A Fairness-Based Preemption Algorithm for LTE-Advanced

Mehdi Khabazian[†], Osama Kubbar[†] and Hossam Hassanein[‡]

[†] Qatar University Wireless Innovation Center (QUWIC),
Doha, Qatar.
[‡] Queens University,
Kingston, Canada.
Email: {mehdik, osamak}@quwic.com, hossam@cs.queensu.ca

***Abstract*—In this paper, we propose a fairness-based preemption algorithm for 3GPP LTE-Advanced. It considers bearers' priorities as well as their QoS over-provisioning with respect to their minimum QoS needs into the partial preemption decision. We define a preemption contribution metric for each established bearer and we evaluate the proposed scheme both in terms of inter- and intra-priority preemption fairness. Through a simulation approach, we conclude substantial improvements in preemption fairness when compared to a traditional preemption approach. We discuss that the proposed scheme does not affect the main performance measurements of the network, i.e. the bearers' blocking and dropping probabilities due to congestion. Finally, we discuss that the preemption contribution metric can be effectively used to variate the total generated revenue.**

I. INTRODUCTION

As one of the *Radio Resource Management* (RRM) functionalities in LTE systems, *call admission control* (CAC) is employed to control the number of LTE bearer requests in order to maintain the QoS of the admitted bearers. However, no guarantee can be provided due to inherently dynamic nature of wireless communication. For example, during congestion periods when several communications experience poor channel quality or high mobility [1], it is highly possible that the network cannot maintain its bearers' QoS requirements. Thus, preemption schemes may be employed to alleviate the situation. Preemption is unavoidable in two circumstances, namely, to manage the resources among bearers when the network is overloaded as a congestion control mechanism or to allocate a high-priority bearer request while sufficient resources are not available. In general, two types of preemption can be configured, namely partial and full preemptions. In partial preemption which is also referred to as service degradation or bandwidth adaptation, resources assigned to one or more low-priority bearers are partially released up to their minimum QoS needs. On the other hand, by full preemption which is also referred to as cutoff process, a high-priority bearer may completely preempt the resources assigned to one or more low-priority bearers. Preemption plays a critical role in support of achievements in the higher layers such as enhancements in scalability of voice and video coding schemes which let the quality of experience stay optimal regardless of network condition variations [2]. Preemption mechanisms has been addressed in several studies such as in [3]–[6]. [3] suggests a joint CAC and partial preemption for heterogeneous wireless networks. By defining a service utility function, they represent a generic measurement of the profit that is gained by the mobile network operator. Their work aims to improve the profit while taking the connection-level QoS of the services into consideration. [5] presents several partial preemption policies for supporting variable bit-rate multimedia. By the concept of node contribution adaptation, they allow each node to dynamically adapt its needed bandwidth and alter its contribution to maintain a particular bandwidth contribution skew across all nodes in the system. [6] offers a partial preemption algorithm which tries to minimize the number of degraded calls and the number of calls receiving lower bandwidth than the requested. Their algorithm has been developed for the case of single service networks and tested for video traffic. Finally, [4] discusses the importance of preemption mechanisms in handling congestion in wireless multimedia networks. They investigate the effect of some QoS factors such as traffic priorities and degradation thresholds on the performance of preemption mechanisms.

The aforementioned studies do not optimally conceptualize the preemption problem for LTE since they have been mostly developed for general wireless network models in which LTE requirements as indicated in 3GPP have not been included. As an example, the bearer level multi-class QoS treatment [7] has been neglected in their modelings. On the other hand, very few studies on LTE preemption have been reported in the literature. [8] proposes an LTE full preemption mechanism. To release target resources $\Delta\rho$, it first sorts the LTE bearers based on their priority and QoS over-provisioning. Then, the bearers with the lowest ranks are one-by-one removed until the target load is released. This method is not necessarily fair since some bearers with lower priorities may be fully preempted while others are still over-provisioned. [9] proposes a partial preemption scheme which aims to maximize the system capacity while maintaining the basic QoS of the admitted bearers. They first

This work was made possible by a National Priorities Research Program (NPRP) grant from the Qatar National Research Fund (Member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

sort the bearers which belong to the LTE non-guaranteed bit rate (Non-GBR) classes according to their priority levels. Then, it selects the non-GBR bearers with the lowest priority one-by-one to be partially preempted to their minimum QoS needs until the target load is released. This algorithm may still be considered unfair since the preemption applies to each flow individually. In other words, some of the bearers may suddenly be degraded to their basic data rate while others from the same priority still hold their maximum data rate. We note that from the application point of view, such a disruptive and sudden bearer degradation is not desirable. We may consider a video streaming service which is suddenly downgraded from its maximum data rate to its basic data rate which is not appealing for a viewer.

In this study, we propose a partial preemption scheme by which each established bearer may be preempted according to its priority as well as the amount of extra allocated resources compared to its basic data rate. Thus, we introduce a preemption contribution metric for each bearer. We compare the fairness of the proposed scheme with a traditional preemption scheme by means of two well-known fairness indices, i.e. Jain's index and min-max index. We also discuss its effect on bearers' blocking and dropping probability and total generated revenue[1].

The remainder of the paper is organized as follows: Section II details the LTE QoS provisioning model, Section III presents the CAC model of the study, Section IV describes the proposed partial preemption mechanism. Fairness evaluation and simulation results are provides in Section V and the final section concludes the paper.

II. LTE QoS Provisioning Model

In the LTE evolved packet system (EPS), QoS provisioning of the established service data flows (SDFs) are treated by the concept of *bearers* classification [10]. The bearer classes are limited and each class is assigned a *priority* level. Bearers from the same priority are treated with similar maximum resource budgets in terms of QoS needs. 3GPP has defined a number of bearer's classes along with their associated QoS attributes and priorities which can be found in [11]. This standard pre-configuration was mainly done to unify multi-vendor deployments and roaming. Two types of bearers were defined as guaranteed bit rate (GBR) and non-guaranteed bit rate (non-GBR) bearers. Resource assignment for GBR bearers are associated with two parameters of maximum bit rate (MBR) and guaranteed bit rate (GBR) which, respectively, they refer to the maximum sustained data rate that the GBR bearer cannot exceed and the minimum reserved data rate that the network guarantees for the GBR bearers. Thus, depending on resource availabilities, a bearer's data rate may be over-provisioned up to its maximum bit rate. On the other hand, resource assignments for non-GBR bearers are associated with a system parameter called aggregate MBR (AMBR) which defines the total amount of bandwidth resources that can be assigned to a group of non-GBR bearers [11].

Each bearer is associated with a key attribute called *allocation and retention priority* (ARP) which determines its priority and capability in full preemption process [7], [11]. ARP is used by CAC and congestion control to establish or to modify the bearers. A bearer ARP consists of three information fields, namely, priority, preemption capability and preemption vulnerability. The priority is used for differentiation purposes and it ensures that the request of the bearer with the higher priority level is preferred. For example, we may consider a video-conference service whose video and voice SDFs were mapped into a low-priority and a high-priority bearers, respectively. Thus, in case of congestion, eNodeB may drop the video bearer while keeping the voice bearer to preserve the service continuity. The second ARP field, i.e., the preemption capability is a flag and when it is set, it authorizes the bearer to fully preempt a preemptable bearer with a lower priority. Finally, the preemption vulnerability field is also a flag and when it is set, it means that an established bearer can be fully preempted by a preempting capable high-priority bearer.

III. CAC Model

In this section, we describe the LTE CAC mechanism which will be jointly considered with the proposed preemption scheme. First, let us define the following notations:

i: an index which denotes the priority of a bearer.

n_i: number of active bearers with priority i.

b_{gi} : guaranteed bit rate of GBR bearers with priority i.

b_{mi} : maximum bit rate of GBR bearers with priority i. In this study, we extend the definitions of b_{gi} and b_{mi} for non-GBR bearers. Thus, b_{gi} denote the minimum data rate at which the non-GBR service can be maintained in an acceptable level. Acceptable data rate for real-time non-GBR bearers maybe defined as the rate at which a minimum quality of experience (QoE) is experienced, while for non real-time non-GBR bearers such as FTP, it may be defined as a data rate at which the file can be transfered within a reasonable time. non-GBR b_{mi} is also defined as the maximum bit rate which may be assigned to the non-GBR bearers with priority i. We note that for a non-GBR service, b_{gi} may be set to zero to present a best-effort service with no QoS restriction while b_{mi} can be set to AMBR parameter as defined earlier.

$a_{i,j}(b)$: allocated OFDM physical radio block (PRB) resources to the j^{th} bearer of priority i to achieve bit rate b. We note that for given allocated resources, the achievable bit rate by the bearer depends on the channel quality and consequently, the employed modulation scheme.

C: total capacity of the LTE cell excluding reserved capacity for hand-off traffic in PRBs.

$\Delta\rho$: the amount of resources which is needed to be released by preemption for admitting a new bearer with priority i.

We assume that a new bearer request is admitted if the required resources to fulfill the application QoS is either

[1]We note that the main contribution of this work is the development of a fair *partial* preemption algorithm; however, we may refer to it by the general term of preemption since two phases of partial and full preemptions are not separable.

available or can be preempted according to the preemption algorithm described in the following section. Thus, denote $\overline{a_{i,j}}$ as the load contribution of the j^{th} bearer of priority i in PRB, the availability of resources may be estimated as $C - \sum_{i,j} \overline{a_{i,j}}$. One way to determine the load contribution of an established bearer is by averaging its recent load over a measurement window [12] as:

$$\overline{a_{i,j}} = min\left(K_c, \frac{m_{i,j}}{C\Delta t}\frac{b_{gi}}{T_{i,j}}\right), \quad 1 \leq j \leq n_i, \tag{1}$$

where $m_{i,j}$ is the total number of PRBs allocated by the MAC scheduler to the j^{th} bearer with priority i over a measurement window of Δt sub-frames and $T_{i,j}$ is the achieved throughput of that bearer over the measurement window. In (1), K_c is a constant with a value close to 2.0 which may be used to keep the load under a reasonable value in case $T_{i,j}$ becomes very small due to an extreme poor channel quality for a particular bearer [12].

Thus, assuming that resource ρ_{i,n_i+1} is needed to establish a new bearer with priority i, i.e., the $(n_i+1)^{\text{th}}$ bearer, we may write its admission and resource allocation criteria as:

$$\rho_{i,n_i+1} = \begin{cases} a_{i,n_i+1}(b) | b_{gi} < b < b_{mi}, & \text{if} \sum_{j=1}^{9}\sum_{k=1}^{n_j} \overline{a_{j,k}} \leq C - a_{i,n_i+1}(b) \\ a_{i,n_i+1}(b_{gi}), & \text{if} \sum_{j=1}^{9}\sum_{k=1}^{n_j} \overline{a_{j,k}} - \Delta\rho \leq C - a_{i,n_i+1}(b_{gi}) \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

where the first condition determines whether the available resources are enough to admit a bearer with a data rate within its QoS interval. The second condition presents the situation where the system does not have enough resources but it may be obtained by preemption. Finally, the bearer request is denied if the required resources are neither available nor preemptable.

IV. Proposed Preemption Algorithm

Traditional preemption mechanisms work in two phases. The first phase allows preemption of resources by reconfiguring the lowest priority bearers to their minimum QoS needs and the second phase allows total preemption of some low-priority bearer resources. In the following, we propose a fairness-based preemption algorithm which is unique and aligned with 3GPP requirements. Similarly, this algorithm is working in two phases. Given that the extra target load $\Delta\rho$ is needed for admission of a new bearer with priority i, first, it tries to partially preempt the resources of the active bearers based on their priorities as well as their over-provisioned resources in a distributed fashion. Let us define $f_{i,j}$ as the partial preemption contribution metric for the j^{th} bearer with priority i as:

$$f_{i,j} = i^{\alpha}(\overline{a_{i,j}} - bg_i)^{\beta}, \quad 1 \leq j \leq n_i, \tag{3}$$

where i is the bearer priority and $(\overline{a_{i,j}} - bg_i)$ denotes the amount of the over-provisioned resources which were assigned to that bearer with respect to the needed resources to provision its guaranteed bit rate. From (3), contribution metric increases when the priority of an established bearer or its QoS over-provisioning increases. α and β are fine tuning knobs which may control the fairness of the metric. Larger α and β result in higher contribution metrics for low priority bearers and highly over-provisioned bearers, respectively. We note that (3) is in the form of Cobb-Douglas production function. In economics, this function is widely used to represent the relationship of an output to inputs. This function has the general form $Y = A^{\alpha}B^{\beta}$ where Y represents the total production (output) and A and B, respectively, represent the labor and capital inputs. Thus, α and β may be interpreted as the output elasticities of labor and capital. Output elasticity measures the responsiveness of output to a change in levels of either labor or capital. Thus, following the Cobb-Douglas function properties, three cases may be considered for tunning knobs in (3). First, if $\alpha+\beta=1$. In this case, production function has constant returns to scale. In other words, doubling priority i and resource over-provisioning $(\overline{a_{i,j}}-bg_i)$ will also double the contribution metric $f_{i,j}$. On the other hand, when $\alpha+\beta<1$ and $\alpha+\beta>1$, returns to scale shows a decreasing and increasing properties, respectively.

Finally, the normalized contribution of each established bearer in partial preemption, denoted by $d_p(i,j)$ is defined as:

$$d_p(i,j) = \frac{f_{i,j}}{\sum_{i,j} f_{i,j}}. \tag{4}$$

Thus, to release target load $\Delta\rho$, we propose the following partial preemption algorithm:

Algorithm 1 Partial Preemption

while Target load $\Delta\rho$ is not released: **do**
  Find $f_{i,j}$ for all established bearers according to (3).
  if $f_{i,j} = 0$ for all i, j **then**
    Break (Target load $\Delta\rho$ cannot be released by partial preemption.)
  end if
  If possible, degrade all active bearers partially for $d_p(i,j)\Delta\rho$. Thus, the maximum assigned resources to the j^{th} bearer of priority i will be degraded to $\max[bg_i, \rho_{i,j} - d_p(i,j)\Delta\rho]$.
end while

In the above, the algorithm is stopped whenever metric $f_{i,j}$ approaches to zero for all active bearers. This indicates that all bearers are already degraded to their guaranteed bit rates. In this case, full preemption is considered in the second phase only if the requesting bearer is preempting capable. Thus, all bearers whose preemption vulnerability flag were set to one will be sorted according to their priority in decreasing order and their resources are fully preempted one-by-one until the target load is achieved. At the end of the second phase, if still the target load is not achieved, the new bearer request will be rejected.

We note that the proposed mechanism not only partially preempt the bearers in a fair and distributed scheme, but also it allows the service provider to control the preemption fairness by fine-tuning the bearers' contributions based on their priorities and QoS over-provisioning factors.

V. Performance Evaluation

A. Simulation Settings

The simulations are done in the downlink of a single-cell LTE network; however, the same principle applies in the uplink direction. For simplicity, we assume that a user assigns only one SDF of a kind to a bearer with priority i^2.

For the physical layer aspects, the same modeling of [13] is considered which is in conformation to 3GPP LTE specifications [14]. Thus, the path loss is modeled by $39.95 + 43.375log(d/10)$ (in dB) where d is the distance between the user and eNodeB in meters. Total eNodeB transmission power and the noise power are set to 43 dBm and -174 dBm/Hz, respectively. The OFDM channel consists of a 5 GHz band and it contains 25 PRBs. Each PRB consists of 12 sub-carriers with 15 kHz spacing bands. Finally, a frequency selective fading channel with six taps is assumed. The perceived data rate at PRB j of user i is given by $R_{i,j} = 12log_2(1 + \beta SNR_{i,j})$, where $\beta = \frac{1.5}{-ln(5BER)}$ and $SNR_{i,j}$ is the average of signal to noise ratios on all 12 sub-carriers of PRB j of user i [13]. A proportional fair algorithm is used to schedule the traffic on each PRB in the time domain, whenever a PRB is partially used to support the requested data rate. It is also assumed that 20% of the resources are reserved for incoming hand-off calls and excluded from the analysis. The mobility model of [1] was incorporated to model users' movements.

Three traffic types, namely, voice, video and FTP are assumed and their characteristics are modeled based on the requirements specified by the Next Generation Mobile Networks (NGMN) consortium [15]. Table I details the modelings. We consider four types of bearers with priorities 1 to 4 whose QoS needs are indicated in Table II. For multimedia, a wide data rate intervals are assumed to support scalable multimedia codings such as voice EVS and video H.264/SVC as suggested for future LTE networks [2].

B. Fairness Evaluation

We use Jain's and min-max indices [16] to compare the proposed preemption mechanism with the traditional approach in terms of fairness. Jain's index rates the fairness of a set of allocations such that x_i portion of a total resource is assigned to the i^{th} user among n total users. Thus, it ranges from $1/n$ as a worst case allocation to 1 as the best case allocation where all users are allocated equally. We use this index to measure the fairness among different priorities of bearers and we refer to it as inter-priority fairness. Jain's index is given by:

$$\text{Jain's fairness index} = \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n\sum_{i=1}^{n} x_i^2}, \tag{5}$$

[2] In reality, a user may initiate two or more data services with similar QoS needs which will be mapped to the same bearer.

where n denote the number of bearer's priorities ($n = 4$ for our settings). To make the definition applicable to our scenario, we define allocation x_i as the average ratio of resource over-provisioning of the bearers of class i as follows:

$$x_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \frac{\overline{a_{i,j}} - a_{i,j}(b_{gi})}{a_{i,j}(b_{mi}) - a_{i,j}(b_{gi})}. \tag{6}$$

A smaller x_i indicates a lower QoS over-provisioning for bearers with priority i and vice-versa. Thus, the measured Jain's index represents the bearers' QoS over-provisioning fairness in an inter-priority fashion.

Next, we employ min-max index to measure the bearers' intra-priority preemption fairness which it may be defined as the QoS over-provisioning fairness among bearers with the same priority. Min-max fairness index is given by:

$$\text{Min-max fairness index} = \frac{\min(y_j)}{\max(y_j)}, \tag{7}$$

where y_j denote the QoS over-provisioning for the j^{th} bearer with pririty i and it can be found as:

$$y_j = \overline{a_{i,j}} - a_{i,j}(b_{gi}). \tag{8}$$

We note that all bearers with the same priority should contribute equivalently in partial preemption process. Thus, minmax index is more appropriate in this case since it is sensitive to service degradation and service unfairness among users within the same class [17].

C. Simulation Results

In the following, we assume $\alpha + \beta = 1$ which as described, it corresponds to a perfect competition scenario in Cobb-Douglas presentation of contribution metric (3). Fig. 1 presents the inter-priority fairness measurements (Jain's index) as a function of traffic arrival rates for the three cases of traditional preemption approach and fairness-based approaches with $\alpha = 0.9$ and $\alpha = 0.1$. We note that

TABLE I: Traffic distributions and parameters

FTP	
File size	Truncated Log-normal distribution with maximum = 5 Mbytes, mean = 2Mbytes and Standard deviation = 0.722 Mbytes
Service time	Until file transfer is completed
Video	
Frame inter-arrival time	100msec (Deterministic)
No. of slices (packets) per frame	8 slices/frame (Deterministic)
Packet (Slice) size	Truncated Pareto Distribution with mean = 100 bytes and maximum = 250 bytes
Packet inter-arrival time in a frame	Truncated Pareto Distribution with mean=6 msec and maximum = 12.5 msec
Service time	Exponentially distributed with mean = 100 Sec
Voice	
Source rate	Variable rates
Talk spurt length	Exponentially distributed with mean = 147 msec
Silent period length	Exponentially distributed with mean = 167 msec
Service time	Exponentially distributed with mean = 100 Sec

TABLE II: LTE Traffic Parameters Settings

Traffic Type	Priority	Guaranteed bitrate (kb/sec)	Maximum bitrate (kb/sec)	Packet error rate	Preemption vulnerability	Preemption capability
Voice, GBR	1	32	64	10^{-2}	0	1
Video, GBR	2	128	256	10^{-3}	0	1
Video, non-GBR	3	10	64	10^{-2}	1	0
FTP, non-GBR	4	10	256	10^{-3}	1	0

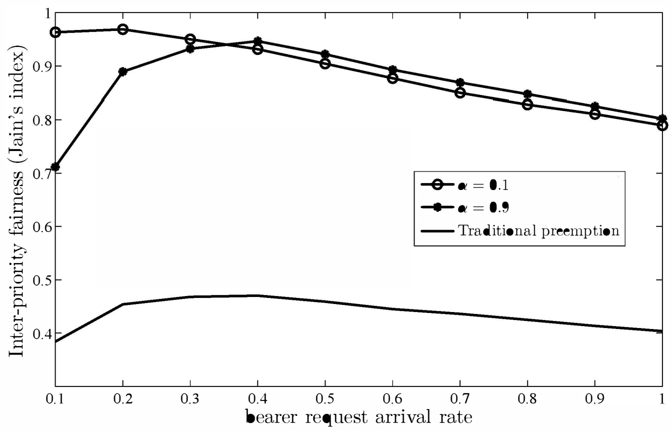

Fig. 1: Inter-priority fairness measurement (Jain's index) as a function of traffic batch arrival rate.

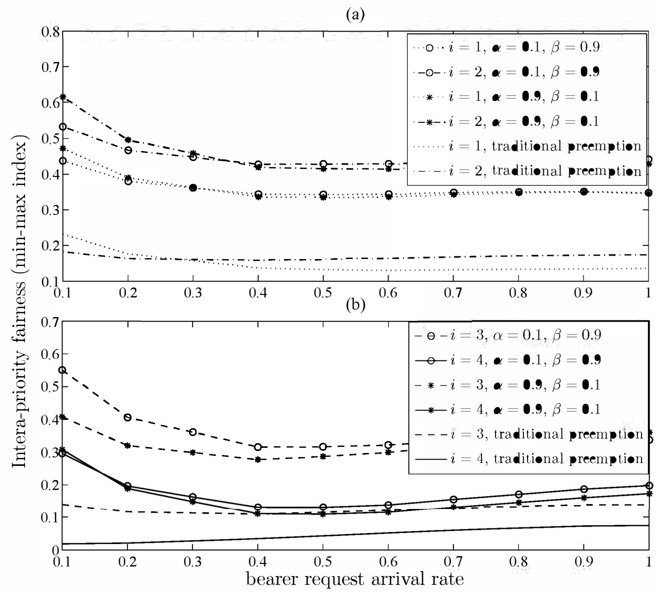

Fig. 2: Intra-priority fairness measurement (min-max index) as a function of traffic batch arrival rate.

$\alpha = 0.9$ and $\alpha = 0.1$ result in bearer's contribution metrics with dominant priority and dominant QoS over-provisioning, respectively. The x axis represents the traffic batch arrival rates for all priorities. Further, arrival rates are chosen such that the network is considered lightly congested for the values less than 0.3 and heavily congested afterward. As it may be seen, the employment of fairness-based preemption will substantially increase the preemption fairness among the four bearer classes. Further, when α is decreasing from 0.9 to 0.1 (middle values are not shown on the figure), the preemption fairness improves in the lightly congested region while they are almost identically decreasing when the network becomes highly congested. We note that inter-priority fairness is not always interesting since a service provider may prefer to value some bearers with higher priorities even for partial preemption. This can be done by excluding those high-priority bearers from algorithm 1, thus not allowing a low-priority bearer to partially preempt the high-priority ones.

Fig. 2 (a) and (b) plot the intra-priority fairness measurements (min-max index) as a function of traffic batch arrival rate for GBR and non-GBR bearer classes, respectively. As before, three cases of traditional preemption approach and fairness-based preemption approach with $\alpha = 0.9$ and $\alpha = 0.1$ are considered. As it may be seen, the partial preemption fairness among the bearers of the same priority has been improved substantially when the proposed algorithm is employed; however, the variation in tuning knobs negligibly affects the fairness index. Thus, while the proposed scheme assures identical QoS over-provisioning for all bearers within the same priority, the tuning knobs may be used effectively to adjust inter-priority fairness among bearers.

Next, we have measured the bearers' blocking and dropping probabilities. These two measurements are defined, respectively, as the probability that a bearer request is not admitted as a result of congestion and the probability that a bearer is dropped when it can not maintain its guaranteed bit rate within congestion avoidance phase. As an important observation, we have noticed that the proposed algorithm affects these measurements negligibly for different values of α and β[3]. For example, for $n_i = 4$ with fairness-based preemption mechanism, two aforementioned probabilities, respectively, increase from 0.12 to 0.125 and from 0.041 to 0.051 for the batch arrival rate of 0.6 bearers/sec and $\alpha = 0.1$ which indicate less than 5% deviations. Thus, we conclude that the achieved fairness both in intra- and inter-priority preemption does not result in major deterioration or improvement in bearers' blocking and dropping probabilities.

Next, we measure the total generated revenue in a period of T hours. We assumed a linear batch arrival rate function which is increasing from 0.1 to 1 bearer requests/sec within a period of $T = 7$ hours and we consider service charges of 1,

[3]The results have not been presented since as aforementioned, the differences in observations were negligible.

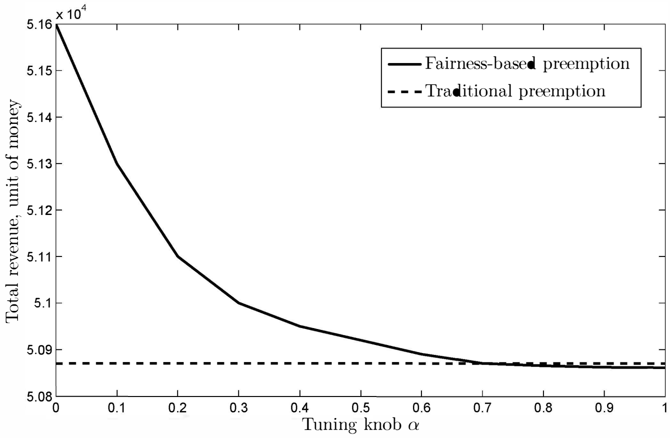

Fig. 3: Total generated revenue as a function of preemption tuning knob α over a period of $T = 7$ hours.

0.15, 0.12 and 0.1 money unit/Mb for bearers with priority 1 to 4, respectively. The generated revenue for the given period is measured and sketched in fig. 3 as a function of tuning knob α (or equivalently $1-\beta$). The result of traditional preemption is also inserted for comparison reason. As it may be observed, the total revenue is decreasing from 50.9k to 51.6k when α is increasing from 0 to 1 which almost shows a 1.4% variation. The result indicates the capability of preemption mechanism in altering generated revenue. Although a deceasing effect is observed for the current network settings, the trend highly depends on the traffic parameters' settings such as designated gauranteed/maximum bit rate and their prioritization schemes. Thus, we may even expect local extrema for some other settings. Thus, definition of an appropriate optimization problem is interesting, especially to adjust the tuning knobs to maximize the generated revenue. This task has been set as a future work.

VI. Conclusion

Variable bit-rate supported applications such as multimedia scalable coding have been recently considered mainly to cope with inherent channel quality and network traffic variations. Thus, it brought new challenges for adaptive QoS provisioning in advanced wide-band communications such as LTE-Advanced. Although in the current version of LTE, the MBR and GBR parameters are assumed equal, they are considered as two main bearers' QoS parameters in 3GPP R10 and afterward. Thus, it is assumed that the bearers are treated with a wide QoS interval which are limited with these two parameters. As a result, resource preemption mechanism and its fairness issue will be a prominent issue since it may directly affect applications' QoS in the higher layers as well as other network attributes such as generated revenue. In this work, we proposed a fairness-based preemption algorithm which take into consideration the bearer priority as well as the amount of bearer's QoS over-provisioning. We defined the contribution of each bearer in preemption through a contribution metric which was presented in the form of Cobb-Douglas production function. The inter-priority and intra-priority fairness of the proposed scheme was evaluated through Jain's and min-max fairness indices, respectively and were compared with a traditional preemption scheme. We discussed that while the proposed algorithm improves the preemption fairness significantly, it does not alter two important performance measures of bearers blocking and dropping probabilities. Finally, we showed its capability to fairly modify the generated revenue by the service providers which the later has been set as a future work.

References

[1] M. Khabazian, O. Kubbar, and H. Hassanein, "Call admission control with resource reservation for multi-service ofdm networks," in *Proc. IEEE International Conference on Computing, Networking and Communications (ICNC'12)*, Feb. 2012, pp. 1–6.

[2] K. Jarvinen, I. Bouazizi, L. Laaksonen, P. Ojala and A. Ramo, "Media coding for the next generation mobile system LTE," *Elsevier Computer Communications*, vol. 1, pp. 1916–1927, April 2010.

[3] A. Klein, C. Lottermann, C. Mannweiler, J. Schneider and H.D. Schotten, "A novel approach for combined joint call admission control and dynamic bandwidth adaptation in heterogeneous wireless networks," in *Proc. EURO-NGI Conference on Next Generation Internet (NGI'11)*, June 2011, pp. 1–8.

[4] A.M. Rashwan, A.E.M. Taha and H.S. Hassanein, "Considerations for bandwidth adaptation mechanisms in wireless networks," in *Proc. Biennial Symposium on Communications*, June 2008, pp. 43–47.

[5] S. Schubert, F. Uyeda, N. Vasic, N. Cherukuri and D. Kostic, "Bandwidth adaptation in streaming overlays," in *Proc. International Conference on Communication Systems and Networks (COMSNETS'10)*, Jan. 2010, pp. 1–10.

[6] N. Nasser and H. Hassanein, "Combined admission control algorithm and bandwidth adaptation algorithm in multimedia cellular networks for QoS provisioning," in *Proc. Canadian Conference on Electrical and Computer Engineering*, vol. 2, May 2004, pp. 1183–1186.

[7] 3GPP TS 23.401 V9.5.0, "Technical specification group services and system aspects; gprs enhancements for e-utran access," June 2010.

[8] R. Kwan, R. Arnott, R. Trivisonno, and M. Kubota, "On pre-emption and congestion control for LTE systems," in *Proc. IEEE Vehicular Technology Conference Fall (VTC'10-Fall)*, Ottawa, Canada, Sep. 2010, pp. 6–9.

[9] M. Qian, Y. Huang, J. Shi, Y. Yuan, L. Tian and E. Dutkiewicz, "A novel radio admission control scheme for multiclass services in LTE systems," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM'09)*, Dec. 2009, pp. 1–6.

[10] 3GPP TS 36.300 V8.12.0, "Technical specification group radio access network; evolved universal terrestrial radio access (e-utra) and evolved universal terrestrial radio access network(e-utran); overall description; stage 2, (release 8)," March 2010.

[11] 3GPP TS 23.203 V8.9.0, "Technical specification group services and system aspects; policy and charging control architecture (release 8)," March 2010.

[12] R. Kwan, R. Arnott and M. Kubota, "On radio admission control for lte systems," in *Proc. IEEE Vehicular Technology Conference (VTC'10-Fall)*, Ottawa, Canada, Sep. 2010, pp. 1–5.

[13] D. K. Kim, D. Griffith, and N. Golmie, "A new call admission control scheme for heterogeneous wireless networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 10, pp. 3000–3005, Oct. 2010.

[14] 3GPP TS45.005 v8.3.0, "Radio access network; radio transmission and reception (release 8)," Nov. 2008.

[15] NGMN Alliance, "Next generation mobile networks radio access performance evaluation methodology," *White Paper*, Jan. 2008.

[16] J. F. Kurose and K. W. Ross, *Computer Networking: A Top-Down Approach*, 5th ed. Addison-Wesley, 2010.

[17] N. A. Ali, P. Dhrona, and H. Hassanein, "A performance study of uplink scheduling algorithms in point-to-multipoint WiMAX networks," *Elsevier Computer Communications*, vol. 32, no. 3, pp. 511–521, Feb. 2009.