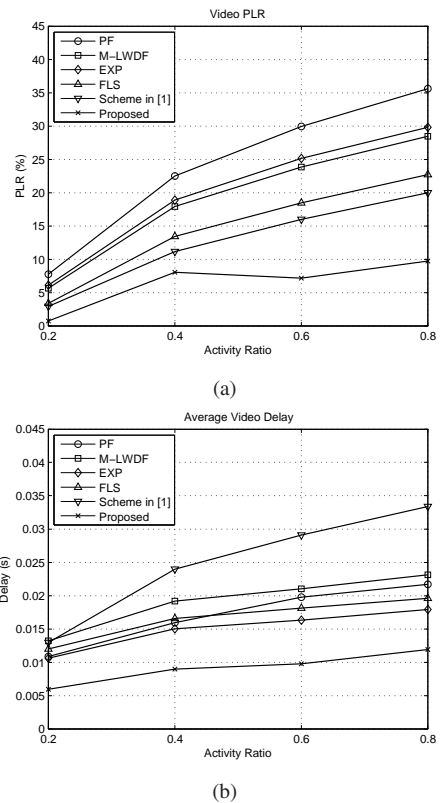
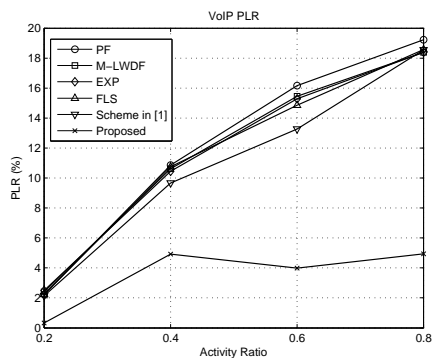


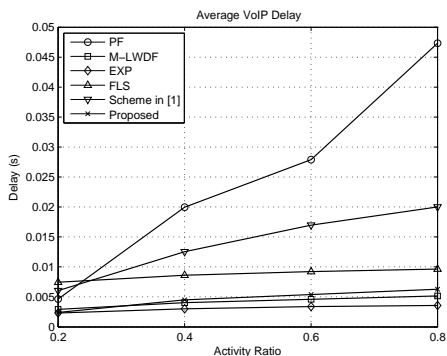
Fig. 4. (a) PLR and (b) packet delay performance of video flows with varying activity ratios.

$r_a = 0.2, 0.4, 0.6$ and 0.8 ; and the CQI feedback is reported every 2 ms.

Fig. 4 shows the packet loss rate (PLR) and packet delay performance of video flows. In Fig. 4(a), the proposed scheme achieves a PLR of less than 10% while other schemes achieve a PLR of less than 21%-36% across all values of r_a . In Fig. 4(b), the proposed scheme outperforms all other schemes by attaining the lowest average packet delay. The superior performance of the proposed scheme is attributed to the proposed admission control algorithm that dynamically admits data flows based on their admission priority and evaluation of constraint (3a), thus ensuring a high degree of QoS satisfaction for all data flows in the femtocell network. In addition, the proposed packet scheduling algorithm provides bounded packet delay for GBR flows, which improves the QoS performance. The proposed power allocation algorithm further reduces the video PLR by up to 10% compared to the scheme in [1] because it further optimizes (5). On the other hand, we can observe that the scheme in [1] incurs the highest average packet delay, albeit with a very low PLR. This is because some data flows may have constantly received PRBPs with low channel quality, thus diminishing their scheduling metric in (7). This reduces the chance of these data flows being scheduled earlier, hence increasing the packet delay. This problem has been remedied by the proposed scheme via power allocation in Algorithm 2 which increases the transmission power on the PRBPs allocated to those data flows in order to increase their achievable rate, thereby increasing the metric in (7).



(a)



(b)

Fig. 5. (a) PLR and (b) packet delay performance of VoIP flows with varying activity ratios.

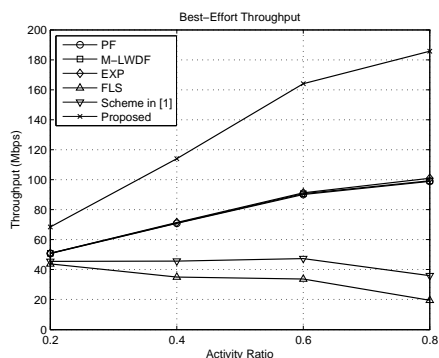
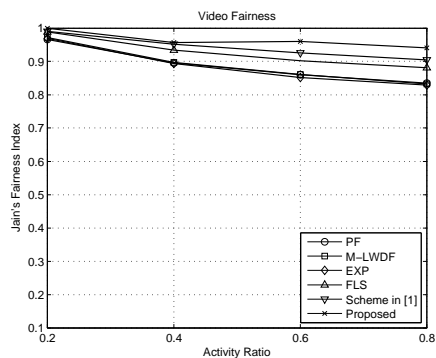


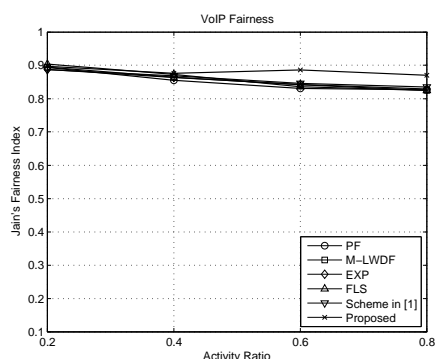
Fig. 6. Throughput performance of best-effort flows with varying activity ratios.

Fig. 5 shows the PLR and packet delay performance of VoIP flows. In Fig. 5(a), the proposed scheme again outperforms the other schemes by achieving the lowest PLR of up to 5% while others achieve PLRs of up to 18%-19% across all values of r_a . It is noted that the addition of power allocation in our proposed scheme improves by up to 13% compared to the scheme in [1]. As depicted in Fig. 5(b), although the proposed scheme does not achieve the lowest average packet delay, it is still comparable with the PF, M-LWDF and EXP schemes on this aspect.

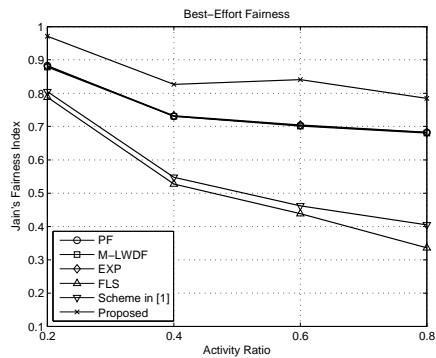
Fig. 6 illustrates the best-effort throughput performance of the schemes. It can be observed that the proposed scheme attain a substantial throughput performance gain over the other schemes. In particular, the difference between the best-



(a)



(b)



(c)

Fig. 7. Fairness performance of (a) video, (b) VoIP and (c) best-effort flows with varying activity ratios.

effort throughput of the proposed scheme and that of the scheme in [1] is huge. This is because the proposed scheme is complemented with power allocation via Algorithm 2 which improves the SINR by increasing the corresponding transmission power level, hence leading to a higher achievable best-effort throughput.

We also evaluate the fairness performance of all traffic classes using Jain's fairness index, which is given as $\frac{(\sum_{c=1}^N x_c)^2}{N \sum_{c=1}^N x_c^2}$ [35] where N is the number of data flows. For fairness evaluation of GBR flows, $x_c = \frac{R_c}{R_c^{req}}$. On the other hand, $x_c = R_c$ for non-GBR flows. Fig. 7 shows the fairness performance of all traffic classes. We observe that the proposed scheme achieves a relatively high degree of fairness for all traffic classes. This is attributed to the weighting coefficients

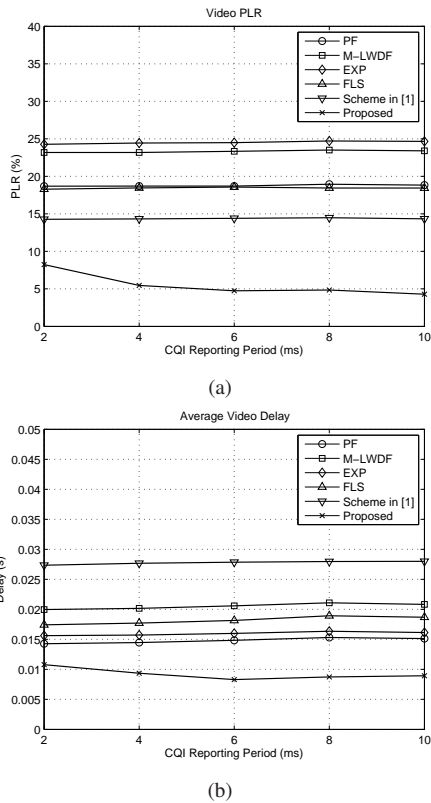


Fig. 8. (a) PLR and (b) packet delay performance of video flows with varying CQI reporting periods.

in (5) which are set to achieve PF. By maximizing (5) with the proposed packet scheduling and power allocation algorithm, a high degree of fairness can be achieved. On the other hand, the scheme in [1] has a significantly lower degree of fairness compared to the proposed scheme. This is because the scheme in [1] does not account for power allocation, resulting in throughput degradation for some best-effort flows which receive low-quality PRBPs. Furthermore, if an urgent GBR flow experiences poor channel conditions, the scheme in [1] will allocate most of the available PRBPs to this data flow until its packet delay constraint is satisfied, thereby resulting in resource starvation for other data flows, especially the non-GBR flows.

B. Performance with Varying CQI Reporting Periods

In this section, we evaluate the proposed scheme with CQI reporting periods of 2, 4, 6, 8 and 10 ms. Here, $r_a = 0.5$ is set for the network scenario.

Fig. 8 shows the video PLR and packet delay performance with varying CQI reporting periods. Overall, both the proposed scheme and the existing schemes show consistent performance across different CQI reporting periods. This is because the channel does not vary rapidly since the UE speed is slow, which is reasonable for those moving in indoor environments such as residential apartments. The proposed scheme remains superior to other schemes due to its dynamic admission control and resource allocation mechanisms.

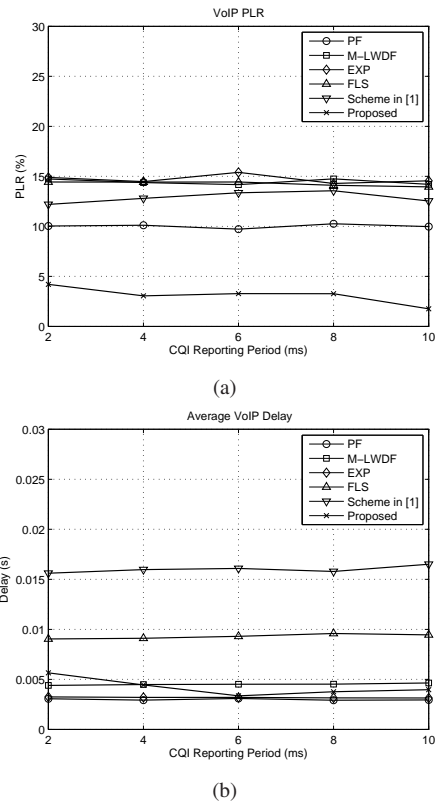


Fig. 9. (a) PLR and (b) packet delay performance of VoIP flows with varying CQI reporting periods.

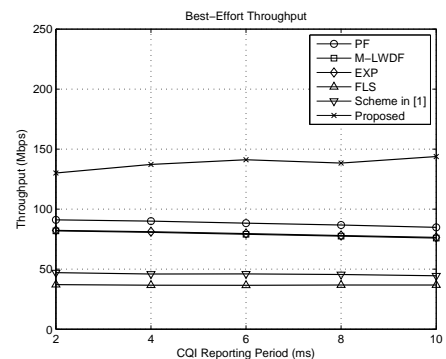


Fig. 10. Throughput performance of best-effort flows with varying CQI reporting periods.

Similarly, all the schemes show consistent VoIP PLR and packet delay performance across different CQI reporting periods, as shown in Fig. 9.

In Fig. 10, consistent best-effort throughput performance is demonstrated by all the schemes with the proposed scheme being the best.

Fig. 11 shows the fairness performance of the schemes across different CQI reporting periods. It can be observed that the trends are similar to those in Fig. 7 wherein the proposed scheme exhibits relatively high fairness over all traffic classes. In particular, the proposed scheme outperforms other schemes in best-effort fairness.

V. CONCLUSION

In this paper, we have proposed a dynamic yet fair resource management scheme for QoS provisioning in LTE-based hybrid access femtocells. The proposed scheme first classifies traffic on the incoming data flows and performs admission control over them with higher priority given to data flows of CSG subscribers, subject to resource availability. The admission control problem is formulated as a lexicographic optimization problem and an optimal greedy solution algorithm is proposed. Packet scheduling and power allocation are performed over the admitted data flows to maximize the weighted sum rate of each femtocell. A sub-optimal delay-bounded packet scheduling algorithm and a dual decomposition-based power allocation algorithm have been devised for the weighted sum rate maximization by taking into account the traffic types, average rate, channel, delay and queue length of each data flow. Simulation results have shown that the proposed scheme achieves substantial PLR and packet delay performance gain for GBR flows such as video and VoIP and significant throughput gain for non-GBR best-effort flows over the existing schemes. The proposed scheme has also attained a relatively high degree of fairness for all traffic classes. This implies that information such as traffic types, channel conditions, average rate, packet delay and queue length are essential in the design of the resource management scheme for hybrid access femtocells in LTE-based cellular networks in order to achieve high throughput, QoS provisioning and high fairness. The proposed scheme is applicable to any LTE-based cellular systems such as 5G.

APPENDIX A PROOF OF PROPOSITION 1

We first define a 0-1 multi-objective knapsack problem as follows:

Definition 3. Suppose that there is a set of N items to be filled in a knapsack with a maximum weight capacity of W . Each item i has a positive weight of w_i and corresponds to J nonnegative profits, i.e., $\{v_{i1}, \dots, v_{ij}, \dots, v_{iJ}\}$ where $i = 1, \dots, N$. The problem is to fill the knapsack such that all J profits are maximized without having the weight of the knapsack exceeding its maximum weight capacity.

From Definition 3, it is obvious in (3) that $i = c, N = |\mathcal{C}_h|, W = |\mathcal{K}_h|, w_i = D_c, J = 4, v_{ij} = 1$ for $i \in \mathcal{C}_{h,j}$ and $v_{ij} = 0$ for $i \notin \mathcal{C}_{h,j}$. This completes the proof.

APPENDIX B PROOF OF LEMMA 1

In lexicographic optimization, the solution may either be obtained when all objective functions are optimized or when a unique solution is found before every objective function is optimized. In this case, we divide the proof into two parts as follows:

Suppose \mathbf{a}^* , which has been obtained when all the objective functions are optimized in (3), is not Pareto optimal. Thus, there exists another solution, $\mathbf{a} \in \mathcal{A}$ such that $\mathbf{F}(\mathbf{a}) \geq \mathbf{F}(\mathbf{a}^*)$, and at least one $f_i(\mathbf{a}) > f_i(\mathbf{a}^*)$ where $i \in \{1, 2, 3, 4\}$. By

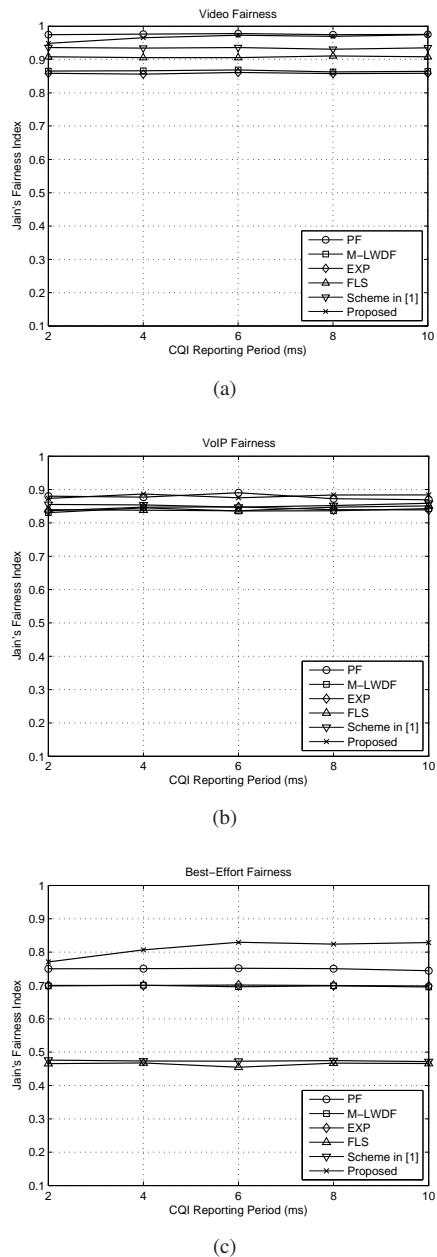


Fig. 11. Fairness performance of (a) video, (b) VoIP and (c) best-effort flows with varying CQI reporting periods.

definition of (3), we cannot have $f_i(\mathbf{a}) > f_i(\mathbf{a}^*)$. Hence, it is only possible that $f_i(\mathbf{a}) = f_i(\mathbf{a}^*)$. This contradicts the assumption of at least one $f_i(\mathbf{a}) > f_i(\mathbf{a}^*)$.

Suppose \mathbf{a}^* , which has been obtained before all the objective functions are optimized in (3), is not Pareto optimal. Thus, there exists another solution, $\mathbf{a} \in \mathcal{A}$ such that $\mathbf{F}(\mathbf{a}) \geq \mathbf{F}(\mathbf{a}^*)$, and at least one $f_i(\mathbf{a}) > f_i(\mathbf{a}^*)$. Assume that \mathbf{a}^* is found when optimizing f_i . This implies that \mathbf{a}^* is a unique solution. Therefore, it is only possible that $f_i(\mathbf{a}) = f_i(\mathbf{a}^*)$. This again contradicts the assumption of at least one $f_i(\mathbf{a}) > f_i(\mathbf{a}^*)$. This completes the proof.

APPENDIX C
PROOF OF THEOREM 1

Assume that, in the first iteration of Algorithm 1 which maximizes f_1 subject to constraints (3a) and (3b), data flow $c^* \in \mathcal{C}_{h,1}$ imposes the smallest resource demand, D_{c^*} . Suppose there exists an optimal solution \mathbf{a} whereby data flow c^* is rejected, $a_{c^*} = 0$. If any data flow that is originally admitted in \mathbf{a} is rejected, and $a_{c^*} = 1$ is set, constraint (3a) remains satisfied and the optimal value of f_1 will not be reduced. Thus, this proves that there always exists an optimal solution with $a_{c^*} = 1$. Next, suppose \mathbf{a} is an optimal solution with $a_{c^*} = 1$ for the maximization of f_1 subject to constraints (3a) and (3b). Then, $\mathbf{a}' = \mathbf{a} \setminus \{a_{c^*}\}$ is an optimal solution for the maximization of f_1 with $\sum_{c \in \mathcal{C}_h \setminus \{c^*\}} a_c D_c \leq |\mathcal{K}| - D_{c^*}$ and constraint (3b). To prove this, we suppose that \mathbf{a}' is not optimal. Then, there exists another solution \mathbf{a}'' such that $\mathbf{a}'' \cup \{a_{c^*}\}$ maximizes f_1 subject to constraints (3a) and (3b). However, this contradicts \mathbf{a} , which is the optimal solution. Thus, after choosing the data flow with the smallest resource demand to be admitted, the optimization problem becomes smaller but remains in the same form as the original problem. Hence, with similar reasoning, the solution obtained by choosing the data flow which has the next smallest resource demand to be admitted is optimal to the smaller problem. By induction, the solution obtained by Algorithm 1 is optimal to f_1 .

In the second iteration where f_2 is maximized, the solution obtained is subject to an additional constraint in which the optimal value of f_1 is retained. Since Algorithm 1 performs the optimization over $\mathcal{C}_{h,2}$ when maximizing f_2 , the solution obtained from the previous iteration which maximizes f_1 remains intact. Thus, the additional constraint will not be violated. Then, with the similar proof of optimality for f_1 , the solution obtained from maximizing f_2 is optimal; likewise for the maximization of f_3 and f_4 in the subsequent iterations.

Since Algorithm 1 follows the lexicographic order in solving (3), the solution obtained is lexicographically optimal by Definition 2. By Lemma 1, it is also Pareto-optimal.

APPENDIX D
DERIVATION OF (12)

Let (10) be rewritten as

$$b_c(l) = u_c(l) + q_c(l) - q_c(l+1). \quad (23)$$

Then, substituting (23) into (11) yields

$$\begin{aligned} & \sum_{n=0}^{M_c-1} u_c(l+n) \\ & \geq q_c(l) + \sum_{n=0}^{M_c-1} \{q_c(l+n+1) - q_c(l+n)\} \\ & = q_c(l) + \{q_c(l+M_c) - q_c(l)\} \\ & = q_c(l+M_c). \end{aligned} \quad (24)$$

Let $q_c(l')$ denote the queue length of data flow c during the l' -th frame where $l' = l + M_c$, (24) can be rewritten as

$$\sum_{n=0}^{M_c-1} u_c(l' - M_c + n) \geq q_c(l'). \quad (25)$$

Equivalently, (25) can be expressed as (12).

APPENDIX E
PROOF OF THEOREM 2

Without loss of generality, we omit the index t from $p_k(t)$ and define $G_k(p_k) = \omega_{ck} \log_2 \left(1 + \frac{p_k g_{ck}}{I_{ck} + N_0}\right)$ and $s_k(p_k) = p_k$. Then, (16) can be rewritten as:

$$\max_{\mathbf{p}} \sum_{k \in \mathcal{K}_h} G_k(p_k) \text{ subject to } \sum_{k \in \mathcal{K}_h} s_k(p_k) \leq S, \quad (26)$$

where $G_k(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$, $\mathbf{p} \in \mathbb{R}^{|\mathcal{K}_h|}$, and U is the number of constraints. By mapping (26) to (16), it can be observed that $S = P_{\max, h}$ and $U = 1$. To prove that the duality gap is zero, we first provide the definition of the time-sharing condition as follows:

Definition 4. Let \mathbf{p}_1^* and \mathbf{p}_2^* be the optimal solutions to (26) with $S = S_1$ and $S = S_2$, respectively. The optimization problem of the form in (26) satisfies the time-sharing condition, if for any $0 \leq \delta \leq 1$, there exists a feasible solution \mathbf{p}_3^* that satisfies the following:

$$\sum_{k \in \mathcal{K}_h} s_k(p_{k,3}^*) \leq \delta S_1 + (1 - \delta) S_2$$

and

$$\sum_{k \in \mathcal{K}_h} G_k(p_{k,3}^*) \geq \delta \sum_{k \in \mathcal{K}_h} G_k(p_{k,1}^*) + (1 - \delta) \sum_{k \in \mathcal{K}_h} G_k(p_{k,2}^*).$$

Firstly, we prove that $\sum_{k \in \mathcal{K}_h} G_k(p_k)$ is a concave function of S , which follows from Definition 4. Let \mathbf{p}_3^* be the optimal solution to (26) with $S_3 = \delta S_1 + (1 - \delta) S_2$ for some $0 \leq \delta \leq 1$. Hence, the time-sharing property indicates that there exists a feasible solution \mathbf{p}_3 such that:

$$\sum_{k \in \mathcal{K}_h} s_k(p_{k,3}^*) \leq S_3 = \delta S_1 + (1 - \delta) S_2$$

and

$$\sum_{k \in \mathcal{K}_h} G_k(p_{k,3}^*) \geq \delta \sum_{k \in \mathcal{K}_h} G_k(p_{k,1}^*) + (1 - \delta) \sum_{k \in \mathcal{K}_h} G_k(p_{k,2}^*).$$

Further, it implies that:

$$\begin{aligned} \sum_{k \in \mathcal{K}_h} G_k(p_{k,3}^*) & \geq \sum_{k \in \mathcal{K}_h} G_k(p_{k,3}) \\ & \geq \delta \sum_{k \in \mathcal{K}_h} G_k(p_{k,1}^*) + (1 - \delta) \sum_{k \in \mathcal{K}_h} G_k(p_{k,2}^*). \end{aligned}$$

Hence, $\sum_{k \in \mathcal{K}_h} G_k(p_k)$ is a concave function of S . Then, it can be shown that the time-sharing condition is always satisfied in a multi-carrier system, such as LTE/LTE-A, with the number of PRBPs approaching infinity, i.e., $|\mathcal{K}_h| \rightarrow \infty$ [36]. Let \mathbf{p}_1^* and \mathbf{p}_2^* be two power allocation solutions. In this case, δ percentage of the total number of subcarriers in PRBP k are allocated $p_{k,1}^*$ and $(1 - \delta)$ percentage of the total number of subcarriers in PRBP k are allocated $p_{k,2}^*$. Therefore, $\sum_{k \in \mathcal{K}_h} G_k$ becomes approximately a linear combination $\sum_{k \in \mathcal{K}_h} \left(\delta G_k(p_{k,1}^*) + (1 - \delta) G_k(p_{k,2}^*) \right)$. When $|\mathcal{K}_h| \rightarrow \infty$, the approximation is exact. Hence, the time-sharing condition is satisfied. This proves that the duality gap approaches zero if the number of PRPBs is sufficiently large.

REFERENCES

- [1] Y. L. Lee, J. Loo, and T. C. Chuah, "A Resource Management Scheme for Hybrid Access Femtocells in LTE/LTE-A Networks," in *Proc. IEEE PIMRC*, Hong Kong, China, Aug.-Sep. 2015, pp. 1134–1138.
- [2] H. Zhang, X. Chu, W. Guo, and S. Wang, "Coexistence of Wi-Fi and Heterogeneous Small Cell Networks Sharing Unlicensed Spectrum," *IEEE Commun. Mag.*, vol. 53, no. 3, pp. 158–164, Mar. 2015.
- [3] H. Zhang, C. Jiang, R. Q. Hu, and Y. Qian, "Self-Organization in Disaster Resilient Heterogeneous Small Cell Networks," *IEEE Netw.*, To appear.
- [4] A. Golaup, M. Mustapha, and L. B. Patanongpibul, "Femtocell Access Control Strategy in UMTS and LTE," *IEEE Commun. Mag.*, vol. 47, no. 2, pp. 117–123, Sep. 2009.
- [5] G. de la Roche, A. Valcarce, D. Lopez-Perez, and J. Zhang, "Access Control Mechanisms for Femtocells," *IEEE Commun. Mag.*, vol. 48, no. 1, pp. 33–39, Jan. 2010.
- [6] "Technical Specification Group Services and System Aspects; Service requirements for Home Node B (HNB) and Home eNode B (HeNB)," Third Generation Partnership Project (3GPP), Sophia-Antipolis, France, Oct. 2014. [Online]. Available: <http://www.3gpp.org/DynaReport/22.200.htm>
- [7] H.-S. Jo, P. Xia, and J. G. Andrews, "Downlink Femtocell Networks: Open or Closed?" in *Proc. IEEE ICC*, Kyoto, Japan, Jun. 2011, pp. 1–5.
- [8] S. Yun, Y. Yi, D.-H. Cho, and J. Mo, "Open or Close: On the Sharing of Femtocells," in *Proc. IEEE INFOCOM*, Shanghai, China, Apr. 2011, pp. 116–120.
- [9] H.-S. Jo, P. Xia, and J. G. Andrews, "Open, Closed, and Shared Access in the Downlink," *EURASIP J. Wireless Commun. Netw.*, vol. 2012, no. 363, pp. 1–16, Dec. 2012.
- [10] Y. L. Lee, T. C. Chuah, J. Loo, and A. Vinel, "Recent Advances in Radio Resource Management for Heterogeneous LTE/LTE-A Networks," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 2142–2180, Jun. 2014.
- [11] L. B. Le, D. Niyato, E. Hossain, D. I. Kim, and D. T. Hoang, "QoS-Aware and Energy-Efficient Resource Management in OFDMA Femtocells," *IEEE Trans. Wireless Commun.*, vol. 12, no. 1, pp. 180–194, Jan. 2013.
- [12] R. Estrada, A. Jarray, H. Otok, Z. Dziong, and H. Barada, "Energy-Efficient Resource Allocation Model for OFDMA Macrocell/Femtocell Networks," *IEEE Trans. Veh. Technol.*, vol. 63, no. 5, pp. 3492–3437, Sep. 2013.
- [13] R. Balakrishnan and B. Canberk, "Traffic-Aware QoS Provisioning and Admission Control in OFDMA Hybrid Small Cells," *IEEE Trans. Veh. Technol.*, vol. 63, no. 2, pp. 802–810, Feb. 2014.
- [14] A. Hatoum, R. Langar, N. Aitsaadi, R. Boutaba, and G. Pujolle, "Cluster-Based Resource Management in OFDMA Femtocell Networks with QoS Guarantees," *IEEE Trans. Veh. Technol.*, vol. 63, no. 5, pp. 2378–2391, Jun. 2014.
- [15] H. Zhang, C. Jiang, N. C. Beaulieu, X. Chu, X. Wen, and M. Tao, "Resource Allocation in Spectrum-Sharing OFDMA Femtocells with Heterogeneous Services," *IEEE Trans. Commun.*, vol. 62, no. 7, pp. 2366–2377, Jul. 2014.
- [16] H. Zhang, C. Jiang, N. C. Beaulieu, X. Chu, X. Wang, and T. Q. S. Quek, "Resource Allocation for Cognitive Small Cell Networks: A Cooperative Bargaining Game Theoretic Approach," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3481–3493, Jun. 2015.
- [17] H. Zhang, C. Jiang, X. Mao, and H.-H. Chen, "Interference-Limited Resource Optimization in Cognitive Femtocells with Fairness and Imperfect Spectrum Sensing," *IEEE Trans. Veh. Technol.*, vol. PP, no. 99, pp. 1–11, Feb. 2015.
- [18] "Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation," Third Generation Partnership Project (3GPP), Sophia-Antipolis, France, Sep. 2012. [Online]. Available: <http://www.3gpp.org/DynaReport/36.211.htm>
- [19] F. Capozzi, G. Piro, L. A. Grieco, G. Boggia, and P. Camarda, "On Accurate Simulations of LTE Femtocells using an Open Source Simulator," *EURASIP J. Wireless Commun. Netw.*, vol. 2012, no. 328, pp. 1–13, Feb. 2012.
- [20] Y. L. Lee, J. Loo, T. C. Chuah, and A. A. El-Saleh, "Fair Resource Allocation with Interference Mitigation and Resource Reuse for LTE/LTE-A Femtocell Networks," *IEEE Trans. Veh. Technol.*, vol. PP, no. 99, pp. 1–15, Jan. 2016.
- [21] F. Capozzi, G. Piro, L. A. Grieco, G. Boggia, and P. Camarda, "Downlink Packet Scheduling in LTE Cellular Networks: Key Design Issues and a Survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 2, pp. 678–700, May 2013.
- [22] V. Pareto, *Manuale di Economica Politica*. Societa Editrice Libreria, 1906.
- [23] M. Ehrgott, *Multicriteria Optimization*. Springer, 2005.
- [24] D. S. Johnson and M. Garey, *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, 1979.
- [25] I. C. Wong and B. L. Evans, "Optimal Downlink OFDMA Resource Allocation with Linear Complexity to Maximize Ergodic Rates," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 678–700, May 2013.
- [26] G. Femenias, B. Danobeitia, and F. Riera-Palou, "Unified Approach to Cross-Layer Scheduling and Resource Allocation in OFDMA Wireless Networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2012, no. 145, pp. 1–19, Apr. 2012.
- [27] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, R. Vijaykumar, and P. Whiting, "CDMA Data QoS Scheduling on the Forward Link with Variable Channel Conditions," Tech. Memo, Bell Labs, Murray Hill, NJ, USA, Tech. Rep., 2000.
- [28] G. Piro, L. Grieco, G. Boggia, R. Fortuna, and P. Camarda, "Two-Level Downlink Scheduling for Real-Time Multimedia Services in LTE Networks," *IEEE Trans. Multimedia*, vol. 13, no. 5, pp. 1052–1065, Oct. 2011.
- [29] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing Quality of Service over A Shared Wireless Link," *IEEE Commun. Mag.*, vol. 39, no. 2, pp. 150–154, Feb. 2001.
- [30] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [31] G. Piro, L. A. Grieco, G. Boggia, F. Capozzi, and P. Camarda, "Simulating LTE Cellular Systems: An Open-Source Framework," *IEEE Trans. Veh. Technol.*, vol. 60, no. 2, pp. 498–513, Feb. 2011.
- [32] "Simulation Assumptions and Parameters for FDD HeNB RF Requirements," Third Generation Partnership Project (3GPP), Sophia-Antipolis, France, Jul. 2009, Project Document R4-092042.
- [33] W. C. Jakes, *Microwave Mobile Communications*. New York: Wiley, 1975.
- [34] S. Shakkottai and A. L. Stolyar, "Scheduling Algorithms for A Mixture of Real-Time and Non-Real-Time Data in HDR," in *Proc. 17th International Teletraffic Congress*, vol. 4, Sep. 2001, pp. 783–804.
- [35] R. Jain, *The Art of Computer Systems: Performance Analysis*. Hoboken, NJ, USA: John Wiley & Sons, 1991.
- [36] W. Yu and R. Lui, "Dual Methods for Nonconvex Spectrum Optimization of Multicarrier Systems," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1310–1322, Jul. 2006.

Ying Loong Lee received the B.Eng. (Hons.) degree in electronics majoring in telecommunications from Multimedia University, Malaysia, in 2012. He is currently working toward the Ph.D. degree at the same university. His research interests include resource management for heterogeneous cellular networks.

Jonathan Loo received the M.Sc. degree in electronics (with distinction) and the Ph.D. degree in electronics and communications from the University of Hertfordshire, Hertfordshire, U.K., in 1998 and 2003, respectively. Since 2010, he has been a Reader in communication networks with the School of Science and Technology, Middlesex University, London, U.K. His research interests include information centric networking, cloud computing, network architecture, communication protocols, wireless and mobile networks, embedded systems, video coding and transmission, and digital signal processing. He has supervised 16 students to completion and has published over 180 journal and conference papers. Dr. Loo has been an Associate Editor for *Wiley International Journal of Communication Systems* since 2011.

Teong Chee Chuah received the B.Eng degree (first-class honors) in electrical and electronic engineering and the Ph.D. degree in digital communications from Newcastle University, U.K., in 1999 and 2002, respectively. Since 2003, he has been with the Faculty of Engineering, Multimedia University, Malaysia. His research interests include advanced signal processing techniques and error control coding for broadband access communications. Since 2008, he has been serving as a Technical Consultant on digital subscriber line technology to the industry.