Received January 23, 2017, accepted February 6, 2017, date of publication February 23, 2017, date of current version March 28, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2672828

Multi-Targeted Downlink Scheduling for Overload-States in LTE Networks: Proportional Fractional Knapsack Algorithm With Gaussian Weights

NASIM FERDOSIAN[1], (Member, IEEE), MOHAMED OTHMAN[1,2], (Member, IEEE), BORHANUDDIN MOHD ALI[3], (Senior Member, IEEE), AND KWEH YEAH LUN[1]

[1]Department of Communication and Network Technology, Universiti Putra Malaysia, Serdang 43400, Malaysia
[2]Computational Science and Mathematical Physics Lab, Institute of Mathematical Science, Universiti Putra Malaysia, Serdang 43400, Malaysia
[3]Department of Computer and Communication Systems, Universiti Putra Malaysia, Serdang 43400, Malaysia.

Corresponding authors: N. Ferdosian (n.ferdosian@gmail.com) and M. Othman (mothman@upm.edu.my).

This work was supported by the Malaysian Ministry of Education under the Fundamental Research Grant Scheme FRGS/02/10 /12/1143/FR.

ABSTRACT In Long-Term Evolution (LTE) and beyond systems, radio resource scheduling mechanism plays one of the main roles in system performance maximization. From this perspective, due to the conflicting quality requirements of different traffic types, providing a compromise among all performance targets for heterogeneous traffic is difficult. Moreover, the centralized scheduling mechanism for the ever growing number of users along with the massive variety of services, especially in overload states, is infeasible due to the extensive cost of information acquisition and computations. In this paper, we design resource scheduling policies for supporting the efficient delivery of heterogeneous traffic in overload states of a cell. To this end, we cast the class-based bearer-level resource distribution problem as a Proportional Fractional Knapsack model. The objective of the formulated problem is to meet Quality of Service (QoS) requirements and provide fairness for all standardized service classes. Since the solution of this problem is computationally expensive, due to the uncertainty and limited information on network and user operation, we develop a Gaussian-based analytical model and drive a formula for simplified computation of the weight of service bearers. Then, we propose Proportional Fractional Knapsack algorithm for guaranteeing effective utilization of resources for heterogeneous traffic. Finally, performance evaluation results are provided and demonstrate that the proposed scheduling approach can provide a significant level of fairness, in balance with the QoS and throughput performance targets, comparable with optimal ones.

INDEX TERMS Fairness, Gaussian weight, heterogeneous traffic, Proportional Fractional Knapsack, Quality of Service, resource scheduling.

I. INTRODUCTION

To cope with the explosive surge of resource demand by an increasing number of users and mobile traffic (services), nowadays, cellular networks have adopted new channel access technologies to improve system performance. However, Long-Term Evolution (LTE) [1] and LTE-Advanced (LTE-A) cellular networks still suffer from numerous shortcomings [2] such as lack of a standard resource scheduling method to technically support the foreseen massive growth of heterogeneous traffic, associated with distinct characteristics, while accomplishing all performance targets.

On one hand, inasmuch as the Quality of Service (QoS) and data rate requirements associated with different types of traffic are in conflict, providing a compromise among performance targets for heterogeneous traffic is a challenging issue. An efficient scheduling approach should satisfy diverse data rate and QoS requirements for all different new-emerging services. On the other hand, there is a give-and-take among the performance targets, including fairness provisioning, QoS support and system throughput maximization. Typical performance-aware scheduling approaches define a set of users and assign appropriate sub-channels or resource blocks to these users, according to a performance-specific target.

In addition to all these issues, system performance would be severely deteriorated if the channel access and resource scheduling approaches are not properly coordinated for overload states of the network. An efficient scheduling approach, answering all performance targets, depends heavily on a big amount of information such as users' experienced throughput, queues and channel quality information, and service-specific characteristics. Access to all this information would be infeasible for the base stations in overload states of the network and particularly in the next generations of mobile communication networks which contain a huge number of users. This introduces further challenges to the resource scheduling problem in overload states, especially for the next generation of cellular networks, conveying heterogeneous traffic.

Therefore, resource scheduling approaches developed to be used in traditional cellular networks are not efficient when applied in forthcoming cellular networks; consequently, it is imperative to look for the new efficient approaches that are specifically tailored to the emerging networking challenges and provide high performance communication systems for the ever-growing users. In the following, the important state-of-the-art of research on these problems are reviewed.

Performance-aware resource scheduling problems in homogeneous traffic cellular networks, for two types of real-time and non-real-time traffics have extensively been studied [3]. Expected increasing application of real-time services, such as online video streaming, over LTE and beyond cellular networks has attracted a considerable amount of research [4]–[7]. These kinds of applications are mostly resource-hungry and critical in the view of scheduling. A radio resource scheduling approach for multicast video streaming applications, in a single-cell scenario, is presented in [8]. The proposed approach is based on the subgrouping techniques [9] which leverages multiuser diversity by grouping users into different groups based on their Channel State Feedbacks. Multicast video transmission in conjunction with user diversity is also considered in [10]. Therein, it is investigated that the explicit provisioning of users' channel conditions feedback to the eNodeB is infeasible. Chiumento *et al.* [11], analyzed and compared a couple of feedback reduction methods, for a wide range of resource scheduling approaches and different scenarios. It was concluded that a high rate of overall system throughput enhancement can be achieved by using a proper feedback reduction solution.

In LTE-A network, multiple Component Carriers (CCs) aggregation technologies are used to provide larger channel bandwidth for resource-hungry services by aggregating licensed and unlicensed spectrum bands [12]. It is clear that a strong dependency between the resource scheduling approach and users' information reduces the method's applicability; since they require excessive amount of information, depend on the number of users and available system resources, imposing large computational cost in LTE-A networks where each user perceives different levels of channel quality on different CCs [13]. In this regard, heuristic resource scheduling solutions, running in polynomial time, are preferred for practical implementation [14]. These solutions make the scheduling decision based on the objective of a formulated optimization problem, including a single or multiple utility function. The utility function is applied to quantify the maximum offered rewards of various radio resource allocations in terms of dissimilar performance objectives, as addressed in [15]. Most works focus on the throughput maximization objective. For instance, Zhang *et al.* [16] optimized the formulated Linear Programming scheduling problem to improve system throughput and spectral efficiency; however, it has been assumed that each user runs only one type of service at the same time. Two-step resource scheduling algorithm, proposed in [17], provides a trade-off between QoS and throughput maximization by considering the performance concerns from the perspective of both user and service provider, regardless of what application the user is running. As discussed in [18], resource scheduling frameworks which rely on the utility maximization concepts, can be used for general radio resource scheduling problems by formulating them in different ways, according to the performance target of the given system.

In addition to the homogeneous services, radio resource scheduling has also been studied for carrying heterogeneous services. These works consider heterogeneous services in different ways of categorizing [19]. Recently, [20] and [21] considers different types of services into the Guaranteed Bit Rate (GBR) and Non-GBR categories. One drawback of these works is that they have not considered all standardized service classes. More works on providing QoS guarantee for GBR and Non-GBR service classes, with emphasis on overload states, include [22] and [23]. The proposed solutions can distinguish intraclass and interclass-based traffic prioritization. Moreover, their applied utilization function is general, including all QoS influential parameters. In addition to capturing QoS requirements of heterogeneous traffic environment, this function contains adjust weight for each parameter to assist service providers to control the level of service provisioning for users as desired. However, in addition to the quality of service degradation, Mobile bandwidth overload may also cause unfair resource allocation, which is imperative to be studied.

Despite recent advancements in LTE/LTE-A radio resource scheduling approaches, most of them mainly concentrate on especial aspects such as interference mitigation with lack of concern for the other important aspects including fairness and QoS provisioning in overload states (in terms of number of users/amount of traffic), information shortage and limited computational capacities. Here, we propose a downlink scheduling approach for overload states of the network, carrying heterogeneous traffic. In the following, distinctive novelty and contributions of this paper are summarized in brief.

- To maximize system efficiency in terms of LTE performance targets, we mathematically formulate the scheduling problem of assigning resource blocks to the service bearers as a multi-objective optimization problem in terms of throughput maximization, fairness provisioning and QoS service guarantee. For adhering well to the detailed specification of LTE standard [24], the problem formulation differentiates the heterogeneous traffic, according to their standardized classification and takes individual service's constraints and characteristics into account. The standard specific QoS and data rate constraints of each service bearer is required to assure the QoS of different types of traffic.
- We cast the bearer-level class-based downlink scheduling of heterogeneous traffic as a Proportional Fractional Knapsack model. We assume that all performance targets have the same importance. Therefore, it leads us to develop an algorithmic solution, called here Class-based Proportional Solution, which provides all these objectives in balance.
- Since resource allocation is opportunistic in nature and users and their applied services appear in the network randomly, their experienced throughput and channel characteristics are random variables. In this paper, we consider LTE downlink scheduling for heterogeneous traffic in overload states where user's channel information is a random variable. We develop a new analytical model for weighting the service bearers under users' channel characteristics uncertainty. We then derive a formula for success probability in this random environment.
- A downlink resource scheduling algorithm, named Proportional Fractional Knapsack algorithm is proposed in two main phases. Through the first phase, resource-to-class distribution is performed in the way that the fair resource portion for each class of bearers is defined proportionally. Through the second phase the resource-to-bearer allocation is performed by allocating the defined fair resource portion of each class to an optimal set of bearers to meet QoS requirements of the bearers.
- Extensive simulation has been conducted assuming the network contains all service classes, possessing a variety of throughput, delay/latency and loss requirements. While comparing with other LTE overload-state downlink scheduling solutions, we have proved the efficacy of our proposed approach. Unlike many previous works, the proposed resource scheduling approach is distributed, does not require information on users' channel characteristics, and does not depend on the specific model of information acquisition; thus it is highly flexible and offers more applicability in comparison with state-of-the-art approaches.

The rest of the paper is organized as follows. Section 2 presents the network configuration, system model and notation followed throughout the paper. In Section 3, we describe the proposed bearer-level class-based downlink scheduling model and analysis for heterogeneous traffic while Section 4 presents the proposed Proportional Fractional Knapsack algorithm. Simulation setup is described in Section 5 and a performance assessment of our proposals is provided in Section 6. Finally, Section 7 concludes the paper.

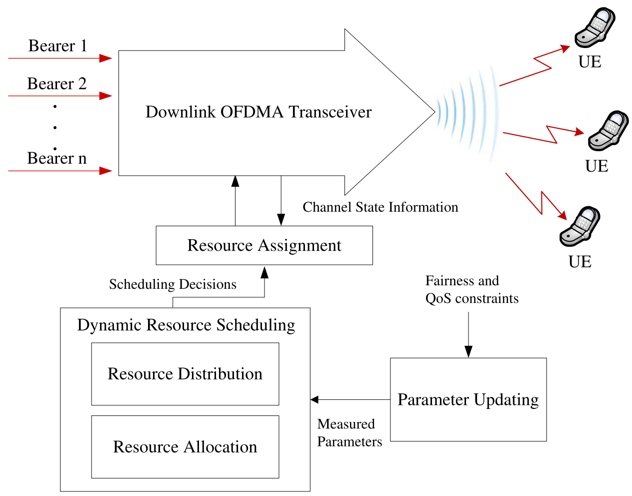

FIGURE 1. **Performance targeted LTE downlink resource allocation.**

II. SYSTEM MODEL AND PROBLEM FORMULATION

Here, we consider a single-cell downlink resource scheduling scenario of the LTE cellular network, where the base station (eNodeB) exploits OFDMA access strategy, to transmit bearers from all different service classes to the units of user equipment as depicted in Fig. 1. The dynamic resource scheduling mechanism, relying on the bearer-level architecture, contains two main phases of resource distribution and resource allocation.

Let $N = \{1, 2, \ldots, \alpha\}$ be set of the bearers waiting to be scheduled by the set of the available system resource blocks $B = \{1, 2, \ldots, \beta\}$ in t^{th} Transmission Time Interval (TTI). Note that we eliminate the time notation t for the simplicity in what follows, unless an ambiguity arises. The 3GPP has classified services and standardized the corresponding characteristics of the service bearers into 9 Quality Class Indicator (QCI) classes in Table 1 [24]. Suppose that $n \in N$ denotes index of a given bearer, waiting for resource scheduling. Each bearer n is assigned to an individual service class with label $c \in C = \{1, 2, \ldots, \gamma\}$ ($\gamma = 9$) and defined as n_c. Table 2 presents the key notations used throughout this paper.

The aim of the eNodeB scheduling algorithm is to allocate resources to these bearers in such a way that optimizes the total system performance. The system performance is determined in terms of three main LTE targets, including throughput maximization, fairness provisioning and QoS service guarantee. Inasmuch as these targets are in conflict with each other, the scheduling algorithm as a flexible solution must

TABLE 1. Standardized QoS characteristics in QCI classes [24].

QCI	Bearer type	Priority	Packet delay budget (ms)	Packet error loss rate	Example services
1	GBR	2	100	10	Conversational Voice
2		4	150	10	Conversational Video (Live Streaming)
3		3	50	10	Real Time Gaming
4		5	300	10	Non-Conversational Video (Buffered Streaming)
5	Non-GBR	1	100	10	Internet protocol Multimedia Subsystem (IMS) Signalling
6		6	300	10	Video (buffered streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
7		7	100	10	Voice, Video (Live Streaming) Interactive Gaming
8		8	300	10	Video (buffered streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p
9		9	300	10	file sharing, progressive video, etc.)

TABLE 2. Notations summary of.

Notation	Definition
β	Total number of available resource blocks
B	Set of available bandwidth resource blocks
B_n	Set of resource blocks allocated to bearer n
N	Set of queued bearers waiting for scheduling
$\bar{r}$	Total system average throughput
$R_{n,rb}(t)$	Throughput of bearer n on rb^{th} resource block at time t
$R_n(t)$	Average throughput of bearer n up to time t
φ_n	Minimum desired fraction of total throughput
$d_{n_c}(t)$	Average delay of bearer n_c up to time t
$\ell_{n_c}(t)$	Average packet loss of bearer n_c up to time t
d_{n_c}	Measured delay of bearer n_c
ℓ_{n_c}	Measured packet loss of bearer n_c
D_c	Standardized delay threshold of QCI class c
L_c	Standardized packet loss threshold of QCI class c
α_c	Number of bearers in QCI class c
ϑ_j	Normalized value of influential metric j
ω_{n_c}	Proportion of bearer n_c selected for scheduling
ρ_{n_c}	Overall weight value of bearer n_c
s_{n_c}	Number of required resource blocks for bearer n_c

provide a compromise among them. Therefore, the scheduling problem of assigning resource blocks to the bearers can be mathematically formulated as a multi-objective optimization problem as follows.

- **Throughput maximization:** In terms of system performance, the most common optimization objective for LTE systems is to distribute resource blocks among different service bearers in such a way that the overall system throughput is maximized, as stated in

$$\forall t, \quad 1 \leq t \leq \Upsilon : \max \sum_{n\in N} \sum_{rb\in B_n} R_{n,rb}(t), \tag{1}$$

where $B_n \subseteq B$ is the set of resource blocks dedicated to the bearer n to transmit data over them and $R_{n,rb}(t)$ is the achieved data rate by bearer n over the rb^{th} resource block at a given time interval t and Υ denotes the number of time intervals.

- **Fairness provisioning:** If an optimal level of the total system performance is provided without any concern for fairness issue, it might prevent a satisfaction level of performance for all users in the network. Therefore, a fairness paradigm is needed to grant each user a long-term and a certain amount of the total system performance, as stated in

$$\forall n \in N : \liminf_{\Upsilon\to\infty} \frac{1}{\Upsilon} \sum_{t=1}^{\Upsilon} R_n(t) \geq \varphi_n \bar{r}, \tag{2}$$

where $R_n(t)$ denotes the overall data rate of bearer n at time t. It is computed by summation of the bearer's data rate over whole set of resource blocks assigned to that bearer at the given time ($R_n(t) = \sum_{rb\in B_n} R_{n,rb}(t)$). φ_n is the minimum fraction of the total system throughput, $\bar{r}$, needed by bearer n, while $\varphi_n \geq 0$ ($\varphi_n = 0$ when bearer n does not have any packet to transmit) and $\sum_{n=1}^{\alpha} \varphi_n \leq 1$.

- **QoS service guarantee:** The third pertinent optimization objective, QoS service guarantee, involves both loss and delay minimization with their respective constraints. It implies that bearers' average loss and delay are intended to be minimized over multiple time intervals defined as

$$\forall n \in N : \min \frac{1}{\Upsilon} \sum_{t=1}^{\Upsilon} \ell_{\alpha_c}(t) \text{ and } \min \frac{1}{\Upsilon} \sum_{t=1}^{\Upsilon} d_{\alpha_c}(t), \tag{3}$$

where $d_{n_c}(t)$ and $\ell_{n_c}(t)$ are measured packet loss and delay of bearer n from QCI class c over time interval t, subject to

$$\forall n \in N, \quad c \in C : d_{n_c} < D_c, \tag{4}$$

$$\forall n \in N, \quad c \in C : \ell_{n_c} < L_c. \tag{5}$$

The constraints in Equations (4) and (5) states the predefined per-class QoS constraints which are needed to be met through the scheduling process. They imply that packet error loss rate ℓ_{n_c} and packet delay budget d_{n_c} experienced by bearer n from QCI class c should be less than predefined loss and delay thresholds, L_c and D_c respectively.

III. MULTI-TARGETED SCHEDULING OF HETEROGENEOUS TRAFFIC

As described in previous section an exact resource distributing solution is required to distribute resource blocks fairly among the bearers, while keeping in check the QoS and data rate requirements for every class of heterogeneous traffic. To address this challenging issue, we formulate the resource distribution problem to the Proportional Fractional Knapsack model and then try to solve it by using a sampling solution through the following sections.

A. PROPORTIONAL FRACTIONAL KNAPSACK MODEL

Class-based resource distribution scenario of heterogeneous traffic can be mapped to a Proportional Fractional Knapsack model with minimal manipulation. Let consider that n bearers are queued at the buffer of the eNodeB, waiting to be resource allocated for transmission. Every bearer n_c has three different characteristics: class index c, size s_{n_c} and weight value ρ_{n_c}. Suppose that there are α_c bearers in each class c, while summation of the bearers in all classes is equal to the total number of bearers waiting for scheduling (i.e. $\sum_{c=1}^{\gamma} \alpha_c = \alpha$). Then size of each class, S_c, and the total size of the whole queued bearers, S, are respectively defined as

$$S_c = \sum_{n=1}^{\alpha_c} s_{n_c} \tag{6}$$

and

$$S = \sum_{c=1}^{\gamma} \sum_{n=1}^{\alpha_c} s_{n_c}. \tag{7}$$

The desired solution must fill the knapsack with a capacity equal to the number of available system resource blocks, β, by the queued bearers, while satisfying the following specifications.

1) The total profit of the chosen bearers should be maximum.
2) The proportion of the number of selected bearers from class c should be close to S_c/S.
3) The total size of the selected bearers should be less than or equal to β.

Then according to these specifications, the corresponding Proportional Fractional Knapsack model of the above-mentioned resource distribution problem can be formulated, in order, through the Equations 8, 9 and 10 in the following definition.

Definition 1: The Proportional Fractional Knapsack model of the fair resource distribution is defined by a finite set of bearers N, a nonnegative real number B and two nonnegative vectors $\rho \in \mathbb{R}^{\mathrm{N}}$ and $s \in \mathbb{R}^{\mathrm{N}}$ verifying

$$\max \sum_{c=1}^{\gamma} \sum_{n=1}^{\alpha_c} \omega_{n_c} \rho_{n_c}, \tag{8}$$

$$\min \left\{ \sum_{c=1}^{\gamma} \left(\frac{\sum_{n=1}^{\alpha_c} \omega_{n_c} s_{n_c}}{\sum_{c=1}^{\gamma} \sum_{n=1}^{\alpha_c} \omega_{n_c} s_{n_c}} - \frac{\sum_{n=1}^{\alpha_c} s_{n_c}}{\sum_{c=1}^{\gamma} \sum_{n=1}^{\alpha_c} s_{n_c}} \right) \right\}, \tag{9}$$

and

$$\forall n \in N, \quad 0 \leq \omega_{n_c} \leq 1 : \sum_{c=1}^{\gamma} \sum_{n=1}^{\alpha_c} \omega_{n_c} s_{n_c} \leq B, \tag{10}$$

where ω_{n_c} is the proportion of bearer n from class c, selected for scheduling. □

Number of the resource blocks, which are required for transferring data packets of the bearer n, defines the bearer size s_{n_c}. Further, the weight value of a given bearer which indicates the maximum offered reward of that bearer for scheduling, is computed by a normalized weighting function, compromising the QoS and throughput influential factors.

Theorem 1: The Proportional Fractional Knapsack problem of resource distribution, defined in Definition 1, is an NP-hard problem.

Proof: A proof method for the NP hardness of the $0-1$ Proportional Knapsack problem has been provided in [25]. Since Definition 1 problem is the fractional version of $0-1$ Proportional Knapsack, with the only difference that individual packets/bytes from the bearer queue can be chosen without any force to choose all or nothing from each bearer, therefore Proportional Fractional Knapsack problem of resource distribution defined in Definition 1, is an NP-hard problem too. □

Since the Proportional Fractional Knapsack problem of resource distribution is NP-hard, tackling it by using a heuristic procedure is a normal way. If we come up with a subset of the highest weighted bearers such that, for each class the proportion of the number of that class's bearer selected for scheduling, and the proportion of the number of that class's bearers waiting for scheduling is almost the same, and the total size of the selected bearers not overloading the available resources, then the problem is solved. Accordingly, we target to find such a solution.

We assume that all three mentioned objectives have the same importance. Therefore, it leads us to a solution that provides all these objectives in balance. Let set $H = \{\omega_1, \omega_2, \ldots, \omega_\alpha\}$ indicates any feasible solution, here called sample. First, we determine the size of each class inside the sample by considering two later objectives (9) and (10); then, find the optimal set of the bearers from each class, fitting the predefined size of that class and satisfying the first objective (8) in order to ensure that the optimal solution will be found.

B. OBJECTIVE SIMPLIFICATION BY Gaussian WEIGHTS

In this section, we first show that the respective problem to the objective of performance maximization stated in Equation (8) is a complicated stochastic version of the Knapsack model and then we try to overcome this problem and decrease the computational complexity, by simplifying the objective function using the summer property of the weight value random variable.

The main uncertainty in this stochastic optimization problem is the weight value of each bearer which is distributed normally. Let consider the weight value of the bearers, at any time interval t, is denoted by a random vector $\rho(t) = (\rho_n(t) : n \in N)$. Therefore, the respective objective function in Equation (8) with stochastic random weight values for every time interval t can be written as

$$\max \sum_{c=1}^{\gamma} \sum_{n=1}^{\alpha_c} \omega_{n_c} \rho_{n_c}(t), \qquad \rho_{n_c} \geq 0. \tag{11}$$

Since $\rho(t)$ are random values, we operate with expected random values rather than the actual values. With this perspective in mind, the expected value of $\rho(t)$ is $\mathbb{E}[\rho(t)] = \int \rho f(\rho) d_\rho$, where $\mathbb{E}$ denotes the expected value operator and f is the density function of the random vector ρ. The main difficulty in every TTI is the number of scenarios to consider for evaluating the objective function, containing many variables and computations.

As already mentioned, the overall weight ρ_n for a given bearer n can be calculated by using an aggregated normalized weighting function which is a combination of the normalized value ϑ_i of influential parameter $i \in \{delay, loss, queuedepth, priority, throughput\}$, and can be defined as [8]

$$\rho_n(t) = \sum_i p_i . \tanh(\vartheta_i). \tag{12}$$

In case of four former influential parameters, the respective normalized values are computed by using the user itself information and here we consider them as a known value $A \geq 0$. However in case of throughput parameter, the normalized value is computed by utilizing the feed-back information including CQI value and the past average throughput. Let model the normalized value of bearers' throughput by a random vector $R(t) = (R_n(t) : n \in N)$, which can take value in a finite set $R(t) \subset [0, +\infty)$ and for all $t \neq t'$, $R_n(t)$ and $R_n(t')$ are independent and have the same distribution. Here, we consider the generalized form of the problem for ease of exposition. Then by separating the known value A from the stochastic random variable $R(t)$, the aggregated normalized weight function (12) can be written as

$$\rho_n(t) = A_n + R_n(t). \tag{13}$$

Then by replacing $\rho_n(t)$ and discarding index k for simplicity, the consequent objective function (11) is redefined as

$$\begin{aligned} &\max \sum_{n=1}^{\alpha} \Big(\omega_n A_n + \omega_n R_n(t) \Big) \\ &= \max \Big(\sum_{n=1}^{\alpha} \omega_n A_n + \sum_{n=1}^{a} \omega_n R_n(t) \Big) \\ &= \max \sum_{n=1}^{\alpha} \omega_n A_n + \max \sum_{n=1}^{a} \omega_n R_n(t), \end{aligned} \tag{14}$$

where $\mathbb{E}[R_n] = \int R_n f(R_n) d_{R_n}$ and f is the density function of the random variable R_n. If we consider $\Omega(\omega)$ as the resource function which computes the expected optimal value of the second stage of the objective function (14), and R_n as the only resource variable with the coefficient equal to 1, then $\Omega(\omega)$ can be defined by

$$\begin{aligned} \Omega(\omega) &= \mathbb{E}\left[\max_{0 \leq \omega_n \leq 1} \sum_{n=1}^{\alpha} \omega_n R_n \right] \\ &= \int \sum_{n=1}^{\alpha} \omega_n R_n \max f(R_1, R_2, \ldots, R_n) d_{R_1} d_{R_2} \ldots d_{R_n}. \end{aligned} \tag{15}$$

In each iteration, the solution technique needs to evaluate many times the resource function which is an integral of n variables. It results in big number of scenarios to consider for evaluating the objective function, containing many variables and computations.

We assume that the sample size of each class, computed in previous section, is sufficiently big so that the central limit theorem can be employed and $R(t)$ of each bearer is according to a known distribution and independent of other bearers. The central limit theorem indicates that big number of independent observations from any distribution tends to be a normal distribution. Therefore, It follows from our assumption and the central limit theorem, that $R_n(t)$s are normally distributed. Consequently according to summation property, if $R_1, R_2, \ldots, R_n$ are independent Gaussian random variables such that $R_n \sim N(\mu_n, \sigma_n^2)$ with mean μ_n and variance σ_n, then $\Omega(\omega) := \sum \omega_n R_n(t)$ follows a Gaussian distribution with mean $\mu(\omega) = \sum \omega_n \mu_n$ and $\sigma_n^2(\omega) = \sum \omega_n^2 (\sigma_n)^2$. Then it is easy to show that since $\Omega(\omega) \sim N(\mu(\omega), \sigma(\omega)^2)$ then we can express Ω as follows [26].

$$\Omega(\omega) = \mu(\omega) \Psi\left(\frac{\mu(\omega)}{\sigma(\omega)}\right) + \frac{\sigma(\omega)}{\sqrt{2\pi}} \exp\left(\frac{-\mu(\omega)^2}{2\sigma(\omega)^2}\right), \tag{16}$$

where Ψ denotes the standard normal cumulative distribution function. Therefore, the optimization problem of (14) with Gaussian weight values can be formulated by the $\Omega(\omega)$ as

$$\max \sum_{n=1}^{\alpha} \omega_n A_n + \left(\mu(\omega) \Psi\left(\frac{\mu(\omega)}{\sigma(\omega)}\right) + \frac{\sigma(\omega)}{\sqrt{2\pi}} \exp\left(\frac{-\mu(\omega)^2}{2\sigma(\omega)^2}\right) \right), \tag{17}$$

which is a concave function of binary variables, therefore it can be computed quickly.

IV. CLASS-BASED BEARER-LEVEL DOWNLINK SCHEDULING

A. PROPORTIONAL DISTRIBUTION SOLUTION

Recall that α_c is the number of bearers in class c which are selected for solution and ω_{n_c} $(0 \leq \omega_{n_c} \leq 1)$ is the proportion of the selected bearer n_c. Therefore, the sample size for each class is denoted by U_c, and can be demonstrated in the following expression as

$$U_c = \sum_{n=1}^{\alpha_c} \omega_{n_c} s_{n_c}, \tag{18}$$

then the total sample size U, which is summation of the size of all γ classes, is defined as

$$U = \sum_{c=1}^{\gamma} U_c. \tag{19}$$

According to objectives (9) and (10), it is desired that

$$U_c/U = S_c/S, \quad \text{for } c = 1, \ldots, \gamma, \tag{20}$$

and

$$\sum_{c=1}^{\gamma} \sum_{n=1}^{\alpha_c} \omega_{n_c} s_{n_c} = \beta. \tag{21}$$

Then, by replacing U with β in Equation (20) we can calculate the sample size U_c, for each class $c = 1, \ldots, \gamma$, as

$$U_c = \lfloor \beta/S \rfloor S_c, \quad \text{for all } c = 1, \ldots, \gamma \tag{22}$$

where $\lfloor \beta/S \rfloor$ denotes the closet integer value to β/S. After calculating the U_c, which is the number of bearers from class k in the desired sample, we continue with the next step to choose the optimal set of the bearers from each service class, according to the first predefined objective, fitting the U_c size.

B. PROPORTIONAL-FRACTIONAL-KNAPSACK (PFK) ALGORITHM

This section introduces a resource scheduling algorithm in two main steps. The detailed procedure is shown in the PFK Algorithm 1. This algorithm initializes the bearers' size s_{n_c}, QoS (ℓ_{n_c}, d_{n_c}) and throughput parameters $R_n(t)$ and updates their values in each iteration. The actual updates of the bearers' parameters in the system is done to take into account the effects of resource allocation, transmission rates, delays and packet drops in the next scheduling iterations.

In the first step through lines 27, fairness-aware class-based resource distribution is performed in the way that a fair resource share for each class of bearers is defined, proportional to their aggregated generated traffic. After classifying the active bearers into standardized QCI classes, the accumulated resource demand by each class is calculated by summation of the bearers' demands in that class, yielding the class-based traffic distribution vector $S = \{S_1, S_2, \ldots, S_9\}$. Then, by applying the proposed proportional resource distribution solution, the fair share for each QCI class is computed.

Further to support the QoS requirements of the bearers in LTE downlink multi-service system, in the second phase, the classes' resource share defined in the first phase are allocated to the bearers of each class based on the bearers' QoS requirements. All the system active bearers are sorted according to their weight value in descending order. The weight value of each bearer is computed by using the normalized weighting function.

Then in each TTI, the bearer with the highest weight value is selected to be resource allocated, if there is any unallocated resources in the corresponding QCI class. The maximum amount of bits from each bearer, which are able to be transmitted using the available resource blocks, is computed.

Algorithm 1 PFK Algorithm

Data: Set of available system resources β, a queue of n bearers waiting for scheduling along with their size s_{n_c}, measured QoS ℓ_{n_c}, d_{n_c} and throughput parameters $R_n(t)$.
Result: Allocating resources to the queued bearers

1 **begin**
2 Classify bearers into QCI classes;
3 **for** *each QCI class c* **do**
4 Compute the accumulated resource demand of each QCI class c, using $S_c = \sum_{i=1}^{\alpha_c} s_i$;
5 Compute the proportional share (U_c) for each QCI class c, using $U_c = \left\lfloor \frac{\beta}{\sum_{j=1}^{\gamma} S_j} \right\rfloor S_c$;
6 **end**
7 calculate weight values of the queued bearers using Equation (17);
8 sort bearers according to their ratio (ρ_{n_c}/s_{n_c}) in decreasing order and index them from 1 to α;
9 **While** all resource quotas are allocated OR no bearers remained unallocated
10 **for** *each queued bearer n indexed from* 1 *to* α **do**
11 **while** $U_c > 0$ ***and*** $n \leq \alpha$ **do**
12 **if** $(U_c - s_{n_c}) \geq 0$ **then**
13 $U_c - = s_{n_c}$;
14 $n + = 1$;
15 **else**
16 $s_{n_c} - = U_c$;
17 $U_c = 0$;
18 **end**
19 **end**
20 **end**
21 **end**

After allocating the resources and transmitting the bearers' packets if there is still any packet whose deadline was over and not transmitted yet (due to lack of resource blocks), it is dropped from the bearer's queue. Then, based on the arrived, transmitted, and dropped number of packets, the sizes, losses and delays of the remaining packets are measured and updated. This loop continues till all the bearers are resource allocated or no resource block is remained unallocated.

V. SIMULATION ENVIRONMENT AND PARAMETERS

We consider an LTE cell under sequences of the overload traffic, where the users are weighted for scheduling by applying the Gaussian weights and with the Knapsack policy being the resource allocation algorithm. The goal of the scheduling system is to efficiently distribute and allocate resources to the users which gains the highest performance rewards in term of QoS and fairness provisioning. Performance of the proposed Gaussian-Knapsack scheduling approach is evaluated in comparison with the performance of Knapsack,

TABLE 3. Simulation parameters.

Parameter	Value
System channel bandwidth	20 MHz
Number of Resource Blocks (RBs)	100 RBs per spectrum allocation (12 subcarrier per RB)
Subchannel bandwidth	180 KHz
Frame structure	TDD
Time slot duration	1 ms
Simulation length	32 min
Scheduling time window	10 ms
Number of data bearers	400
Hurst parameter for data traffic	0.9
Number of voice bearers	1200
Voice Activity Factor for voice generator	0.5
Mean talk spurt duration	5 s
Voice codec encoding frame size	20 ms
Subcarrier spacing	15 kHz
Modulation and coding scheme	QPSK, 16QAM and 64QAM
Normal, overload run intervals in sequence	(1,5,3,2,1,2,1,10,3,5 and 3)

Greedy-Knapsack and Priority-only algorithms as the reference approaches. Priority-Only algorithm serves the queued bearers waiting for scheduling, according to the priority queuing approach, so that the bearer with higher QoS priority is served preferentially. The same simulation platform and parameters, as stated in Table 3, are applied for all scheduling strategies and performances are evaluated in different scenarios. This simulation environment was implemented based on the LTE cellular network characteristics, defined in 3GPP verification framework, while comprehending scheduling aspects of eNodeB. The assumed 20MHz bandwidth consists of 100 resource blocks per spectrum allocation in time-frequency domain. The voice and data traffic were modeled by means of the exponential distribution function and aggregate self-similar pattern respectively.

Heterogeneous traffic scenarios are imposed, where there is a mixture of different traffic types including all QCI classes except QCI class 5, which has the highest Priority and independent from the access network. The simulation analysis considers transmission of multi-class bearers through the eNodeB toward the users in its cell. To be realistic models, the implemented scenarios are composed of a stochastic distribution of normal and overload time-intervals. During the normal period, there is a set of 330 active users in a cell with different kinds and number of service bearers. For implementing the overload time interval, 50 users with single data bearer are added to the existed traffic and they are eliminated at the beginning of the normal state. The implemented overload time intervals model scenarios, imposing extreme limiting factors, such as a higher overhead for users' channel quality reporting, more computations for resource scheduling and higher overhead for announcing the scheduling decisions.

VI. PERFORMANCE RESULTS AND DISCUSSION

To investigate the effectiveness of the proposed scheduling algorithm, we compare it with other scheduling approaches tailored for the overload-state downlink resource allocation of LTE networks. These approaches have been adapted to our considered heterogeneous environment, covering all different classes of services. Then, we provide per-class results, representing the performance evaluation of the proposed multi-targeted scheduling approach in terms of fairness, QoS and throughput. Inasmuch as, it is outlined in [22], no typical traditional algorithm is optimal for all application classes during the heavy load states of the LTE network; therefore, a comparison of the proposed algorithm with the traditional fairness- and QoS-ware approaches is not fine in this context.

A. PER-CLASS FAIRNESS

The first simulation analysis evaluates the fairness provisioning of the PFK algorithm for heterogeneous services. Inasmuch as each kind of service has dissimilar data rate requirements, the level of fairness, provided by each overload-state scheduling algorithm, is evaluated per service class. In this regard, we measure fairness for each QCI class separately by computing the Cumulative Distribution Function of the per-class Jain's fairness index in two categories of GBR and Non-GBR classes. As shown in Figures 2 and 3 respectively, PFK algorithm provides a better level of fairness for all QCI classes, in compare with the reference scheduling algorithms, while the most significant improvement belongs to the GBR classes.

For QCI class 1 in Figure 2(a), the fairness index value obtained by PFK is higher than 0.89, which is the biggest range obtained by the algorithms, indicating that PFK algorithm provides an optimal Quality of Experience (QoE) for VoIP bearers, even when there is an excessive load on the system.

In case of QCI class 2 in Figure 2(b), 85 % of the bearers received a fairness of 0.73 or less when PFK was used, while they received a fairness of 0.63 or less when the Priority-Only and Knapsack algorithms were used. As Figures 2(c) and 2(d) show, the PFK algorithm obtained the best fairness value for QCI class 3 and compared to the other algorithms, which all produced a smaller amount of fairness especially for the higher percent of the bearers.

In contrast, for the Non-GBR QCI classes, as can be seen in Figures 3(a)-3(d), the difference between the PFK and alternative scheduling algorithms is much more. For instance, 85 % of the bearers from QCI class 6 experience a fairness of up to 0.67 by using the PFK, while for the Knapsack and Priority-Only algorithms they experience fairnesses up to 0.30. The main reason for this big improvement is because of the distribution of resources among different services, proportional to their traffic distribution over the network. Since, the Non-GBR bearers, including TCP-based services, occupy the biggest volume of the current systems' traffic, the highest fairness level by the PFK algorithm has been provided for them by allocating the fair portion of resources proportional to their size.

Overall, we can see that PFK outperforms the Priority-Only algorithm for almost all QCI classes, with the highest amount of improvements in contrast to the Knapsack and Greedy-Knapsack algorithms. For the Knapsack and Greedy-Knapsack algorithms, the fairness results for most of the QCI classes are very close. For the most part, the Greedy-Knapsack algorithm benefits from the

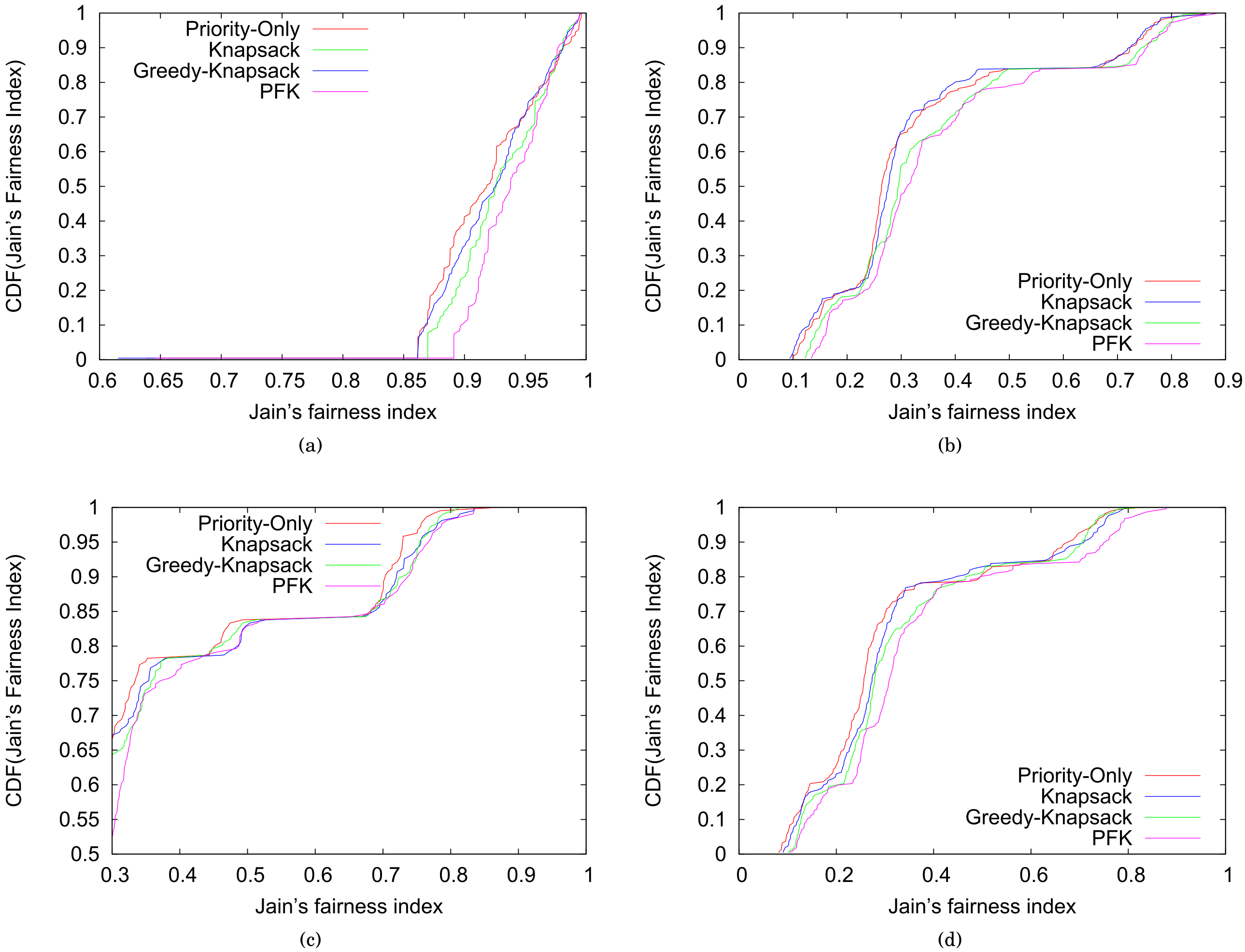

FIGURE 2. **Per-class fairness for bearers from GBR classes: (a) QCI class 1, (b) QCI class 2, (c) QCI class 3, and (d) QCI class 4.**

throughput-aware ranking function and produces better results than the Knapsack algorithm, especially for QCI classes 6-9.

The PFK improvements essentially benefit from the class-based proportional technique applied to resource distribution, where the QCI classes, with dissimilar characteristics and different amounts of resource demand, negotiate, and cooperate with each other to make a fair decision. In contrast, in the reference algorithms, the resource competition among different service classes is accomplished in a selfish manner, where each class's performance desire is in conflict with that of the others. Therefore, by combining the results of Figures 2 and 3, it can be concluded that the division rule applied in the PFK algorithm obtains an outcome with fair service and ubiquitous coverage. Consequently, the throughput of the system is not biased in favor of the specific kinds of services.

B. PER-CLASS QoS AND THROUGHPUT

The QoS performance results are presented in terms of average packet loss and delay in addition to throughput for each individual QCI class. Tables 4 to 11 show the percentage of improvement obtained by the PFK algorithm with respect to the reference algorithms. The average throughput, packet loss, and delay for GBR classes 1-4, shown in Tables 4 to 7, indicate that the proposed algorithm performs well enough to ensure that GBR QCI classes 1-4 meet their QoS constraints in terms of loss and delay, leading to a strong QoE for GBR traffic.

As can be seen in these tables, the PFK algorithm obtains improvements for GBR application classes, especially in compare with the Knapsack and Priority-Only algorithm. Moreover, the VoIP bearers, which correlate to QCI class 1, are scheduled with no loss and almost no delay. The conversational video traffic from QCI class 2 and the rest of the GBR bearers from QCI classes 3 and 4 also experience near zero loss. In addition, the Non-GBR traffic for QCI classes 6 to 9 experience reasonable levels of loss and delay, although they are impacted by the overload.

With respect to throughput performance, the PFK algorithm achieves a less notable throughput improvement with respect to the Greedy-Knapsack scheduling algorithm

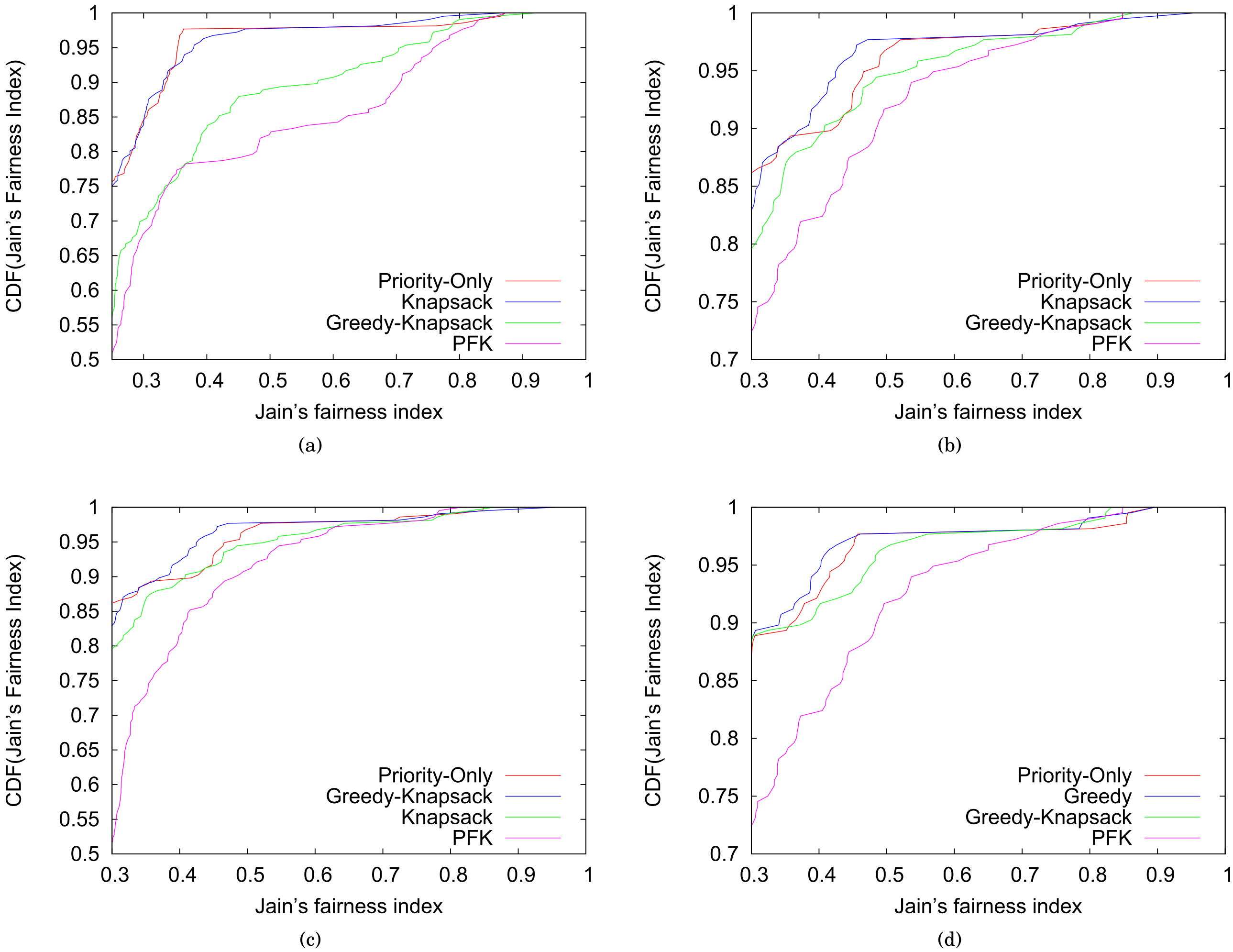

FIGURE 3. Per-class fairness for bearers from Non-GBR classes: (a) QCI class 6, (b) QCI class 7, (c) QCI class 8, and (d) QCI class 9.

TABLE 4. QoS and throughput evaluation, QCI1.

Scheduler	Avg. throughput	Avg. latency (ms)	Avg. loss (Mbps)
PFK	21.19	4.5	0.00
Greedy-Knapsack	21.23	4.5	0.00
Improved percentage(%)	-0.2	0.0	0.0
Knapsack	21.14	4.5	0.00
Improved percentage(%)	0.2	0.0	0.0
Priority-Only	21.15	4.5	0.00
Improved percentage(%)	0.2	0.0	0.0

TABLE 5. QoS and throughput evaluation, QCI2.

Scheduler	Avg. throughput	Avg. latency (ms)	Avg. loss (Mbps)
PFK	11.97	68.3	0.15
Greedy-Knapsack	12.09	65.0	0.09
Improved percentage(%)	-1.0	-5.1	-66.7
Knapsack	11.84	68.5	0.17
Improved percentage(%)	1.1	0.3	11.8
Priority-Only	10.99	95.1	0.24
Improved percentage(%)	8.9	28.2	37.5

TABLE 6. QoS and throughput evaluation, QCI3.

Scheduler	Avg. throughput	Avg. latency (ms)	Avg. loss (Mbps)
PFK	18.57	58.8	0.09
Greedy-Knapsack	18.62	52.1	0.06
Improved percentage(%)	-0.3	-12.8	-50.0
Knapsack	18.96	63.4	0.10
Improved percentage(%)	-2.0	7.2	10.0
Priority-Only	17.90	70.7	0.34
Improved percentage(%)	3.7	16.8	73.5

TABLE 7. QoS and throughput evaluation, QCI4.

Scheduler	Avg. throughput	Avg. latency (ms)	Avg. loss (Mbps)
PFK	10.32	69.3	0.17
Greedy-Knapsack	10.67	68.3	0.14
Improved percentage(%)	-3.3	-1.5	-21.4
Knapsack	10.16	70.9	0.19
Improved percentage(%)	1.6	2.2	10.5
Priority-Only	10.31	70.6	0.21
Improved percentage(%)	0.1	1.8	19.0

because both of them use the same throughput-aware policy, which is applied in an aggregate ranking function. Moreover, the better improvement in throughput acquired by PFK with respect to the Priority-Only and Knapsack algorithms indicates that these algorithms are not throughput-optimal. The GBR service bearers, especially those from QCI classes 4 and 9, maintained a high level of QoE in terms of throughput in the PFK scheduling algorithm. However, some tradeoffs were also perceived. For instance, QCI classes 6, 7 and 8 were allocated less throughput, but the service traffic did not

TABLE 8. QoS and throughput evaluation, QCI6.

Scheduler	Avg. throughput	Avg. latency (ms)	Avg. loss (Mbps)
PFK	5.77	2,052.6	23.63
Greedy-Knapsack	5.97	1,386.4	20.45
Improved percentage(%)	-3.3	-48.0	-15.5
Knapsack	5.83	1,423.1	18.69
Improved percentage(%)	-1.0	-44.2	-26.4
Priority-Only	7.59	20.94	0.33
Improved percentage(%)	-24.0	-9702.3	-7060.6

TABLE 9. QoS and throughput evaluation, QCI7.

Scheduler	Avg. throughput	Avg. latency (ms)	Avg. loss (Mbps)
PFK	3.98	705.2	5.42
Greedy-Knapsack	5.02	612.6	4.02
Improved percentage(%)	-20.7	-15.1	-34.8
Knapsack	4.74	550.3	3.41
Improved percentage(%)	-16.0	-28.1	-58.9
Priority-Only	5.12	183.5	1.51
Improved percentage(%)	-22.3	-284.3	-258.9

TABLE 10. QoS and throughput evaluation, QCI8.

Scheduler	Avg. throughput	Avg. latency (ms)	Avg. loss (Mbps)
PFK	6.57	1,959.5	35.21
Greedy-Knapsack	6.64	1,861.9	34.66
Improved percentage(%)	-1.0	-5.2	-1.6
Knapsack	6.62	1,887.1	33.49
Improved percentage(%)	-0.7	-3.8	-5.1
Priority-Only	6.65	1,578.1	32.61
Improved percentage(%)	-1.2	-24.2	-8.0

TABLE 11. QoS and throughput evaluation, QCI9.

Scheduler	Avg. throughput	Avg. latency (ms)	Avg. loss (Mbps)
PFK	5.80	2,017.5	36.15
Greedy-Knapsack	5.55	2,582.0	42.89
Improved percentage(%)	4.5	21.9	15.7
Knapsack	5.71	2,537.0	39.65
Improved percentage(%)	1.6	20.5	8.8
Priority-Only	4.23	4,574.8	58.51
Improved percentage(%)	37.1	55.9	38.2

starve, even under overload states in the network. Everyday TCP traffic, which is assigned to QCI class 9, experienced a significant improvement in throughput compared to the results of the Knapsack scheme.

The PFK's strong bias toward providing fairness for all classes of QCI excessively compromised the experienced throughput and QoS over the region of interest (GBR classes) and lower priority services (Non-GBR classes). Consequently, results of the PFK algorithm indicate that this scheduling strategy is the most effective for fine-tuning performance targets across the various service classes.

VII. CONCLUSION

The presence of a huge number of mobile users, running heterogeneous traffic, increases the radio resource management criticalities, due to the high dissimilarity of users' channel conditions and heterogeneity of data rate and QoS requirements of the different service classes. This paper addressed LTE resource scheduling of heterogeneous traffic in overload states of the networks where a huge number of traffic bearers need to be served by the limited available radio resources. An efficient resource scheduler in base station, answering all system performance targets, needs enough information on users' operations and specified requirements such as user's throughput gains and service constraints. Access to all this information would be infeasible for base stations, particularly in emerging hierarchical networks containing a huge number of users. To cope with the difficulty in information acquisition and high computations, we proposed a Gaussian-based bearer weighting method which simplifies the performance maximization objective. The Proportional Fractional Knapsack scheduling algorithm was proposed, in which the highest valued bearers from different service classes are scheduled for resource allocation proportional to their distribution in a fair manner. The simulation results demonstrate that fair resource allocation in the downlink scheduling scheme among the QCI classes with the same number of bearers is achieved along with the QoS provisioning in terms of system loss and delay. Future works will address the analytical validation of the proposed scheduling approach and its investigation through multi-operator shared systems.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Michael Brehm and Prof. Dr. Ravi Prakash for their valuable contributions from the University of Texas at Dallas.

REFERENCES

[1] "Evolved universal terrestrial radio access (E-UTRA) and evolved universal terrestrial radio access network (EUTRAN), Rel. 11," 3GPP, Sophia Antipolis, France, Tech. Rep., Sep. 2012.

[2] T. O. Olwal, K. Djouani, and A. M. Kurien, "A survey of resource management toward 5G radio access networks," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1656–1686, 3rd Quart., 2016.

[3] 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, 4th Quart., 2014.

[4] G. Ku and J. M. Walsh, "Resource allocation and link adaptation in LTE and LTE advanced: A tutorial," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 3, pp. 1605–1633, 3rd Quart., 2015.

[5] T.-C. Liu, K. Wang, C.-Y. Ku, and Y.-H. Hsu, "QoS-aware resource management for multimedia traffic report systems over LTE-A," *Comput. Netw.*, vol. 94, pp. 375–389, Jan. 2016.

[6] M. Carlesso, A. Antonopoulos, F. Granelli, and C. Verikoukis, "Uplink scheduling for smart metering and real-time traffic coexistence in LTE networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 820–825.

[7] N. Zorba and C. Verikoukis, "Energy optimization for bidirectional multimedia communication in unsynchronized TDD systems," *IEEE Syst. J.*, vol. 10, no. 2, pp. 797–804, Jun. 2016.

[8] M. Condoluci, G. Araniti, A. Molinaro, and A. Iera, "Multicast resource allocation enhanced by channel state feedbacks for multiple scalable video coding streams in LTE networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 2907–2921, May 2016.

[9] R. O. Afolabi, A. Dadlani, and K. Kim, "Multicast scheduling and resource allocation algorithms for OFDMA-based systems: A survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 240–254, 1st Quart., 2013.

[10] A. Tassi, C. Khirallah, D. Vukobratović, F. Chiti, J. S. Thompson, and R. Fantacci, "Resource allocation strategies for network-coded video broadcasting services over LTE-advanced," *IEEE Trans. Veh. Technol.*, vol. 64, no. 5, pp. 2186–2192, May 2015.

[11] A. Chiumento, C. Desset, S. Polling, L. Van der Perre, and R. Lauwereins, "The value of feedback for LTE resource allocation," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2014, pp. 2073–2078.

[12] R. Zhang, Z. Zheng, M. Wang, X. Shen, and L.-L. Xie, "Equivalent capacity in carrier aggregation-based LTE-A systems: A probabilistic analysis," *IEEE Trans. Wireless Commun.*, vol. 13, no. 11, pp. 6444–6460, Nov. 2014.

[13] K. I. Pedersen, F. Frederiksen, C. Rosa, H. Nguyen, L. G. U. Garcia, and Y. Wang, "Carrier aggregation for LTE-advanced: Functionality and performance aspects," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 89–95, Jun. 2011.

[14] M. E. Aydin, R. Kwan, and J. Wu, "Multiuser scheduling on the LTE downlink with meta-heuristic approaches," *Phys. Commun.*, vol. 9, pp. 257–265, Dec. 2013.
[15] T. Taleb, Y. Hadjadj-Aoul, and K. Samdanis, "Efficient solutions for enhancing data traffic management in 3GPP networks," *IEEE Syst. J.*, vol. 9, no. 2, pp. 519–528, Jun. 2016.
[16] Q. Zhang, Q. Wang, Z. Feng, and T. Yang, "Design and performance analysis of a fairness-based license-assisted access and resource scheduling scheme," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 11, pp. 2968–2980, Nov. 2016.
[17] T. Hori and T. Ohtsuki, "QoE and throughput aware radio resource allocation algorithm in LTE network with users using different applications," in *Proc. IEEE 27th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Sep. 2016, pp. 1–6.
[18] F. Shams, G. Bacci, and M. Luise, "A survey on resource allocation techniques in OFDM(A) networks," *Comput. Netw.*, vol. 65, pp. 129–150, Jun. 2014.
[19] S. E. Middleton and S. Modafferi, "Scalable classification of QoS for real-time interactive applications from IP traffic measurements," *Comput. Netw.*, vol. 107, no. 1, pp. 121–132, Oct. 2016.
[20] P. Ameigeiras *et al.*, "3GPP QoS-based scheduling framework for LTE," *EURASIP J. Wireless Commun. Netw.*, vol. 78, pp. 1–14, Dec. 2016.
[21] P. Oliver-Balsalobre, M. Toril, S. Luna-Ramírez, and J. M. R. Avilés, "Self-tuning of scheduling parameters for balancing the quality of experience among services in LTE," *EURASIP J. Wireless Commun. Netw.*, vol. 7, pp. 1–12, Jan. 2016.
[22] M. Brehm and R. Prakash, "Overload-state downlink resource allocation in LTE MAC layer," *Wireless Netw.*, vol. 19, no. 5, pp. 913–931, Jul. 2013.
[23] N. Ferdosian, M. Othman, B. M. Ali, and K. Y. Lun, "Greedy–knapsack algorithm for optimal downlink resource allocation in LTE networks," *Wireless Netw.*, vol. 22, no. 5, pp. 1427–1440, Jul. 2016.
[24] *3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Study on RAN Improvements for Machine-type Communications; (Release 11), V11.0.0*, Document TR 37.868, 3GPP, sep. 2011.
[25] M. Penn, D. Hasson, and M. Avriel, "Solving the 0–1 proportional knapsack problem by sampling," *J. Optim. Theory Appl.*, vol. 80, no. 2, pp. 289–297, Feb. 1994.
[26] B. Fortz, M. Labbé, F. Louveaux, and M. Poss, "Stochastic binary problems with simple penalties for capacity constraints violations," *Math. Program.*, vol. 138, no. 1, pp. 199–221, Apr. 2013.
[27] A. M. Cohn and C. Barnhart, "The stochastic knapsack problem with random weights: A heuristic approach to robust transportation planning," in *Proc. Triennial Symp. Transp. Anal.*, 1998, pp. 1–13.

NASIM FERDOSIAN received M.Tech. degree (Hons.) in computer science from Jawaharlal Nehru Technological University, India, in 2010. She is currently pursuing the Ph.D. degree with the Faculty of Computer Science and Information Technology, Universiti Putra Malaysia. She has authored a number of articles in high impact factor scientific journals. Her current research interests are mainly focused on the study of new radio resource management algorithms for the next generations of cellular networks and quality-of-service in wireless communications and cellular networks.

MOHAMED OTHMAN received the Ph.D. degree (Hons.) from the Universiti Kebangsaan Malaysia. He is currently a Professor with the Faculty of Computer Science and Information Technology, Universiti Putra Malaysia (UPM). He is also an Associate Researcher with the Laboratory of Computational Science and Mathematical Physics, Institute of Mathematical Research, UPM. He published over 160 International journals and 230 proceeding papers. His main research interests are in the fields of high speed network, parallel and distributed algorithms, wireless network (MPDU- and MSDU-Frame aggregation, TCP performance, MAC layer, resource management, and traffic monitoring) and scientific telegraph equation and modeling. He received the Best Ph.D. Thesis Award in 2000 by Sime Darby Malaysia and Malaysian Mathematical Science Society.

BORHANUDDIN MOHD ALI (SM'–) received the B.Sc. degree (Hons.) in electrical and electronics engineering from the Loughborough University of Technology in 1979, and the M.Sc. and Ph.D. degrees in electromagnetic engineering, from the University of South Wales, in 1981 and 1985, respectively. He served as the Director of the Center of Excellence, Institute of Multimedia and Software, with Universiti Putra Malaysia (UPM), from 2002 to 2006, spent one year with Celcom Research and Development as a Visiting Scientist in 1995, and two years with Might as a Senior Manager from 1996 to 1997, charged with coming up with research and policy direction for Malaysian Telecommunication Industry. He was on secondment to MIMOS Bhd, a government research laboratory on ICT, heading the Center of Wireless Communications. He is currently a Full Time Professor with the Department of Computer and Communication Engineering, Faculty of Engineering, UPM. His research interest spans wireless and broadband communications, and network engineering. In 1996, he helped to realize the formation of Teman project, and later was made the Chairman of the MyREN Research Community, a National Research Testbed. He is a Chartered Engineer and a member of the IET. He was the Chair of ComSoc Chapter, from 1999 to 2002, and the Chair of the IEEE Malaysia Section from 2002 to 2004.

KWEH YEAH LUN received the M.S. and Ph.D. degrees in the field of Parallel and Distributed Computing in computer science from Universiti Putra Malaysia in 2004 and 2011, respectively. He has been a Senior Lecturer with the Department of Communication Technology and Networking, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, since 2011. His research interests include distributed algorithm, peer-to-peer networks, and distributed computing.