

Article

A Novel Algorithm for Efficient Downlink Packet Scheduling for Multiple-Component-Carrier Cellular Systems

Yao-Liang Chung

Department of Communications, Navigation and Control Engineering, National Taiwan Ocean University, Keelung City 20224, Taiwan; ylchung@email.ntou.edu.tw; Tel.: +886-2-2462-2192 (ext. 7224)

Academic Editor: William Holderbaum

Received: 23 September 2016; Accepted: 8 November 2016; Published: 15 November 2016

Abstract: The simultaneous aggregation of multiple component carriers (CCs) for use by a base station constitutes one of the more promising strategies for providing substantially enhanced bandwidths for packet transmissions in 4th and 5th generation cellular systems. To the best of our knowledge, however, few previous studies have undertaken a thorough investigation of various performance aspects of the use of a simple yet effective packet scheduling algorithm in which multiple CCs are aggregated for transmission in such systems. Consequently, the present study presents an efficient packet scheduling algorithm designed on the basis of the proportional fair criterion for use in multiple-CC systems for downlink transmission. The proposed algorithm includes a focus on providing simultaneous transmission support for both real-time (RT) and non-RT traffic. This algorithm can, when applied with sufficiently efficient designs, provide adequate utilization of spectrum resources for the purposes of transmissions, while also improving energy efficiency to some extent. According to simulation results, the performance of the proposed algorithm in terms of system throughput, mean delay, and fairness constitute substantial improvements over those of an algorithm in which the CCs are used independently instead of being aggregated.

Keywords: efficiency; multiple component carriers; proportional fair; 4G/5G cellular systems

1. Introduction

The development and deployment of the 4th generation (4G) cellular system was undertaken by the International Telecommunication Union because user data demands have continued to grow at a rapid rate, thus necessitating further enhancement of cellular network capacities. Indeed, the advancement of cellular technologies towards 5th generation (5G) systems is already underway in order to accommodate still further increases in user needs. In order to be effective, 4G and 5G systems must be capable of supporting data requirements within multi-user environments in a way that meets both economic and service demands. Additionally, the two key aims of 5G systems are spectral efficiency and energy efficiency. However, the challenge of how to couple and integrate these disparate objective performance metrics efficiently has yet to be resolved. Previously, complete overviews for 4G systems were published by Parkvall and Astely [1] and Martín-Sacristán et al. [2], while articles by Andrews et al. [3], Wu et al. [4], Gupta and Jha [5], Hossain and Hasan [6], and Han et al. [7] surveyed recent advances in key 5G system technologies in detail.

Among those advances, the ability of base stations (BSs) to make use of multiple component carriers (CCs) simultaneously during data transmissions is among the key characteristics of 4G/5G systems that makes them more powerful than less advanced networks, allowing them to achieve substantially higher capacities. In other words, the aggregation of multiple CCs in a single BS allows 4G/5G systems to support much broader transmission bandwidths than the more basic networks are capable of supporting.

One of the key technical challenges of 4G/5G systems is the design of ideal packet schedulers, since any effective packet scheduler design must meet the following requirements: (1) It must effectively provide packet scheduling in environments with multiple CCs; (2) It must support the necessary quality-of-service (QoS) requirements even while handling various kinds of traffic, such as real-time (RT) traffic (i.e., QoS-sensitive traffic) and non-RT (NRT) traffic (i.e., QoS-non-sensitive traffic); (3) It must achieve high overall system throughput; and (4) it must maintain the necessary level of fairness among users.

There have already been considerable research efforts dedicated to resource management designs for use in cellular systems, with much of those efforts focused on systems that make use of multi-user single-CC BSs [8–14]. For example, a scheme consisting of enhanced efficiency cross-layer packet scheduling and resource management was proposed by Jeong et al. [8], while various fast algorithms designed to provide optimal resource allocation that would, in turn, maximize overall system utility were proposed by Madan et al. [9]. In 2012, a study by Ng et al. [10] focused on the question of how resources should be allocated in order to ensure energy-efficient communication in single-cell environments with a high number of transmitting antennas. In a subsequent study, the same authors went further by also addressing that question with regard to multi-cell environments with cooperative BSs [11]. In a 2013 study, Ng et al. investigated the use of a type of BS that engages in hybrid energy harvesting for data transmissions; in the BS in question, an energy harvester was used in combination with a constant energy source provided from a non-renewable resource in order to supply the system with the energy necessary for its operation [12]. Taking on the issue of energy efficiency from another perspective, the development of heterogeneous network deployments involving both small cells and macrocells was investigated by both Morosi et al. [13] and Chung [14] as a means of enhancing energy efficiency. Unfortunately, while all these studies (i.e., [8–14]) contributed some important insights, the issue of using multiple-CC BSs for packet scheduler designs was not specifically considered by any of them.

A range of key issues affecting multiple-CC BS packet scheduler designs have, however, been investigated in various studies published in recent years [15–25]. This research has addressed the two key aims of 5G systems: spectral efficiency [15–20] and energy efficiency [21–25]. The performance of both throughput and delay for elastic traffic in systems utilizing the mechanism of multiple-CC transmission was previously analyzed by Lei and Zheng [15]. Various scheduling algorithms for CCs and scheduler structures were proposed in papers by, respectively, Chen et al. [16] and Takeda et al. [17], while a user grouping-based algorithm for resource scheduling was proposed in a study by Songsong et al. [18]. Additionally, a method of minimizing packet transmission delay was proposed by Chung and Tsai [19], who described a quantized water-filling packet scheduling scheme with multiple-CC transmission. A study by Wang et al. [20] presented a number of CC allocation schemes addressing uplink environments. In a 2013 study, the issue of how to configure users on a subset of CCs in order to provide energy-efficient transmissions was investigated by Sundaresan and Rangarajan [21], who specifically addressed the issue by considering both CC selection and packet scheduling. Also in 2013, Chen et al. [22] published a study in which they proposed an energy-efficient coordinated scheduling mechanism that could be adapted for use in multi-cell environments. More recently, another study by Chung [23] addressed the problem of minimizing energy at BS transceivers while maintaining required QoS and fairness for all users, using a rate-and-power control scheme for two-CC transmission. A further study by Chung [24] took a system-level perspective to address energy-efficient two-CC transmission, in an extended framework which was proposed and analyzed. Chung [25] next proposed an efficient multi-CC transmission power-saving scheme with greater flexibility than those considered in his previous studies [23,24] for application to BSs aggregating more than two CCs.

However, while the first eight of these previous investigations (i.e., [15–22]) of multiple-CC BS packet scheduler designs made significant contributions to the literature, they did not address the issue of how systems deal with simultaneous RT and NRT traffic. Consequently, the previous studies may not meet ideal scheduler requirements, since RT and NRT traffic forms differ considerably from each

other with regard to both required QoS levels and also traffic characteristics. In addition, however, none of Chung's schemes [23–25] addressed in detail the issue of QoS for different kinds of traffic. Thus far, most of the previous studies have not, to the best of our knowledge, given full consideration to the aforementioned challenges in terms of how they might affect packet scheduler efficiency in 4G/5G systems utilizing multiple-CC transmission. Additionally, given the trend among cellular system operators of co-locating the BSs for mobile networks utilizing various CCs [26], the importance of designing efficient packet schedulers for use in such co-location environments will only increase.

With these points in mind, we herein propose a novel packet scheduling algorithm for use in such environments with being focused on downlink transmissions, where the simultaneous transmission of both RT and non-RT traffic is supported; more specifically, this algorithm is based on the classic proportional fair (PF) criterion that has been proposed by Kelly et al. [27] and ensures that certain resources are used exclusively for the handling of RT packets. Furthermore, the performance of a system utilizing the proposed algorithm is compared with that of a system utilizing the independent CC mechanism, a mechanism in which the CCs are used independently rather than working in concert.

The remainder of this paper is organized as follows. First, the problem description is presented in Section 2. Next, a detailed description and presentation of the proposed efficient packet scheduling algorithm is provided in Section 3. Numerical examples and a demonstration of various simulation results are given in Section 4, followed by a presentation of our conclusions in Section 5.

2. Problem Description

2.1. System Model

For the remainder of this paper, we consider the case of a BS operating in a single cell environment, with our focus being on the specific issue of downlink transmission. For these purposes, the cell in question is comprised of a BS and n user terminals, with each user terminal being indexed as user- k (including both RT users and NRT users). In addition, we assume for the sake of simplicity that there are c adjacent CCs available to be taken for aggregation in the BS within the same frequency band. Next, assume that all the CCs have equal bandwidth and also that the i th ($i = 1, 2, \dots, c$) CC has b_i resource blocks (RBs), where a single RB is herein designed to consist of the smallest allocation unit to be used in resource scheduling. As such, given the aggregation of multiple CCs for transmission, there are a total of $b_{\text{total}} = \sum_{i=1}^c b_i$ RBs that can be utilized for packet transmissions. From a system-level perspective, the design of the conceptual operation of our proposed system model is illustrated in Figure 1.

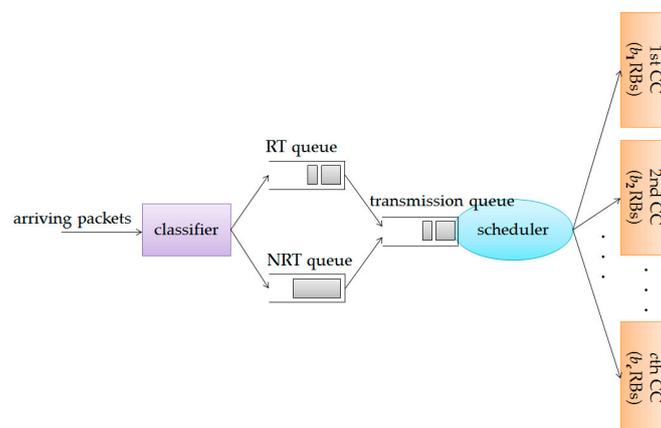


Figure 1. The proposed system model, which consists of a classifier, a real-time (RT) queue, a non-RT (NRT) queue, a transmission queue, a scheduler, and c component carriers (CCs) with the i th CC containing b_i resource blocks (RBs), where $i = 1, 2, \dots, c$, for downlink transmission.

In this system, the first step consists of the classification of all of the arriving packets by the classifier as either RT packets or NRT packets. The packets are then sent on, according to their classification, to either the RT queue or the NRT queue on a first-come-first-served basis. It is then that the proposed packet scheduling algorithm (which will be detailed in Section 3) employed in the scheduler comes into play, with the packets in both the RT queue and the NRT queue being scheduled as appropriate for simultaneous parallel transmission from CCs according to the requirements of the algorithm. As noted above in Section 1, the classic PF criterion [27], which is itself detailed in Section 3, is the foundation upon which the proposed packet scheduling algorithm is based. In addition, the buffer size is assumed to be infinite for each queue shown in Figure 1.

Because of its high robustness against multi-path fading, as well as its high spectral efficiency and bandwidth scalability, orthogonal frequency division multiplexing access (OFDMA) radio technology [28] is utilized as the underlying downlink radio access scheme in the proposed system model. Furthermore, we define the duration of a given OFDMA downlink frame, which is denoted as t_{OFDMA} , to consist of a round for scheduling packets, where, it should be noted, the current round at a given time is identified by the notation s .

2.2. Goal

This study aims to maximize overall system throughput to the greatest extent possible while also ensuring that both the required QoS for the RT traffic and fairness among all the users are maintained. A novel algorithm is proposed to achieve this goal efficiently and effectively, with this aim including improved energy efficiency for the overall system. The algorithm will be described in Section 3.

3. Proposed Algorithm

In this section, the concept of the standard PF criterion for the scheduling of packets is first described in brief. Next, the proposed efficient packet scheduling algorithm for which said PF criterion provides the basis is presented. Thereafter, a baseline scheduler design is introduced for the purposes of comparison. Moreover, we assume that on each RB the transmitted power is the same. It is further assumed that the users in the cell can periodically estimate and then send those estimates of the average signal-to-noise ratios for all the RBs on the c CCs back to the BS. Additionally, in order to obtain the best possible transmission data rate for each transmitted packet, the adaptive modulation and coding mechanism is further utilized in the physical layer [29]. The corresponding downlink data rates are performed with table lookup.

3.1. Classic Proportional Fair Criterion

The ratio of a user's instantaneous data rate to his or her average data rate is indicated by the standard PF utility function. This utility function is important, because in the context of the PF scheduling process the user with the highest utility value is selected. This function can be formally expressed by [27]

$$k^* = \operatorname{argmax}_k \frac{R_k}{\bar{R}_k} \quad (1)$$

where R_k is the instantaneous transmittable data rate at the current time instant of user k and \bar{R}_k is the average data rate of the previous time of user k . According to Kelly et al. [27], the idea behind Equation (1) is to ensure a balance between the maximization of system throughput and the maintenance of fairness among all users.

Meanwhile, Jalali et al. [30] have noted that the PF scheme possesses certain advantages for NRT traffic; specifically, it is capable of achieving substantially greater system throughput than other packet scheduling schemes, such as the round-robin (RR) scheme. In the same study, Jalali et al. [30] also proved that the PF scheme can provide, at least on average, the same level of fairness as the RR scheme.

3.2. Proposed Efficient Packet Scheduling Algorithm

For the purpose of ensuring that the required QoS for the RT traffic is maintained, we set the ratio of b_{NRT} , which consists of the number of RBs that the NRT users can share, to b_{total} , at a specified level η ; i.e., $\eta = b_{\text{NRT}}/b_{\text{total}}$, where $b_{\text{NRT}} \in \{1, 2, \dots, b_{\text{total}}\}$. It should be noted that the selection of those RBs (i.e., the b_{NRT} RBs) begins with the RB on the CC with the index $i = 1$ and then proceeds in sequence. More precisely, all the RT packets can be transmitted over b_{total} RBs, whereas the NRT packets can only be transmitted over ηb_{total} RBs. Consequently, the NRT packets are scheduled on the basis of the PF criterion only with regard to the ηb_{total} RBs available; however, for RT packets, b_{total} RBs are available. In other words, $(1 - \eta)b_{\text{total}}$ RBs are always kept in reserve exclusively for all the RT packets. As a result, the ability of all the RT packets to be transmitted can be protected to some extent when the system is faced with an excess amount of NRT packets.

In addition, in the proposed system the NRT packets buffered in the NRT queue (as shown in Figure 1) are delivered on a periodic basis into the transmission queue for transmission. The length of time for each period between such deliveries is denoted as t_{th} , where that term is defined as consisting of a specified integer multiple of a round from the first round. Meanwhile, the RT packets buffered in the RT queue are delivered into the transmission queue on a round-by-round basis. Moreover, δ_{RT} is defined as the delay constraint for each RT packet.

As the procedure utilized by the proposed algorithm begins, all the average data rates of the users are initialized to a constant value. Then, the algorithm follows the following steps for the current scheduling round s . (It should be noted that “Step 1” constitutes the first operation to be executed at the start of every round for the scheduling of packets.)

Step 1: If the epoch of the current scheduling round is an integer multiple of t_{th} , then those NRT packets from previous rounds still buffered in the NRT queue (if any) are delivered to the transmission queue. It should be noted, again, that only ηb_{total} RBs are available for the delivered NRT packets.

Step 2: Those RT packets from previous rounds still buffered in the RT queue (if any) are delivered to the transmission queue.

Step 3: Those RT packets in the transmission queue for which delay constraints are violated with regard to the threshold δ_{RT} (if any) are dropped.

Step 4: In scheduling packets in the transmission queue, the unused RB on the CC with the index $i = 1$ is started from.

Step 5: The *fairness vector* is then calculated. The *fairness vector*, which is denoted as $\text{tuple}(i^*, j^*, k^*)$, modified from Equation (1), is precisely calculated as follows

$$\text{tuple}(i^*, j^*, k^*) = \underset{j, j, k}{\text{argmax}} \frac{R_k(i, j, s)}{\bar{R}_k(s)}. \quad (2)$$

It should be noted that in Equation (2), $R_k(i, j, s)$ denotes the instantaneous data rate, in the current scheduling round s , from the BS to the user k over the resource RB j on CC i , while $\bar{R}_k(s)$ is the user k 's average downlink data rate up to round $s - 1$.

Step 6: The packet of user k^* is then transmitted over the resource RB j^* on CC i^* . It should be noted that if the RB's size is smaller than that of the packet under consideration, then said packet will be partitioned. The remaining portion of the packet may then be transmitted within the same round or within the next round.

Step 7: The impact of any short-term undesired fluctuations is reduced by updating the average data rate according to the following smoothing exponential filtering equation, for every user k

$$R_k(i, j, s) = wR_k(i, j, s) + (1 - w)\bar{R}_k(s) \quad (3)$$

where w is a weighting factor used to average the data rate of user k .

Step 8: A check is then conducted to determine if there are sufficient resources remaining that can be allocated in the current round. If there are, the algorithm returns to Step 4; otherwise, it returns to Step 1 for preparing for the start of the next round.

A flow chart of the proposed algorithm is also presented in Figure 2. It should be note that the proposed algorithm is employed in the BS for operation.

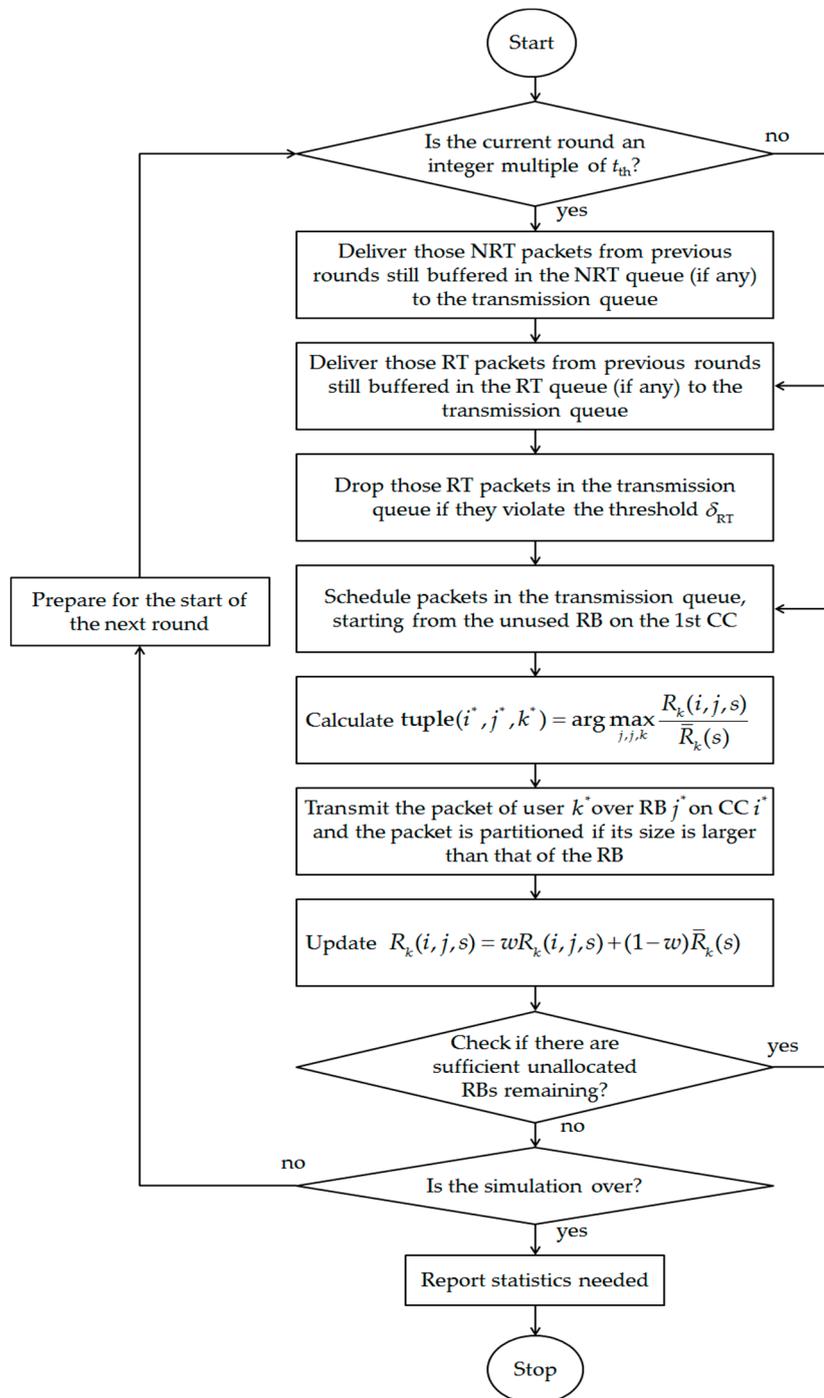


Figure 2. Flow chart of the proposed algorithm.

Figure 3 shows a typical transmission scenario for the algorithm in a case of a partitioned packet; more specifically, the illustrated case involves six users and two CCs aggregated in a BS. It is assumed that $\eta = b_1/b_{total}$. It is then further assumed that user-5 is an NRT user, while the other users are

RT users. Thus, according to our definition, only the 1st CC can be used by user-5 for transmission, whereas the RT users can use both CCs for parallel transmissions. However, since the size of the packet for user-5 is larger than that of the considered RB, it is partitioned during its first transmission. The remaining portion of this packet is then provided a RB for transmission during the next round. It should be noted that all those packets scheduled in the current round illustrated in Figure 3 would have already been buffered in the transmission queue illustrated in Figure 1.

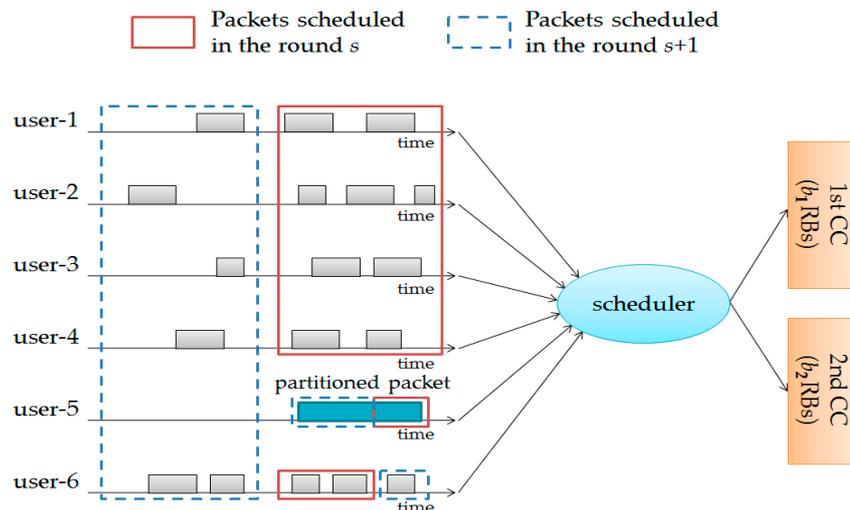


Figure 3. An example of downlink transmissions in a case involving six users and two CCs aggregated in a base station (BS). More specifically, both RT packets (for user-1, user-2, user-3, user-4 and user-6) and NRT packets (user-5) are included in the cell.

Last, but not least, the proposed algorithm's worst-case performance with regard to the time complexity as a function of the number of CCs c , the number of the i th CC's RBs b_i , with $i = 1, 2, \dots, c$, and the number of users n is provided as follows. The dominant cost is on the operation of the *fairness vector* calculation, which iterates $\sum_{i=1}^c b_i$ times and requires n comparisons in each stage, and thus the worst-case cost can be calculated as $O(n \sum_{i=1}^c b_i)$. It is important to note that, in current environments, the value of c typically ranges from 2–5, with the likelihood of it being equal to or lower than 3 being particularly high. As such, the corresponding level of complexity is also typically low.

3.3. Baseline Scheduler

A comparison of the performance of the proposed scheduler versus that of a baseline scheduler design is now presented in order to demonstrate the advantages provided by the proposed mechanism of simultaneously transmitting a packet via multiple CCs.

For the baseline scheduling algorithm, the key working principle is that the various CCs cannot be simultaneously used for the transmission of a user packet, although they can be used independently. More specifically, each RT user will at first be randomly assigned to only a single specific CC out of all the CCs for the transmission of his/her packets. At the same time, the random assignment of each NRT user to only a single specific CC among all the CCs designated for NRT users under the definition of η will also be made. For the case illustrated in Figure 3, for example, if the baseline algorithm is used, then only the 1st CC can be used for transmissions by all the NRT users. Otherwise, the steps of the baseline algorithm are exactly the same as those listed above for the proposed algorithm.

4. Results and Discussions

This section provides an examination of the long-term throughput, delay, and fairness of the system when using the proposed algorithm, as well as how they compare with the same characteristics when the baseline algorithm is used.

For these purposes, we assume an urban macro-cell environment for which the radius is equal to 1 km, and in which two adjacent CCs in the 2 GHz (gigahertz) frequency band are aggregated within the BS for packet transmissions. The underlying path-loss model selected for this study is the COST 231 (Cooperation in Science and Technology) model [31,32]; this model is widely used for making predictions of the path-loss for mobile wireless systems. The function (in dB)

$$L_{\text{path_loss}} = 69.55 + 26.16\log_{10}f - 13.82\log_{10}h_b - \varphi(h_d) + (44.9 - 6.55\log_{10}h_b)\log_{10}d \quad (4)$$

is used to express the COST 231 path-loss model for an urban macro-cell operating at the frequency f equal to 2000 MHz (megahertz). In Equation (4), d denotes the distance between the BS and the user terminal (in km), h_b denotes the effective BS antenna height (in m), h_d denotes the user-terminal antenna height (in m), and $\varphi(h_d)$ is a correction factor for the user-terminal antenna height, which is given by

$$\varphi(h_d) = 3.2\log_{10}(11.75h_d)^2 - 4.97 \quad (5)$$

In addition, we also assume independent Rayleigh slow-fading channels are employed [29], with those channels being corrupted by additive white Gaussian noises. Related to this, the speed of the fading variation is assumed to be slower than the packet transmission speed.

For each CC, the bandwidth is set at 5 MHz, with each CC also containing 512 subcarriers. The fast Fourier transform size is set at 512. Values of 100 and 1 m are set for h_b and h_d , respectively. The setting for η is equal to 1/2, while that for w is equal to 1/6. A value of 5 msec is set for t_{OFDMA} , with 48 OFDMA symbols, while a simulation time of 10^8 OFDMA frames is also set. For the considered cell, it is assumed that there are eight RT users and eight NRT users, and that the distribution of all of these different users generated is uniform. Additionally, the speed at which every RT user moves is assumed to be equal to 3 km/h, while they are further assumed to move in random directions with a uniform distribution. For the NRT users, meanwhile, it is assumed that they mostly access the network from fixed locations. Values of 3 and 20 rounds are set for t_{th} and δ_{RT} , respectively.

Meanwhile, in order to generate traffic sources for both the RT and NRT packets, the ON-OFF Poisson traffic model [33] is used. More specifically, all of the users' traffic is generated independently according to this ON-OFF model in order to provide simulation results. For the RT and NRT packets, the OFF durations have exponential distributions with means of 0.03 and 0.05 s, respectively, while the ON durations for the RT and NRT packets also have exponential distributions, with respective means of 0.01 and 0.1 s. Under the ON condition, packet sizes with means of 100 and 300 bytes, respectively, and with truncated geometric distributions are generated for the RT and NRT packets. Furthermore, it should be noted that if the size of a generated packet is larger than 1500 bytes, then the packet size will be regenerated.

For the purpose of demonstration, the packet generation rate is allowed to vary. To facilitate later description, the ratio of the total packet arrival rate to the maximum system service rate is used as the definition of the system load. For the considered case, the maximum system service rate, after averaging the service rates under different fading conditions, was about 14.8 Mbps (megabits per second). It should be noted that in this context, the system load is dependent upon the change in the packet generation rate.

4.1. Throughput Comparison

Figure 4 presents a comparison between the proposed algorithm and the baseline algorithm in terms of the system throughput versus the system load. As can be seen in Figure 4, the proposed

algorithm provided better system throughput performance than the baseline algorithm did, particularly when the system load was heavy. For example, the system throughput of the proposed algorithm was about 12% better than that of the baseline algorithm when the system load was equal to 1. This was because by aggregating two CCs for transmission, the proposed scheduler had a better capacity to avoid/skip temporarily faded users when the system load was larger. To put that another way, the proposed scheduler is better able to achieve location (spatial) diversity. In summary, by aggregating multiple CCs for transmission the proposed algorithm provides clear improvements in system throughput, especially when the system load is large.

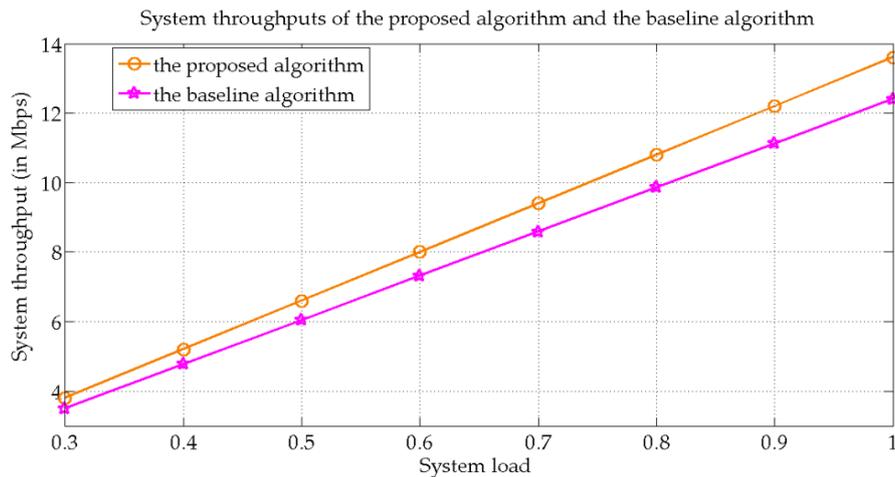


Figure 4. Comparison between the proposed algorithm and the baseline algorithm for the system throughput versus the system load.

Furthermore, that the throughput of the system is significantly better when the proposed algorithm (as opposed to the baseline algorithm) is used suggests that a greater amount of data can be transmitted in an equivalent amount of time, even as the amount of energy consumed remains almost the same, due to the fact that each RB is provided with the same amount of output power for transmission, with that output power being low in comparison to the input power level [24]. Accordingly, the proposed algorithm's efficient design not only allows it to adequately use spectrum resources for the purposes of transmissions, but also gives it greater energy usage efficiency than that provided by the baseline algorithm. With these points in mind, it can be concluded that the proposed algorithm serves as an efficient means by which to provide great spectral efficiency, even as it also addresses the problem of energy efficiency to some degree.

4.2. Mean Packet Delay Comparison

In this section, an examination of the mean packet delay versus the system load is presented. It should be noted that, for these purposes, the measurement of the packet delay starts from the time when a given packet arrives into the classifier shown in Figure 1 and then lasts until the transmission of the packet is fully completed. However, it should be noted that the packet processing time in the classifier itself is assumed to be ignored. Given these conditions, Figures 5 and 6 present comparisons between the proposed algorithm and the baseline algorithm of the mean packet delay for RT and NRT packets, respectively.

As can be seen from both Figures 5 and 6, the proposed algorithm provided better performance than the baseline algorithm in terms of the mean packet delays for both RT and NRT packets. It can also be seen that for both the proposed algorithm and the baseline algorithm, the system performance in terms of the RT packet delay was better than the performance in terms of the NRT packet delay.

This explains why the proper setting of η can effectively protect RT packets from competition by NRT packets.

Furthermore, Figure 5 also shows that there was no monotonic increase in the mean RT packet delay for either the proposed algorithm or the baseline algorithm. Rather, for high system loads, the maximum of the mean RT packet delay was always equal to 0.1 s because the RT packet would be dropped if it violated the threshold of $\delta_{RT} = 20$ rounds (i.e., 0.1 s).

Additionally, Figure 6 further shows that the proposed algorithm yielded a significant overall improvement over the baseline algorithm in terms of the mean NRT packet delay, particularly when the system load was high. This is because under the proposed algorithm, when the mechanism of multiple CCs being aggregated for transmission is enabled, the loads of the RT users can be distributed over the two CCs for parallel transmissions, which in turn provides the NRT users with more opportunities to transmit their packets via the 1st CC.

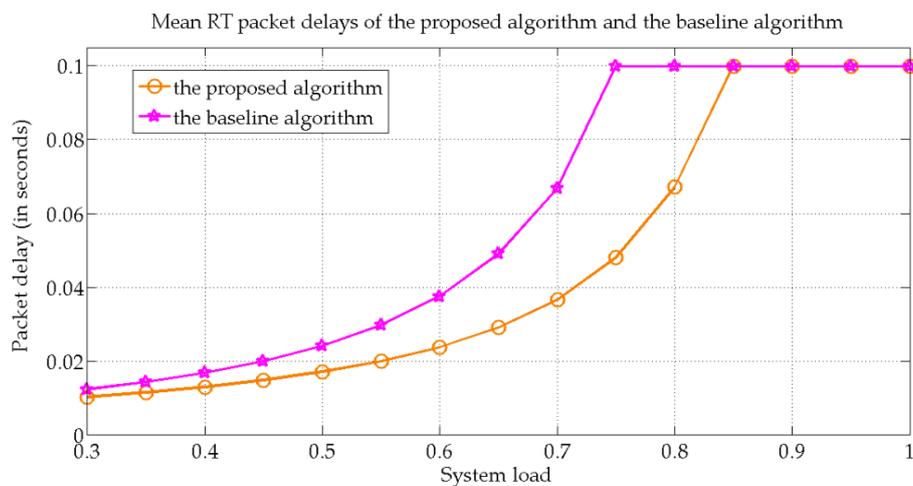


Figure 5. Comparison between the proposed algorithm and the baseline algorithm for the mean RT packet delay versus the system load.

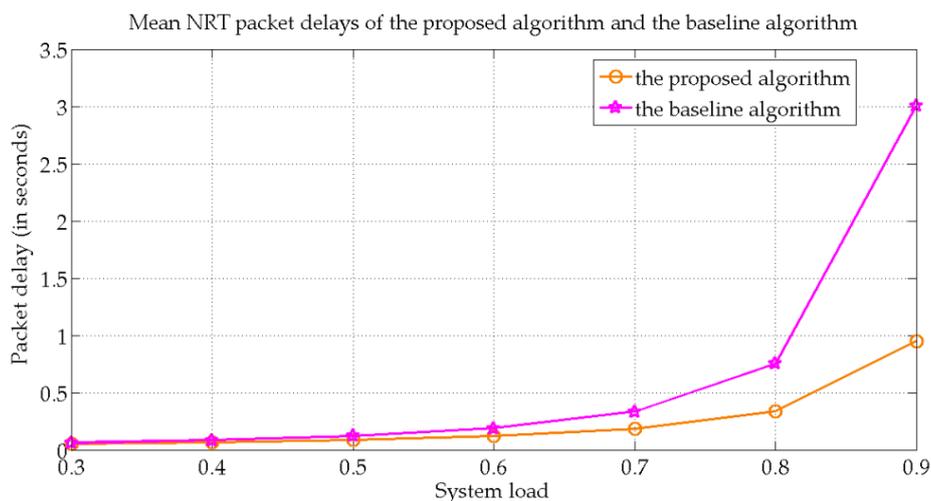


Figure 6. Comparison between the proposed algorithm and the baseline algorithm for the mean NRT packet delay versus the system load.

Finally, as shown in Table 1, we used values of the parameter $\eta = 1/4, 1/2$ and $3/4$, respectively, in order to determine the effects that these different values would have on the mean RT packet delay and mean NRT packet delay when using the proposed algorithm with the system load set at

respective values of 0.3, 0.5, 0.7 and 0.9. The effects of the different values on the mean RT packet delay were observed first. In that regard, the results did not actually differ significantly when different η values were used under the various system loads noted above. This lack of variability in the results presumably resulted from the fact that the number of available RBs that could be reserved for RT packets remained nearly the same under a given system load value even with different values of η . As for the mean NRT packet delay, the differences in the observed results were increasingly significant as the system loads were increased, with the best, middle, and worst performances for any given system load being at $\eta = 3/4, 1/2$ and $1/4$, respectively. This order from best to worst performance was due to the fact that by adjusting $\eta = 1/4$ to $\eta = 3/4$, 50% more resources could be reserved by the system for NRT packets.

Table 1. The mean RT packet delay and mean NRT packet delay when using the proposed algorithm given three different values of η and four different system loads.

Mean Delay (in Seconds)		System Load			
		0.3	0.5	0.7	0.9
RT packets	$\eta = 1/4$	0.010	0.016	0.032	0.1
	$\eta = 1/2$	0.010	0.017	0.037	0.1
	$\eta = 3/4$	0.012	0.020	0.045	0.1
NRT packets	$\eta = 1/4$	0.0624	0.1295	0.3572	1.9830
	$\eta = 1/2$	0.0512	0.0849	0.1826	0.9487
	$\eta = 3/4$	0.0274	0.0337	0.1095	0.7152

4.3. Fairness Comparison

In order to provide a full picture, this section presents a comparison of the fairness of the two algorithms. To that end, we employ the fairness index of R. Jain et al. [34], which is defined as

$$F = \lim_{s \rightarrow \infty} \frac{\left(\sum_{k=1}^n \bar{R}_k(s) \right)^2}{n \sum_{k=1}^n \left(\bar{R}_k(s) \right)^2} \quad (6)$$

in order to quantify the fairness among all users. It should be noted that in Equation (6) the value of $\bar{R}_k(s)$ is updated to the long-term stationary value. Furthermore, all packets violating δ_{RT} are excluded if users in question are included as part of the RT traffic.

In Figure 7, a plot of the fairness index versus the system load is shown. As can be seen, since both the proposed algorithm and the baseline algorithm were designed on the basis of PF, the fairness indexes for both were high. However, the fairness index provided by the baseline algorithm was still slightly lower than that provided by the proposed algorithm. This was presumably because, with the proposed algorithm, each RT user could use all the resources made available by the aggregation of two CCs for transmission, whereas they could only use a single CC under the independent CC mechanism used by the baseline algorithm. In other words, with the independent CC mechanism, the probability of a delay violation with respect to δ_{RT} was higher, which in turn indicated that the variation in $\bar{R}_k(s)$ was larger. As a result, F was smaller. In addition, Figure 7 also indicates that the fairness indexes for both algorithms were almost equal to 1 when the system load was low.

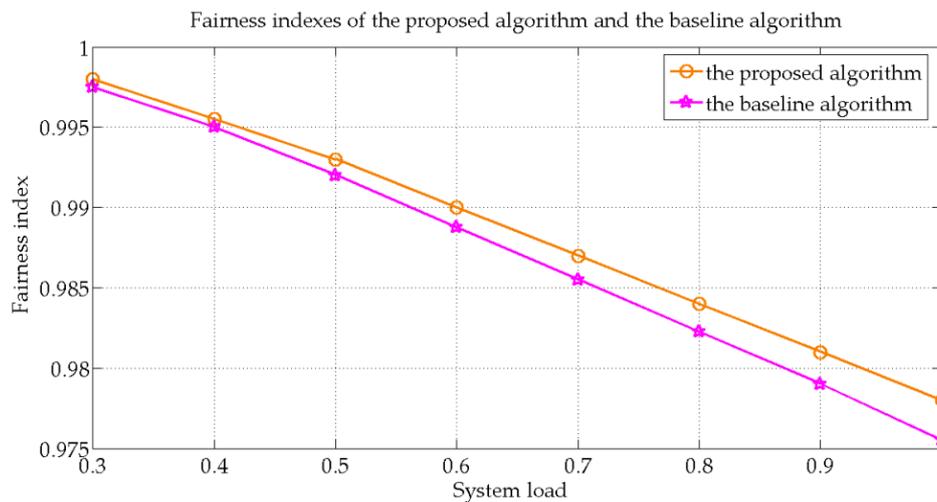


Figure 7. Comparison between the proposed algorithm and the baseline algorithm for the fairness index versus the system load.

5. Conclusions

This paper proposes and presents a novel packet scheduling algorithm, which utilizes the mechanism of multiple CCs aggregated for transmission, for efficient downlink transmissions in 4G/5G cellular systems. The simulation results detailed above indicate that, as long as η and t_{th} are properly adjusted, the proposed scheduler design can cope well and simultaneously with both RT and NRT traffic. In addition, the scheduler also ensures a certain level of fairness among all the users because it is based on the PF criterion. Compared with a baseline algorithm in which multiple CCs are not aggregated for transmission, the proposed algorithm was able to achieve significant improvements in system throughput, mean delay, and fairness performance. In particular, the proposed algorithm provided substantial improvement in terms of the delay performance under the condition of high system loads. Meanwhile, because it flexibly and effectively utilizes all CCs, the proposed packet scheduling algorithm strongly supports the timely delivery of RT traffic and NRT traffic in a simultaneous manner. As a result of this efficiency in its design, the proposed algorithm can also improve overall energy efficiency to some degree. We therefore believe that the proposed scheme constitutes an excellent solution for the packet scheduling of downlink transmissions in multiple-CC cellular systems. Relatedly, we recommend that the cellular industry consider applying the proposed algorithm in developing 5G cellular systems, given the fact that such systems must take both spectral efficiency and energy efficiency into account. In future studies, it would be worthwhile to consider the QoS requirements of various real-world traffic patterns. Further investigation may be required in order to determine how to adjust η and t_{th} in the scheduling process in an adaptive manner based on the fluctuation of RT and NRT packet traffic loads so as to achieve the best overall system performance possible.

Acknowledgments: The author wishes to express his appreciation for the financial support provided by the Ministry of Science and Technology (MOST) of Taiwan, R.O.C., under the Contracts MOST 104-2221-E-019-053-, MOST 104-2218-E-019-004- and MOST 105-2628-E-019-002-MY3.

Conflicts of Interest: The author declares no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Parkvall, S.; Astely, D. The evolution of LTE towards IMT-advanced. *J. Commun.* **2009**, *4*, 146–154. [[CrossRef](#)]

2. Martín-Sacristán, D.; Monserrat, J.F.; Cabrejas-Peñuelas, J.; Calabuig, D.; Garrigas, S.; Cardona, N. On the way towards fourth-generation mobile: 3GPP LTE and LTE-advanced. *EURASIP J. Wirel. Commun. Netw.* **2009**, 2009. [[CrossRef](#)]
3. Andrews, J.G.; Buzzi, S.; Choi, W.; Hanly, S.V.; Lozano, A.; Soong, A.C.K.; Zhang, J.C. What will 5G be? *IEEE J. Sel. Areas Commun.* **2014**, *32*, 1065–1082. [[CrossRef](#)]
4. Wu, J.; Zhang, Y.; Zukerman, M.; Yung, E.K.-N. Energy-efficient base-stations sleep-mode techniques in green cellular networks: A survey. *IEEE Commun. Surv. Tutor.* **2015**, *17*, 803–826. [[CrossRef](#)]
5. Gupta, A.; Jha, R.K. A survey of 5G network: Architecture and emerging technologies. *IEEE Access.* **2015**, *3*, 1206–1232. [[CrossRef](#)]
6. Hossain, E.; Hasan, M. 5G cellular: Key enabling technologies and research challenges. *IEEE Instrum. Meas. Mag.* **2015**, *18*, 11–21. [[CrossRef](#)]
7. Han, F.; Zhao, S.; Zhang, L.; Wu, J. Survey of strategies for switching off base stations in heterogeneous networks for greener 5G systems. *IEEE Access.* **2016**, *4*, 4959–4973. [[CrossRef](#)]
8. Jeong, S.S.; Jeong, D.G.; Jeon, W.S. Cross-layer design of packet scheduling and resource allocation in OFDMA wireless multimedia networks. In Proceedings of the IEEE Vehicular Technology Conference 2006-Spring, Melbourne, Australia, 7–10 May 2006.
9. Madan, R.; Boyd, S.; Lall, S. Fast algorithms for resource allocation in wireless cellular networks. *IEEE/ACM Trans. Netw.* **2010**, *18*, 973–984. [[CrossRef](#)]
10. Ng, D.W.K.; Lo, E.S.; Schober, R. Energy-efficient resource allocation in OFDMA systems with large numbers of base station antennas. *IEEE Trans. Wirel. Commun.* **2012**, *11*, 3292–3304. [[CrossRef](#)]
11. Ng, D.W.K.; Lo, E.S.; Schober, R. Energy-efficient resource allocation in multi-cell OFDMA systems with limited backhaul capacity. *IEEE Trans. Wirel. Commun.* **2012**, *11*, 3618–3631. [[CrossRef](#)]
12. Ng, D.W.K.; Lo, E.S.; Schober, R. Energy-efficient resource allocation in OFDMA systems with hybrid energy harvesting base station. *IEEE Trans. Wirel. Commun.* **2013**, *12*, 3412–3427. [[CrossRef](#)]
13. Morosi, S.; Piunti, P.; del Re, E. Improving cellular network energy efficiency by joint management of sleep mode and transmission power. In Proceedings of the 24th Tyrrhenian International Workshop on Digital Communications-Green Information and Communication Technology, Genova, Italy, 23–25 September 2013.
14. Chung, Y.-L. Energy-saving transmission for green macrocell-small cell systems: A system-level perspective. *IEEE Syst. J.* **2016**, *99*, 1–11. [[CrossRef](#)]
15. Lei, L.; Zheng, K. Performance evaluation of carrier aggregation for elastic traffic in LTE-advanced systems. *IEICE Trans. Commun.* **2009**, *92*, 3516–3519. [[CrossRef](#)]
16. Chen, L.; Chen, W.; Zhang, X.; Yang, D. Analysis and simulation for spectrum aggregation in LTE-advanced system. In Proceedings of the IEEE Vehicular Technology Conference 2009-Fall, Anchorage, AK, USA, 20–23 September 2009.
17. Takeda, K.; Nagata, S.; Kishiyama, Y.; Miki, N.; Tanno, M. Effects of wideband scheduling and radio resource assignment in OFDMA radio access for LTE-advanced downlink. In Proceedings of the IEEE Vehicular Technology Conference 2009-Fall, Anchorage, AK, USA, 20–23 September 2009.
18. Songsong, S.; Chunyan, F.; Caili, G. A resource scheduling algorithm based on user grouping for LTE-advanced system with carrier aggregation. In Proceedings of the International Symposium on Computer Network and Multimedia Technology, Wuhan, China, 18–20 December 2009.
19. Chung, Y.-L.; Tsai, Z. A quantized water-filling packet scheduling scheme for downlink transmissions in LTE-advanced systems with carrier aggregation. In Proceedings of the IEEE Software, Telecommunications and Computer Networks, Split, Croatia, 23–25 September 2010.
20. Wang, H.; Rosa, C.; Pedersen, K.I. Radio resource management for uplink carrier aggregation in LTE-advanced. *EURASIP J. Wirel. Commun. Netw.* **2015**, 2015. [[CrossRef](#)]
21. Sundaresan, K.; Rangarajan, S. Energy efficient carrier aggregation algorithms for next generation cellular networks. In Proceedings of the IEEE International Conference on Network Protocols, Gottingen, Germany, 7–11 October 2013.
22. Chen, H.; Jiang, Y.; Xu, J.; Hu, H. Energy-efficient coordinated scheduling mechanism for cellular communication systems with multiple component carriers. *IEEE J. Sel. Areas Commun.* **2013**, *31*, 959–968. [[CrossRef](#)]
23. Chung, Y.-L. Rate-and-power control based energy-saving transmissions in OFDMA-based multicarrier base station. *IEEE Syst. J.* **2015**, *9*, 578–584. [[CrossRef](#)]

24. Chung, Y.-L. Energy-efficient transmissions for green base stations with a novel carrier activation algorithm: A system-level perspective. *IEEE Syst. J.* **2015**, *9*, 1252–1263. [[CrossRef](#)]
25. Chung, Y.-L. A novel power-saving transmission scheme for multiple-component-carrier cellular systems. *Energies* **2016**, *9*, 265. [[CrossRef](#)]
26. Chung, Y.-L.; Tsai, Z.; Yang, C.-H. A study of quota-based dynamic network selection for multimode terminal users. *IEEE Syst. J.* **2014**, *8*, 759–768. [[CrossRef](#)]
27. Kelly, F.P.; Maulloo, A.; Tan, D. Rate control for communication networks: Shadow prices, proportional fairness and stability. *J. Oper. Res. Soc.* **1998**, *49*, 237–252. [[CrossRef](#)]
28. Lee, S.-B.; Choudhury, S.; Khoshnevis, A.; Xu, S.; Lu, S. Downlink MIMO with frequency-domain packet scheduling for 3GPP LTE. In Proceedings of the IEEE Conference on Computer Communications, Rio de Janeiro, Brazil, 19–25 April 2009.
29. Goldsmith, A. *Wireless Communication*; Cambridge University Press: Cambridge, UK, 2005.
30. Jalali, A.; Padovani, R.; Pankaj, R. Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system. In Proceedings of the IEEE Vehicular Technology Conference 2000-Spring, Tokyo, Japan, 15–18 May 2000.
31. Agrawal, D.P.; Zeng, Q.-A. *Introduction to Wireless & Mobile Systems*, 4th ed.; CENGAGE Learning: Boston, MA, USA, 2016.
32. European Commission. *COST Action 231: Digital Mobile Radio towards Future Generation Systems: Final Report*; European Commission: Brussels, Belgium, 1999.
33. Hayes, J.F.; Babu, T. *Modeling and Analysis of Telecommunication Networks*, 2nd ed.; Wiley-Interscience: Hoboken, NJ, USA, 2004.
34. Jain, R.; Chiu, D.-M.; Hawe, W. A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems. Available online: <http://www.cse.wustl.edu/~jain/papers/ftp/fairness.pdf> (accessed on 1 April 2016).



© 2016 by the author; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).