

Research Article

New Courteous Algorithm for Uplink Scheduling in LTE-Advanced and 5G Networks

Chafika Tata ^{1,2}, Nassima Fellag,² and Michel Kadoch ²

¹Faculty of Science, University Alger 1, Algiers, Algeria

²Department of Electrical Engineering, École de Technologie Supérieure, Montreal, QC, Canada

Correspondence should be addressed to Chafika Tata; cha_aoudia@yahoo.fr

Received 18 March 2019; Revised 22 October 2019; Accepted 3 December 2019; Published 23 January 2020

Academic Editor: Djamel F. H. Sadok

Copyright © 2020 Chafika Tata et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The fast evolution of the number of wireless users and the emergence of new multimedia services have motivated third-generation partnership project (3GPP) to develop new radio access technologies. Thus, the carrier aggregation (CA) was introduced from version 10 long-term evolution (LTE), known as long-term evolution-advanced (LTE-A), to meet the increasing demands in terms of throughput and bandwidth and to ensure the Quality of Service (QoS) for different classes of bearers in LTE networks. However, such solution stills inefficient until implementing good resources management scheme. Several scheduling mechanisms have been proposed in the literature, to guarantee the QoS of different classes of bearers in LTE-A and 5G networks. Nevertheless, most of them promote high-priority bearers. In this study, a new approach of uplink scheduling resources has been developed. It aims to ensure service fairness of different traffic classes that allocates bearers over LTE-A and 5G networks. Also, it raises the number of admitted users in the network by increasing the number of admitted bearers through a dynamic management of service priorities. In fact, the low-priority traffic classes, using low-priority bearers, are favoured during a specific time interval, based on the average waiting time for each class. Simulation results show that the QoS parameters were much improved for the low-priority classes without significantly affecting the QoS of high priority ones.

1. Introduction

Long-term evolution-advanced (LTE-A) and 5th generation (5G) networks have become the most promising telecommunication technologies, thanks to the aggregation of carrier components. This offers a greater bandwidth with a higher throughput, meets the growing demands of mobile traffic, and maximizes the number of admitted users in the network. After accepting a type of the bearer, the evolved Node B (eNodeB) should allocate necessary resource blocks (RBs) to guarantee the requested Quality of Service (QoS) of the admitted bearers. In order to achieve high utilization of radio resources, eNodeB has to perform efficient scheduling scheme, when serving different types of bearers. This way, it could assure a good level of QoS for each bearer's class. Certainly, the throughput delays and blocked calls are one of the most important measurements of the level of QoS in the network. Therefore, bandwidth increasing stills to be an insufficient solution for maintaining a good QoS level in the

whole network, especially during congested situations, unless if it deals with the access management of network resources. Several scheduling approaches are proposed in the literature for the uplink and downlink [1–3], focusing on different aspects, such as queuing and priority management. Some research studies consider the delay tolerance in the queue management, where low-tolerant classes have the highest priority of service. The proportional fair algorithm presented in [4] privileges real-time (RT) traffic compared with the nonreal time (NRT) one. However, the solution does not define any priority mechanism to efficiently manage the network resources. Thus, the NRT-accepted packets would be retained in the queue until the transmission of all RT packets. This scheme will cause significant delays for NRT data, which may exceed the tolerated delay and so that could increase the packet loss rate for this kind of traffic. Moreover, during traffic congestion situations, the bearers allocated to NRT traffic will be the firsts to be preempted. In addition, all new network access requests will be

rejected. Consequently, low-priority traffics will suffer from a high packet loss rate and a high level of blocking calls. Furthermore, several resource management schemes promote higher priority calls over lower priority ones and do not implement any priority management mechanism. Henrique da Mata and Guardieiro [5] propose a solution based on genetic algorithm to assure the QoS-Guaranteed Bit Rate (GBR) traffic, but causing lower pressure on non-GBR application. The approach presented in [6] is based on resource reservation to reduce the blocking probability of handoff calls, especially during the congestion situations. As well, the proposed approach in [7] introduces a mechanism that keeps the blocking probability of handoff calls lower than the blocking probability of new ones, when network resources are lacking. The solution given in [8] improves the QoS for low-priority traffics by restraining the number of accepted new calls when their admitted threshold number is met. In this last approach, the blocking probability of handoff calls and new calls are selected based on threshold values. The solution in [9] deals with the priority management. This is introduced to improve the resource network allocation. However, when the cell is overloaded, LTE-A users will be blocked before the ones holding the highest priority. The approach proposed in [10] deals with the resource allocation in multicell D2D communications by using Games theory and by considering the whole resources existing within all cells.

Besides, our paper proposes a new approach of uplink-scheduling resources. It aims to ensure service fairness for different traffic classes allocating different bearers over LTE-A and 5G networks. Also, it increases the number of admitted users in the network by increasing the number of admitted bearers through a dynamic management of service priorities. In fact, the low-priority traffic classes, using low-priority bearers, are favoured during a specific time interval, based on the average waiting time for each class.

This paper is organized as follows. Section 2 contains materials and methods; it explains how the new courteous algorithm contributes to improve the scheduling management system in LTE-A and 5G networks, based on the native courteous algorithm [11]. It also presents the algorithm developed in this study and describes the mathematical model of our new scheme. Finally, it shows how our algorithm is applied in LTE-A and 5G networks. Section 3 illustrates results of our experiments. We conclude with a conclusion.

2. Materials and Methods

2.1. The New Courteous Algorithm Compared with the Native Courteous Algorithm. The native courteous scheduling network resources algorithm was developed in [11]. It is designed to improve the network performance in WiMAX networks. It proposes a fair management of resources and optimizes the service of low-priority traffic by transmitting their packets instead of high priority ones. This is only possible when the QoS of the high-priority traffic is acceptable. Applying this approach had reduced the average waiting time of low-priority packets and minimized the

packets loss for their classes. Although, this solution has improved the transmission service for disadvantaged classes, this scheme does not bring improvement in the high-priority classes and does not improve the throughput and blocking probability rate in the system. In addition, the native courteous algorithm does not deal with the handoff's traffics because it has been used for fixed Wimax networks. Thus, a new approach for scheduling mechanism in the LTE-A uplink and 5G networks has been developed in this paper, namely, the new courteous algorithm for uplink scheduling in LTE-advanced networks to resources management depending on the three traffic classes, handoff, RT, and NRT in order to ensure the service fairness to different classes, even in mobile situations. This solution is suitable for both moderate load traffic and network congestion situations. It increases the number of satisfied customers and reduces the average waiting time of the underprivileged traffic classes without deteriorating the QoS of high-priority classes. Furthermore, it minimizes blocking and packet loss rates in the system, thus improving the throughput.

2.2. New Courteous Algorithm. This approach is based on a scheme, which considers three classes of traffic, namely, handoff traffic, new RT, and NRT calls. The handoff class has the highest priority and is always served first, to avoid the pre-emption of bearers that are already established, when a congestion situation occurred in the system. RT class will get the second priority, and the NRT class will obtain the lowest one. The new courteous algorithm is applied in order to reorganize the management access to network resources and temporarily assign the highest priority to the two other classes for a limited and defined duration, when they suffer from a congestion situation. This should guarantee fairness of service for the whole packets in the system. Note that the native and the new courteous algorithms are based on offering a high priority to the lowest priority traffic instead of the high-priority class, when the QoS of this one is good. This will occur during a time, called tolerance time, and predefined by the courteous and the new courteous algorithms. This time depends to the Quality of Service threshold time of the high-priority traffic. In addition, it ensures that during courteous mechanism application, the high-priority traffic will not reach the critical threshold of packet loss and will always guarantee an acceptable level of QoS.

Even if the application conditions of the two courteous algorithms converge, it is important to indicate that the two algorithms are not alike, in terms of structures and instructions. In fact, regarding of our solution, two concepts have been developed. In the first design, the RT class is given the highest priority when the average waiting time of RT packet is reached. However, RT packet transmission period should not exceed the average waiting time of handoff packets to maintain the good QoS level of handoff class. In the second design, the period dedicated to the transmission of RT packets in the first design is shared between RT and NRT classes. NRT packets take the highest

priority when the average waiting time divided by a chosen x value is reached.

2.2.1. Conditions of Application of New Courteous Algorithm

Condition 1. The first condition is used for defining an initial sequence of service for the different classes, namely, Handoff, RT, and NRT, by assigning successively the priorities PR_1 , PR_2 , and PR_3 as follows:

$$PR_1 > PR_2 > PR_3. \quad (1)$$

Condition 2. The handoff packets are served first for a period T_{r1} , less than or equal to a time, called threshold RT (Th_{RT}), which is the average waiting time in the RT queue. The value of Th_{RT} is justified by the possibility of reducing RT packet loss due to the waiting time in the RT queue, which exceeds the maximum waiting time allowed. Then, if it is possible, and if handoff class tolerates it, from $t = Th_{RT}$, the RT packets will get high priority during the period of transmission T_{r2} equivalent to the average waiting time of handoff packets that will regain the highest priority at time $t = Th_{HD} + Th_{RT}$, where Th_{HD} is threshold of handoff class (Figure 1):

$$Th_{RT} \leq T_{r2} \leq Th_{HD} + Th_{RT}. \quad (2)$$

Condition 3. The period of transmission T_{r2} is shared between the RT and NRT packets, and the NRT packets get high priority when the average waiting time in the queue NRT reaches threshold NRT or Th_{NRT} , which is equal to the average waiting time of NRT packet divided by a chosen value x . The aim of this division is to have a value of Th_{NRT} close to Th_{HD} and Th_{RT} since the NRT class tolerates higher delays. Thus, RT packets get the highest priority during a time T_{r3} less than or equal to Th_{NRT} within the interval T_{r2} , and the NRT packets have priority during T_{r4} (Figure 1):

$$Th_{NRT} \leq T_{r4} \leq Th_{HD} + Th_{RT}. \quad (3)$$

2.2.2. The New Courteous Algorithm Description. The new courteous algorithm is applied in order to consecutively assign the highest priority to RT and NRT traffics during tolerance time of handoff class and during RT and NRT congestions. When the threshold waiting time of RT packets is reached, RT traffic gets highest priority and will be served according to the Priority Queuing (PQ) scheme, while the handoff and NRT classes will be served according to Class-Based Weighted Fair Queuing (CBWFQ). Similarly, when the threshold waiting time of NRT packets is reached, the NRT class will be served according to PQ, while the two other classes will be served according to CBWFQ. The whole added service time of both RT and NRT classes according to PQ must not result in reaching the handoff waiting time threshold, and this constraint assures keeping a good level of QoS for the handoff traffic.

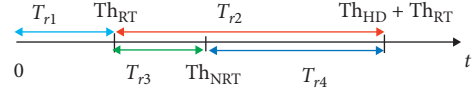


FIGURE 1: Packets NRT service interval following PQ.

The structure of the new courteous algorithm is given below, and the distinction between the different priorities is illustrated through the weights assigned to each class, and the weight given to the handoff class in the basic scenario is equal to 1. The basic scenario or the basic scheme is the one where no courtesy is applied for scheduling packets.

Step 1: packet transmission beginning.

At this stage, we suppose that no congestion occurred. The system will then serve the packets following the basic scheme. In other words, the handoff traffic will be served according to PQ politic, and new RT and new NRT traffics will be served following CBWFQ queuing system. So, the new courteous algorithm applies the following instructions:

- (1) Attribution of the different weights for the queues as follows:
 - (a) $w_1 + w_2 = 1$ with $w_1 > w_2$
 - (b) $w_{rt} = w_1$
 - (c) $w_{nrt} = w_2$
- (2) Determination of the generating matrix Q
- (3) Calculation of P_b , the blocking probabilities
- (4) Checking system stationarity
 - If $P_b \leq 1$ then
 - Stationary system
 - Else
 - Give other values for w_1 and w_2
 - Repeat instructions 3 and 4 of Step 1
 - End if

Step 2: courteous scheme for new RT packets

- (5) Calculating _Thresholds (Th_{HD} , Th_{RT} , Th_{NRT})
- (6) Calculating _ Durations (T_{r1} , T_{r2})
- (7) While ($T_{r1} > Th_{RT}$ and $Th_{RT} \leq T_{r2} \leq Th_{HD} + Th_{RT}$) do
 - (a) Attribution of new weights w'_1 and w'_2
 - (i) $w'_1 = 1$
 - (ii) $w'_1 + w'_2 = 1$
 - (iii) $w'_1 > w'_2$
 - (iv) $w'_{hd} = w'_1$
 - (v) $w'_{nrt} = w'_2$
 - (b) Determining the generating matrix Q
 - (c) Calculation of P_b , the blocking probabilities
 - (d) Checking system stationarity
 - If $P_b \leq 1$ then
 - Stationary system
 - Else
 - Give other values for w'_1 and w'_2
 - Repeat instructions 7(c) and 7(d) of Step 2

```

    End if
(8) End of While
(9) Calculating Duration ( $T_{r3}$ )
(10) While ( $T_{r3} > TH_{NRT}$  and  $TH_{RT} \leq T_{r2} \leq TH_{HD} + TH_{RT}$ ) do
    (a) Attribution of new weights  $w_0''$  and  $w_2''$ 
        (i)  $w_2 = 1$ 
        (ii)  $w_0'' + w_2'' = 1$ 
        (iii)  $w_0'' > w_2''$ 
        (iv)  $w_{hd} = w_0'$ 
        (v)  $w_{rt} = w_2'$ 
    (b) Determining the generating matrix Q
    (c) Calculation of Pb, the blocking probabilities
    (d) Checking system stationarity
        If  $Pb \leq 1$  then
            Stationary system
        Else
            Give other values for  $w_0'$  and  $w_2'$ 
            Repeat instructions 10(c) and 10(d) of Step 2
        End if
(11) End of While
Step 3:
    If there are any packets to be transmitted
        Go to Step 1
    Else
        Go to Step 4
    End If
Step 4: end of transmission

```

The description of the new courteous algorithm shows that the courteous scheme is applied only if the lowest priority traffic class is suffering from a long waiting time that exceeds the threshold for this class and only if the QoS of the higher priorities classes is guaranteed. These two constraints are explained by the condition instruction of the While loop. In addition, once the advantaged class of courteous mechanism reaches an acceptable waiting time, the courteous scheduling system will stop the courteous scheme transmission for that class. Thereafter, the algorithm checks whether there are any packets to be transmitted in the queuing system; if so, the algorithm returns to the first step; otherwise, the execution of the new courteous algorithm halts.

2.3. Mathematical Model of the New Courteous Scheduling System. This section describes the mathematical model we developed corresponding to the new courteous algorithm, which improves the network resources scheduling over the LTE-A and 5G networks. Note that, in this study, we call a basic scheme a classical scheduling system without applying any courteous mechanism. This scheme is considered when defining the mathematical model and when simulations are performed.

2.3.1. Description of System M/M/S/K with PQ_CBWFQ. The model shown in Figure 2 consists of S similar servers and a unique buffer with the capacity K , shared between three virtual waiting queues: Q_{HD} , Q_{RT} , and Q_{NRT} related to handoff traffic and new traffics RT and NRT, respectively. The handoff class is served according to non-pre-emptive Priority Queuing (PQ) policy, which treats the handoff packets without interruption of service when a new packet RT or NRT is being transmitted. The RT and NRT classes are served following CBWFQ with the weight w_1 for Q_{RT} greater than the weight w_2 of Q_{NRT} which means, the quantity of bandwidth dedicated to RT traffic is the highest [12]. For our scheme, as in [13, 14, 15], the calls arriving, in the three queues, follow Poisson distribution of average λ_p , $p = 0, 1, 2$, and require exponential service time $1/s\mu$ with $\mu = \mu_p$, $w_1\mu_1$ or $w_2\mu_2$. Note that this is the number of available servers which varies from 1 to the maximum S value and must be less or equal to the capacity of the buffer K . The service conservation policy is considered in this study where the server may not be in the idle mode (IDLE) as long as there are packets awaiting in queues.

In addition, only one server serves a single packet [16]. Furthermore, the arrivals in the queue Q_{HD} consist of λ_{01} and λ_{02} , where λ_{01} and λ_{02} indicate successively the arrivals of RT handoff packets and NRT handoff packets. The management within the handoff queue follows the CBWFQ policy and handoff packets RT always have the highest priority.

2.3.2. Analytical Model of M/M/S/K System with PQ_CBWFQ. Figure 3 shows the steady state diagram of queuing system M/M/S/K with PQ_and, CBWFQ which is modeled as a three-dimensional discrete Markov chain in continuous time with the state (i, j, k) , where i, j, k denote, respectively, the number of packets in each state of the handoff, RT, and NRT queues. The number of packets existing simultaneously in the three queues must not exceed the total capacity of the buffer K . Packets arriving after reaching the capacity K are blocked. The handoff class has the highest priority and is served with the $s\mu_0$ rate following the PQ politic, with $s = 0, 1, \dots, S$. The approach CBWFQ is applied to serve RT and NRT packets. When Q_{NRT} is empty, the RT packets are treated with the $s\mu_1$ rate. However, in the opposite case, the NRT packets are served with the $s\mu_2$ rate. Otherwise, if the packets exist in both the queues at the same time, the link service rate will be shared between the two classes based on weight w_1 and w_2 by giving the highest weight w_1 to RT class. The resolution of the Markov chain requires to first establishing the set of equations for a steady state obtained by matching the incoming and outgoing traffic in each state. Then, the matrix geometric process is applied in order to calculate the steady state probabilities $\pi(i, j, k)$. The matrix geometric process is a mechanism based on the principle of grouping statements in levels according to the value of i and allows the subsequent transitions:

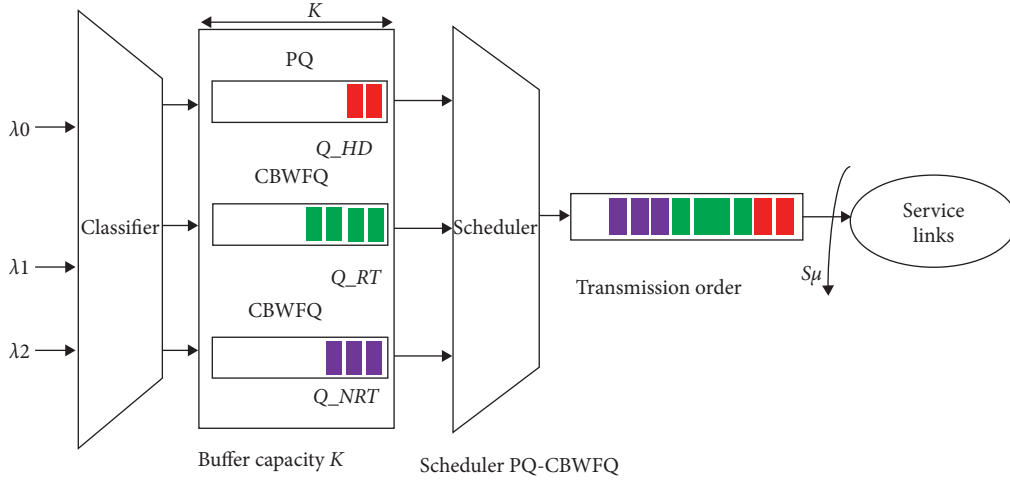
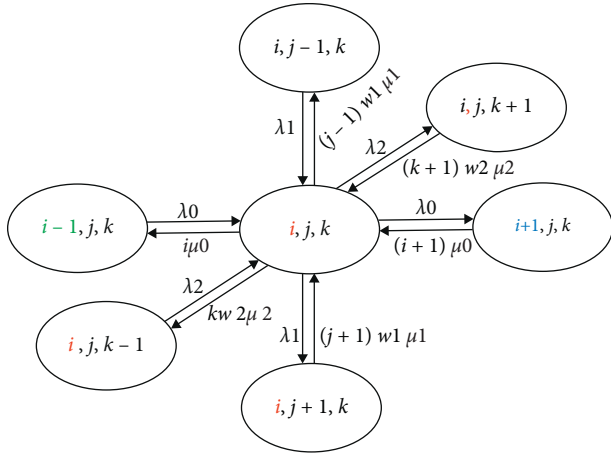


FIGURE 2: System diagram M/M/S/K with PQ-CBWFQ.



- (i) In the same level i , from (i, j, k) to $(i, j+1, k)$, $(i, j-1, k)$, $(i, j, k+1)$, or $(i, j, k-1)$
- (ii) Towards a state of a higher level, from (i, j, k) to $(i+1, j, k)$
- (iii) Towards a state of a lower adjacent level, from (i, j, k) to $(i-1, j, k)$

As the Markov chain is three-dimensional, level i is two-dimensional as shown in Figure 4 [17].

The matrix geometric of the Markov chain M/M/S/K with PQ-CBWFQ is

$$Q = \begin{bmatrix} Q_{0,0} & Q_{0,1} & \bar{0} & \dots & \dots & \bar{0} \\ Q_{1,0} & Q_{1,1} & Q_{1,2} & \dots & \dots & \bar{0} \\ \bar{0} & Q_{2,1} & Q_{2,2} & Q_{2,3} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & Q_{k-1,k-2} & Q_{k-1,k-1} & Q_{k-1,k} \\ \bar{0} & \dots & \ddots & \bar{0} & Q_{k,k-1} & Q_{k,k} \end{bmatrix}_{(j,k)}, \quad (4)$$

where $\bar{0}$ is the matrix zero. The matrices $Q_{i,i-1}$, $Q_{i,i+1}$ are rectangular matrices, while the matrixes $Q_{i,i}$ along the diagonal are nonsingular square ones having the same form as the matrix Q .

The infinitesimal generator of Markov's process is given by

$$Q(i, j, k)(i+1, j, k) = \lambda_0,$$

$$Q(i, j, k)(i-1, j, k) = i\mu_0, \quad i \geq 1,$$

$$Q(i, j, k)(i, j, k) = -\left(\sum \lambda_p + i\mu_0 + jw_1\mu_1 + kw_2\mu_2\right), \quad i+j+k < k,$$

$$Q(i, j, k)(i, j, k) = -\left(\sum \lambda_p + i\mu_0 + j\mu_1\right), \quad k = 0,$$

$$Q(i, j, k)(i, j, k) = -\left(\sum \lambda_p + i\mu_0 + k\mu_2\right), \quad j = 0,$$

$$Q(i, j, k)(i, j, k) = -(i\mu_0 + jw_1\mu_1 + kw_2\mu_2), \quad i+j+k = k,$$

$$Q(i, j, k)(i, j, k+1) = \lambda_2,$$

$$Q(i, j, k)(i, j, k-1) = k\mu_2, \quad j = 0,$$

$$Q(i, j, k)(i, j, k-1) = kw_2\mu_2, \quad j \neq 0,$$

$$Q(i, j, k)(i, j+1, k) = \lambda_1,$$

$$Q(i, j, k)(i, j-1, k) = j\mu_1, \quad k = 0,$$

$$Q(i, j, k)(i, j-1, k) = jw_1\mu_1, \quad k \neq 0,$$

(5)

The steady-state probabilities $\pi(i, j, k)$ can be estimated by solving the normalization formula $\sum \pi_i = 1$ with $\pi Q = 0$, where π is a probability vector partitioned into subvectors π_i , $\pi = [\pi_1, \pi_2, \pi_3, \dots, \pi_k]$ [18].

The resolution of $\pi Q = 0$ leads to the system of formulas as follows:

$$\begin{aligned} \pi_0 Q_{0,0} + \pi_1 Q_{1,0} &= 0, \\ \pi_0 Q_{0,1} + \pi_1 Q_{1,1} + \pi_2 Q_{2,1} &= 0, \\ \pi_0 Q_{1,2} + \pi_1 Q_{2,2} + \pi_2 Q_{3,2} &= 0, \\ \pi_{i-2} Q_{i-2,i-1} + \pi_{i-1} Q_{i-1,i-1} + \pi_i Q_{i,i-1} &= 0, \quad si \ 2 < i \leq K-1, \\ \pi_{K-1} Q_{K-1,K} + \pi_K Q_{K,K} &= 0. \end{aligned}$$

(6)

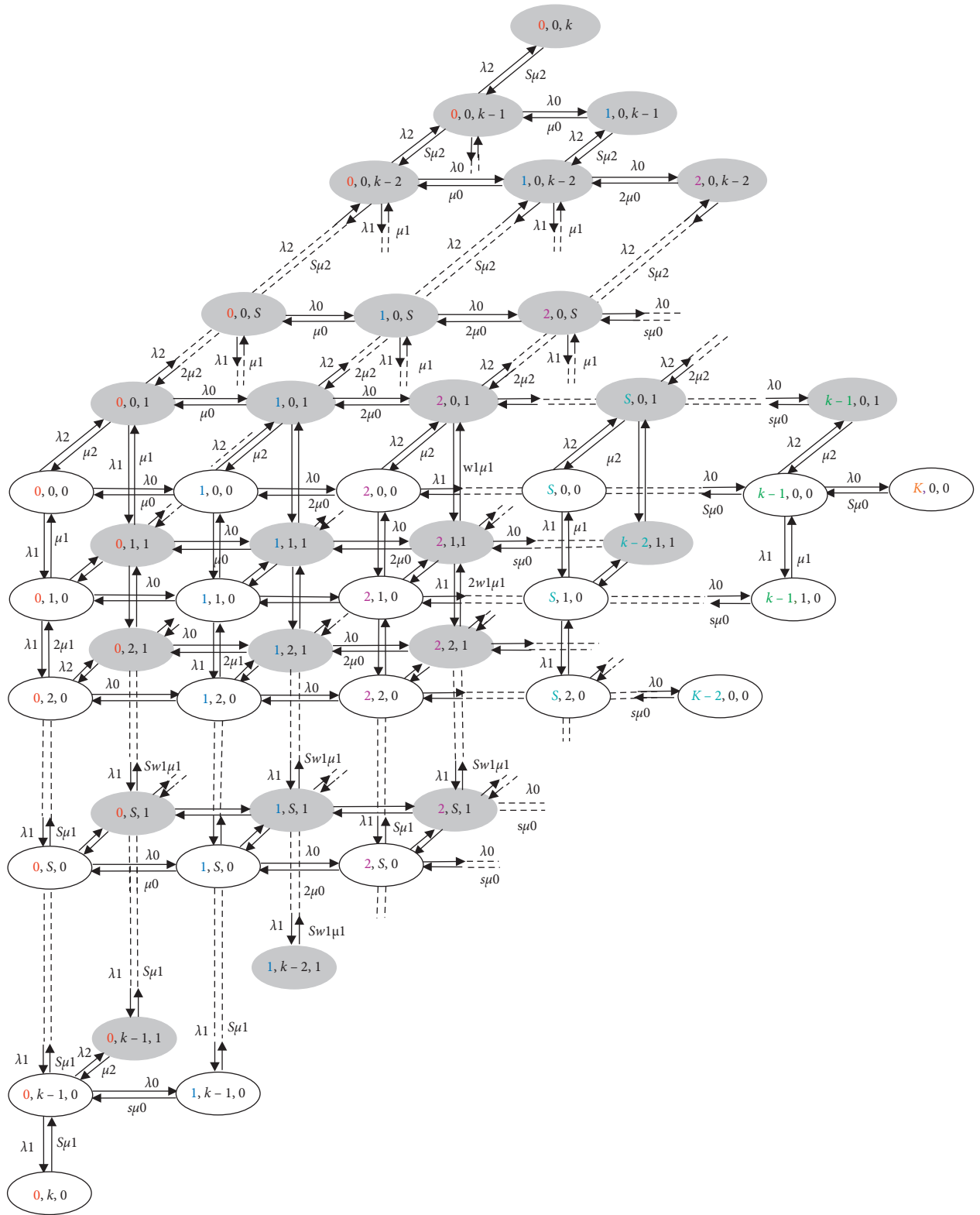
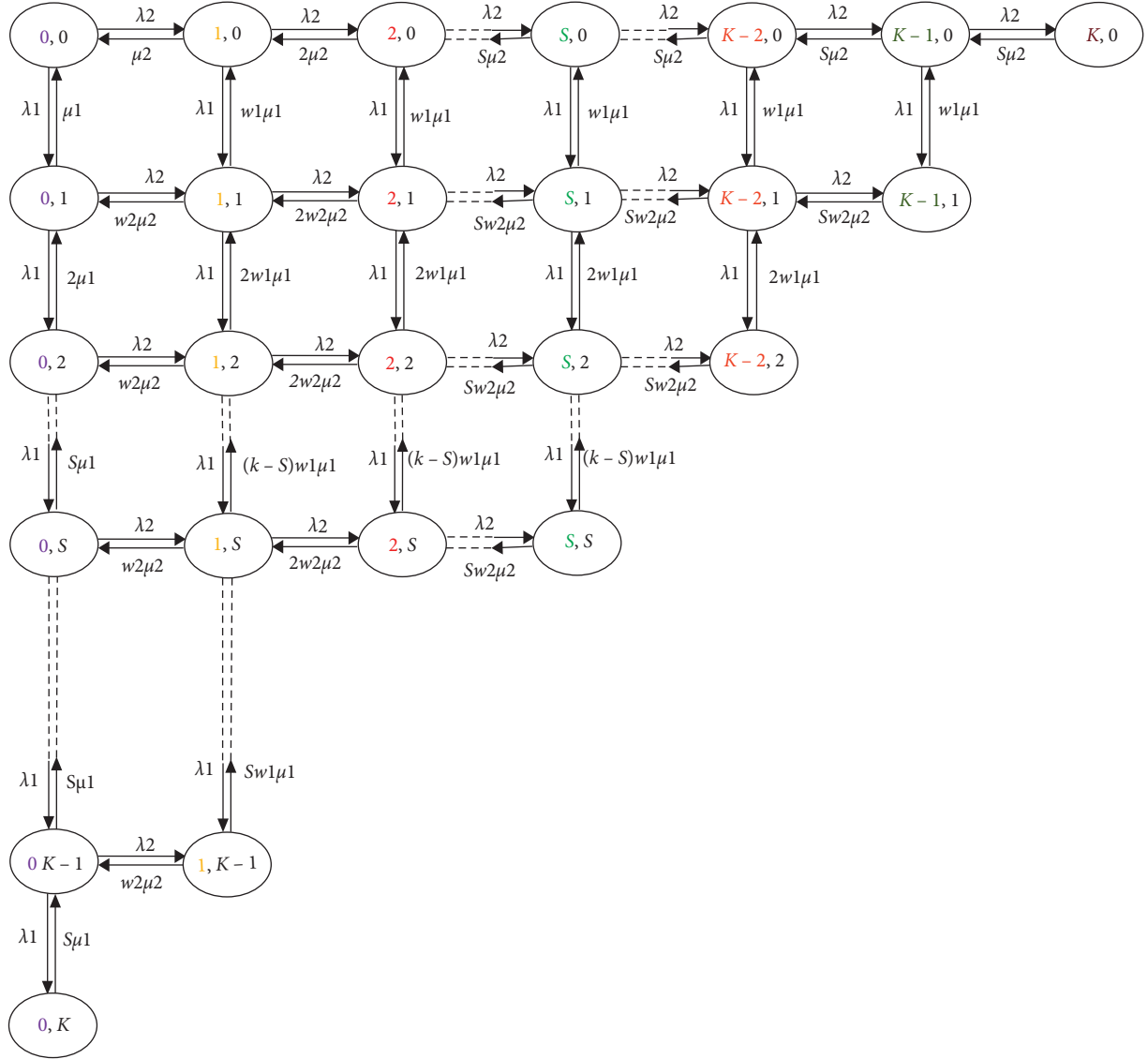


FIGURE 3: Steady-state diagram of queuing system M/M/S/K with PQ_CBWFQ.


 FIGURE 4: Transition diagram of states (j, k) per level (i) .

The formulas of the stationary probabilities can be generalized as follows:

$$\pi_{\xi} = \pi_0 R_{\xi}, \text{ pour } \xi \geq 1. \quad (7)$$

The R_{ξ} matrices are calculated according to the following system:

$$\begin{cases} R_0 = I, \\ R_1 = -Q_{0,0}Q_{0,0^{-1}}, \\ R_{\xi} = -(R_{\xi-2}Q_{\xi-2,\xi-1} + R_{\xi-1}Q_{\xi-1,\xi-1})Q_{\xi,\xi-1^{-1}}, \end{cases} \quad (8)$$

where I is the identity matrix.

Equation (6) can be described as follows:

$$\pi_0 (R_{k-1}Q_{k,k} + R_k Q_{k,k}) = 0. \quad (9)$$

Thus, π_0 is obtained by solving equation (9) with the previous normalization formula expressed in terms of π_0 .

$$\pi_0 \sum_{d=0}^K R_d e_d = 1, \quad (10)$$

where e_d are column vectors of 1.

The estimation of thresholds can be obtained from the calculation of the average waiting times in the three queues.

Thus, Th_{RT} corresponds to the average waiting time of a packet RT in the queue Q_{RT} :

$$\text{Th}_{\text{RT}} = \frac{L_{q1}}{\lambda_1 (1 - \pi^*)}. \quad (11)$$

L_{q1} represents the length of the RT queue:

$$L_{q1} = \sum_{j=0}^K \sum_{i=0}^{K-j} \sum_{k=0}^{K-i-j} j \pi(i, j, k), \quad (12)$$

Th_{HD} is the retention interval of handoff packets in the queue Q_{HD} :

$$\text{Th}_{\text{HD}} = \frac{L_{q0}}{\lambda_0 (1 - \pi^*)}, \quad (13)$$

with π^* is the fraction of arrivals turned away for lack of space in the buffer and π^* is the blocking probability in the system:

$$\pi^* = \pi(i, j, k) \text{ with } i + j + k = K, \quad (14)$$

And, L_{q0} is the length of the handoff queue:

$$L_{q0} = \sum_{i=0}^K \sum_{j=0}^{K-i} \sum_{k=0}^{K-i-j} i\pi(i, j, k). \quad (15)$$

Th_{NRT} is defined by the average waiting time of NRT packets in the queue Q_{NRT} divided by x :

$$\text{Th}_{\text{NRT}} = \frac{L_{q2}}{x\lambda_2 (1 - \pi^*)}. \quad (16)$$

L_{q2} is the queue length of the NRT class, it is given by

$$L_{q2} = \sum_{k=0}^K \sum_{j=0}^{K-j} \sum_{i=0}^{K-j-k} k\pi(i, j, k). \quad (17)$$

The intervals T_{r1} , T_{r2} , and T_{r3} are determined by using the queue analysis of L . Kleinrock of multiple priorities label any packet accessing the system. Then, to estimate the total waiting time w_p in each queue [19, 20],

$$w_p = w_{\text{resp}} + \frac{1}{s\mu} (np + mp), \text{ with } p = 0, 1, 2, \quad (18)$$

w_{resp} : service time of a packet being transmitted since the system is not pre-emptive. $1/s\mu$: average service time of a packet of the priority class PR_p $\mu = \mu_p$, $w_1\mu_1$, or $w_2\mu_2$, and S represents the number of servers. n_p : number of packets belonging to the priority class p , already existing in the queue and receiving a service before the labeled packet. m_p : number of packets of class p , arriving while the tagged packet is waiting and transmitted first. The sum of $n_p + m_p$ can be obtained from the calculation of the queue length for each class.

The service time of a packet being transmitted is given by

$$w_{\text{resp}} = \frac{1}{s\mu} (1 - \pi(0, 0, 0)), \quad (19)$$

where $\pi(0,0,0)$ represents the probability of finding the system IDLE; thus, $(1 - \pi(0,0,0))$ is the probability of finding the system busy. RT packet must wait the remaining service time w_{resp} , adding the service time of the whole handoff packets within Q_{HD} . T_{r1} is calculated as follows:

$$T_{r1} = \frac{1}{\mu_0} \left(\frac{1 - \pi(0, 0, 0)}{i} + \sum_{i=0}^K \sum_{j=0}^{K-i} \sum_{k=0}^{K-i-j} \pi(i, j, k) \right). \quad (20)$$

When PQ is applied to RT traffic, a handoff packet has to wait for the service of RT packets, which is given by the following formula:

$$T_{r2} = \frac{1}{\mu_1} \sum_{j=0}^K \sum_{i=0}^{K-j} \sum_{k=0}^{K-i-j} \pi(i, j, k). \quad (21)$$

Besides, an NRT packet has to wait the remaining service time w_{resp} , adding to the service time of both handoff and RT packets within Q_{HD} and Q_{RT} queues.

$$T_{r3} = \frac{1 - \pi(0, 0, 0)}{i\mu_0} + \frac{1}{\mu_0} \sum_{i=0}^K \sum_{j=0}^{K-i} \sum_{k=0}^{K-i-j} \pi(i, j, k) + \frac{1}{\mu_1} \sum_{j=0}^K \sum_{i=0}^{K-j} \sum_{k=0}^{K-i-j} \pi(i, j, k). \quad (22)$$

The new matrices generated after applying the new courteous algorithm in the both conceptions have the same form and dimensions as the matrixes obtained in the basic scheme. Nevertheless, the service rates change. In the first conception, the RT class is served with the rate $s\mu_1$, while the handoff and NRT classes share the throughput of the link following the rates w'_0 and w'_2 ; the infinitesimal generator of the Markov process related to the first conception is

$$\begin{aligned} Q(i, j, k)(i + 1, j, k) &= \lambda_0, \\ Q(i, j, k)(i - 1, j, k) &= i\mu_0, \quad k = 0, \\ Q(i, j, k)(i - 1, j, k) &= iw_0\mu_0, \quad k \neq 0, \\ Q(i, j, k)(i, j, k) &= -\left(\sum \lambda_p + iw_0\mu_0 + j\mu_1 + kw_2k\mu_2\right), \quad i + j + k < K, \\ Q(i, j, k)(i, j, k) &= -\left(\sum \lambda_p + i\mu_0 + j\mu_1\right), \quad k = 0, \\ Q(i, j, k)(i, j, k) &= -\left(\sum \lambda_p + iw_0i\mu_0 + kw_2k\mu_2\right), \quad j = 0, \\ Q(i, j, k)(i, j, k) &= -\left(\sum \lambda_p + j\mu_1 + k\mu_2\right), \quad i = 0, \\ Q(i, j, k)(i, j, k) &= -(iw_0\mu_0 + j\mu_1 + kw_2k\mu_2), \quad i + j + k = K, \\ Q(i, j, k)(i, j, k + 1) &= \lambda_2, \\ Q(i, j, k)(i, j, k - 1) &= k\mu_2, \quad i = 0, \\ Q(i, j, k)(i, j, k - 1) &= kw_2\mu_2, \quad i \neq 0, \\ Q(i, j, k)(i, j + 1, k) &= \lambda_1, \\ Q(i, j, k)(i, j - 1, k) &= j\mu_1. \end{aligned} \quad (23)$$

In the second conception, NRT class is served with μ_2 rate during the interval T_{r4} . While handoff and RT classes share the throughput of the link according to the rates w_0'' and w_1'' . The infinitesimal generator of the Markov process of the second conception is

$$\begin{aligned}
Q(i, j, k)(i+1, j, k) &= \lambda_0, \\
Q(i, j, k)(i-1, j, k) &= i\mu_0, \quad j=0, \\
Q(i, j, k)(i-1, j, k) &= iw_0\mu_0, \quad j \neq 0, \\
Q(i, j, k)(i, j, k) &= -\left(\sum \lambda_p + iw_0\mu_0 + jw_1\mu_1 + k\mu_2\right), \quad i+j+k < K, \\
Q(i, j, k)(i, j, k) &= -\left(\sum \lambda_p + iw_0\mu_0 + jw_1\mu_1\right), \quad k=0, \\
Q(i, j, k)(i, j, k) &= -\left(\sum \lambda_p + i\mu_0 + k\mu_2\right), \quad j=0, \\
Q(i, j, k)(i, j, k) &= -\left(\sum \lambda_p + i\mu_1 + k\mu_2\right), \quad i=0, \\
Q(i, j, k)(i, j, k) &= -(iw_0\mu_0 + jw_1\mu_0\mu_1 + k\mu_2), \quad i+j+k = K, \\
Q(i, j, k)(i, j, k+1) &= \lambda_2, \\
Q(i, j, k)(i, j, k-1) &= k\mu_2, \\
Q(i, j, k)(i, j+1, k) &= \lambda_1, \\
Q(i, j, k)(i, j-1, k) &= j\mu_1, \quad i=0. \\
Q(i, j, k)(i, j-1, k) &= jw_1\mu_1, \quad i=0.
\end{aligned} \tag{24}$$

2.4. New-Courteous Algorithm Adaptation to the LTE-Advanced Technology. As mentioned above, in LTE-A and 5G networks, after accepting a type of bearers, the eNodeB has to allocate required resource blocks (RBs) to guarantee the requested QoS of the admitted bearers. In order to achieve high utilization of the radio resource, eNodeB also must perform efficient scheduling of the radio resources between different types of bearers to assure a good level of QoS for each bearer's class. Certainly, the throughput, delays, and blocked users are one of the most important measurements of the level of QoS in the network. An efficient management of the radio resources in the cell improves the QoS of the different bearers. Our goal is to assure fairness in RBs management, and we aim to offer some RBs of the high-priority bearers to the low-priority bearers if their QoS is affected.

Otherwise, LTE-A and 5G networks offer differentiated services according to the priorities assigned to the different types of traffic. Prioritization is based on a set of criteria, such as length queue tolerance or whether traffic is new or transferred from a source eNodeB to a target one. Usually, RT traffic has priority over NRT traffic, and handoff calls are preferred over new calls, resulting in high blocking and packet loss rates experienced by low-priority traffic, especially in the case of network congestion.

As a result, the new courteous algorithm is developed in order to ensure the fairness of service for the various traffic classes, through a new priority management. This one allows less-priority traffic to be privileged, when the Quality of Service of high-priority traffic is not affected. It also aims to optimize throughput in the system, minimize blocking and packet loss, and reduce the waiting time for disadvantaged classes. This is possible by adapting a new queuing system that manages the resource blocks by applying the new

courteous algorithm developed in this paper. In fact, when the bearers are admitted in the 5G networks and LTE-A cell, by applying any admission control mechanism, the new courteous queuing system will be used to improve the RBs scheduling by assuring a fairness resources allocation, even for the low-priority traffic classes.

The application of the new courteous algorithm will be done by considering the bearers admission mechanism. In fact, we consider the bearers which are established for assuring the handoff calls, the bearers depending of the new RT calls, and the ones regarding the new NRT calls. First, each class of bearers will allocate a quantity of RBs depending on their needs. After that, the new courteous algorithm will be used for achieving our aims in terms of uplink scheduling.

3. Results and Discussion

The simulation model consists of an LTE cell bandwidth equal to 80 MHz established after aggregating 4 carrier components LTE of 20 MHz, each one using the modulation technique single-carrier frequency division multiple access (SC-FDMA) in the uplink. In addition, each carrier component has a throughput of 50 Mbit/s, hence an aggregated throughput of 200 Mbits/s. The servers considered are a group of carriers where each server contains three carrier components represented by 12 servers. Otherwise, the experiments consider a throughput of 17 Mbit/s per server. Besides, the performance evaluation is made by using MATLAB language.

In this work, three simulation scenarios were performed. The first scenario is related to the simulation of the scheduling scheme without applying any courteous algorithm; this is known as the basic scenario, where the handoff traffic will have the higher priority and will be served according to the PQ politic. The second scenario, called the first conception, implements the case of applying the new courteous algorithm for the RT traffic only, which gives the higher priority to this traffic. Finally, the third scenario, named the second conception, applies the new courteous scheme to RT traffic and to the NRT traffic, so the RT and NRT packets will be served according to PQ system in different periods. It means that the RT will take the higher priority for a while, and after that it offers its higher priority to NRT traffic class, when the average waiting time of this last class exceeds the tolerated threshold. Note that for the three scenarios, the PQ mechanism is always used for the higher priority traffic; therefore, the two other traffics will be served according to the CBWFQ politic.

In addition, to illustrate the new courteous scheme in our solution, we adopt a dynamic queue management system, which means that the server changes policy of the queue management according to the congestion situations of RT and NRT traffics. During congested situations, handoff applies PQ, and RT and NRT apply CBWFQ, which is the basic scenario. In the case where RT packets suffer from a congestion transmission, but not handoff ones, RT will then get the priority of the handoff class. This will occur during the tolerance time, which we call the first conception of

courtesy. Otherwise, if NRT traffic suffers from a congestion, but not RT and handoff traffics, it is NRT that will get the handoff priority during the tolerance time. The tolerance time is defined by the duration between the current waiting time and the handoff waiting threshold.

Note that RT traffic is illustrated by the VoLTE traffic and NRT traffic by the FTP traffic. The service rates μ_p are maintained unchanged during the simulation scenarios. Otherwise, the values of interarrivals are increasing by λ_p ; thus, one simulation time unit corresponds to one λ_p unit, it is important to mention that congestion was observed from time $t = 2$ s. Table 1 summarizes the simulation parameters used in the three scenarios.

In case where the PQ policy is adopted, the priority class is always served with a rate equal to 1. If CBWFQ is used, the rates of sharing service depends mainly on the stability of the system, and the highest weight is always assigned to the class detaining the highest priority with the condition that the sum of the weights equal to 1; hence, in the basic scheme scenario if $w_1 = 0.6$ and $w_1 + w_2 = 1$, $w_2 = 0.4$.

In the first conception, $w'_0 = 0.9$ and $w'_2 = 0.1$, and in the second design $w_0 = 0.9$ and $w_1 = 0.1$, the weights are chosen equal in the two designs in order to eliminate the impact of change the weight on the results. The bandwidth sharing weights assigned to the arrivals of handoff RT and handoff NRT are $w_{HD1} = 0.7$ and $w_{HD2} = 0.3$. The different service rates are calculated as follows:

$$\begin{aligned} \mu_0 &= w_{HD1}\mu_1 + w_{HD2}\mu_2, \\ \mu_1 &= \frac{\text{service rate (bit/s)}}{\text{packet length RT (bit)}}, \\ \mu_2 &= \frac{\text{service rate (bit/s)}}{\text{packet length NRT (bit)}}. \end{aligned} \quad (25)$$

Table 2 indicates the threshold values of the states of transitions corresponding to T_{r2} and T_{r4} , the periods of transmission of RT and NRT packets according to PQ in the first and the second design, respectively. Thus, the components (i_{RT}, j_{RT}, k_{RT}) are obtained when $t = Th_{RT}$ and $(i_{NRT}, j_{NRT}, k_{NRT})$ are calculated at $t = Th_{NRT}$. (i_{HD}, j_{HD}, k_{HD}) indicate the end of the transmission of RT or NRT packets following PQ and the beginning of the transmission of handoff packet following PQ, and these components are calculated when $t = Th_{RT} + Th_{HD}$.

Figures 5–7 compare, consecutively, the results of the of queues lengths L_{q0} , L_{q1} and L_{q2} for the first, the second, and the third scenarios, respectively. L_{q0} is the handoff queue length, L_{q1} is the RT queue length, and L_{q2} is the NRT queue length.

Figure 5 shows that the first scenario gives the lowest values of L_{q0} , the handoff queue length. L_{q0} increased after the application of the first conception (scenario 2) of the new courteous algorithm because during a defined time, the priority of transmitting is given to RT traffic even if there is a handoff packet in the handoff queue. Furthermore, L_{q0} increased following the application of the second conception

TABLE 1: Simulation parameters.

Parameter	Value
Aggregation configuration	4 CCs of 20 MHz, (180 kHz per PRB)
Number of cells	1
Number of antennas	1 × 1
Number of PRB in uplink	90 PRB
PRB duration	1 TTI
Number of subcarriers per resource block	12
Number of symbols per subcarrier	7
Handoff interarrival (λ_0)	15626 packets/s
RT interarrival (λ_1)	50000 packets/s
NRT interarrival (λ_2)	3333.4 packets/s
System size (k)	12 packets
VoLTE packet size	300 bits
FTP packet size	512 bytes
CQI index	7
Modulation	16 QAM
Coding rate	378/1024
Simulation time	1000TTI

TABLE 2: Threshold values for the various periods of transmission.

Parameter	Results									
Simulation time (s)	1	2	3	4	5	6	7	8	9	10
i_{RT}	1	1	1	1	1	1	1	1	1	1
j_{RT}	1	1	1	1	1	1	1	1	1	1
k_{RT}	0	0	0	0	0	0	0	0	0	0
i_{NRT}	1	1	1	1	1	1	1	1	1	1
j_{NRT}	1	1	1	1	1	1	1	1	1	1
k_{NRT}	1	1	1	1	1	1	1	1	1	1
i_{HD}	11	11	11	11	11	0	0	0	1	0
j_{HD}	11	11	11	11	11	5	7	7	8	7
k_{HD}	11	11	11	11	11	4	2	1	0	0

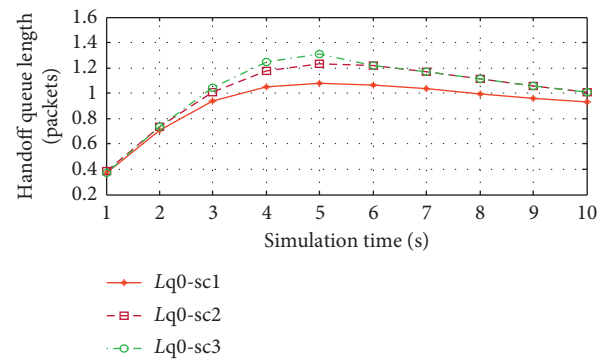


FIGURE 5: Handoff queue lengths vs. simulation time.

(scenario 3), especially, between 2 s and 5 s. Besides, from $t = 6$ s, the graphics L_{q0} -sc2 and L_{q0} -sc3 are closely similar. The increasing of the handoff queue length is affected by a decrease in the length of queues RT and NRT to maintain the system stationary since the three queues share the same buffer K as designed in the mathematical model. Thus, our results show that the application of the new courteous

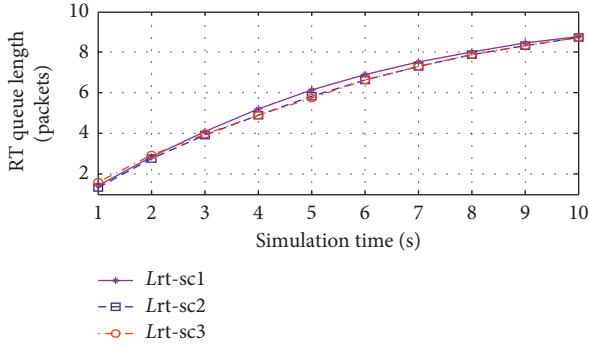


FIGURE 6: RT queue lengths vs. simulation time.

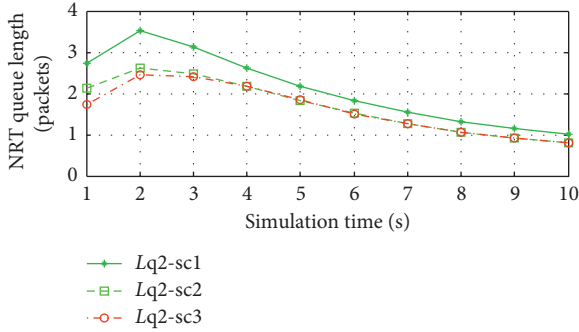


FIGURE 7: NRT queue length vs. simulation time.

algorithm improves the queuing length for the lower priorities classes, but this does not affect the quality of service of the highest priority class.

Figure 6 shows that Scenario 3 gives the highest value of L_{q1} , between 1 s and 2 s. This means that when courteous algorithm is applied to NRT traffic and when handoff and RT are served according to CBWFQ politic by giving a higher priority to handoff with respect to RT, the RT queue will have the biggest number of waiting packets. This increase is due to the reduction of the VoLTE packet transmission period T_{r2} following PQ to give a priority to FTP packets. Then, L_{q1-sc3} starts to decrease and reaches approximately the value of L_{q1-sc2} at $t = 6$ s.

However, the changes in the NRT queue length are influenced by the variations in the lengths of the handoff and RT queues, as shown in Figure 7, for t between 1 s and 4 s, Scenario 3 gives the best result of the NRT queue length. From $t = 4$ s, the graphics of the two designs become almost similar.

After studying the queue lengths, we consider the throughput herein. The throughput is defined by the average rate of packets that are transmitted through the system at steady state:

$$T_p = \lambda_p (1 - \pi^*), \text{ with } p = 0, 1, 2. \quad (26)$$

Figures 8–10 show the throughput T_0 , T_1 , and T_2 related to handoff, RT, and NRT classes. The results show first that as the throughputs increase when the number of interarrivals in the network is greater. In addition, the throughputs are increased in the second and third scenario. However, it is

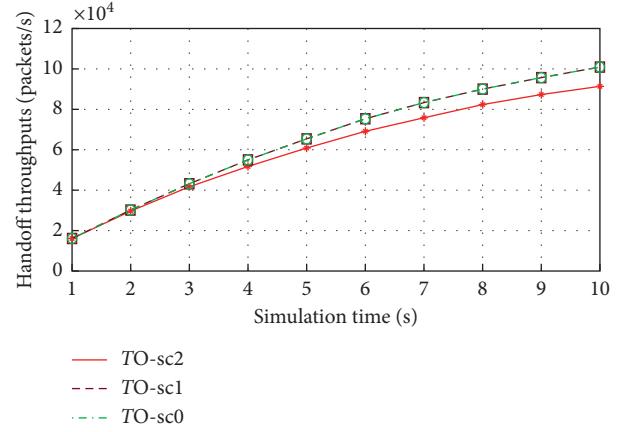


FIGURE 8: Handoff throughputs vs. simulation time.

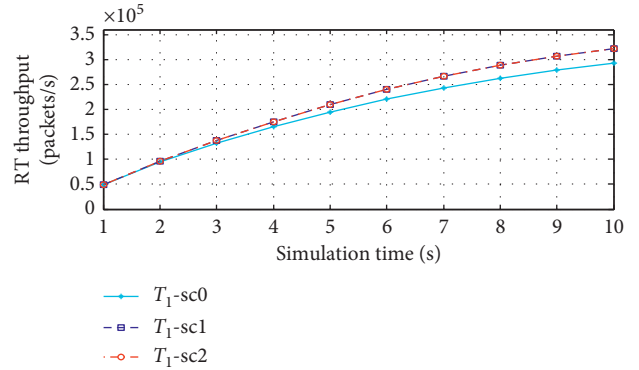


FIGURE 9: RT throughputs vs. simulation time.

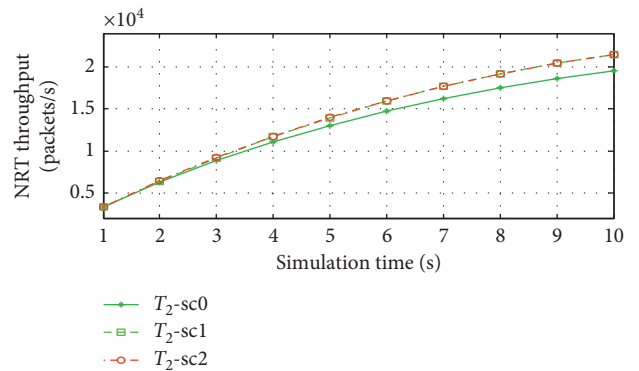


FIGURE 10: NRT throughputs vs. simulation time.

difficult to observe the effect of the third scenario on the number of interarrivals accepted and served, as the graphics are almost similar.

Thus, Table 3 represents the throughput values of the second and third scenarios, and it shows that when t is included between 2 s and 5 s, the third scenario offers the best values for the three classes, as NRT packet transmission intervals following PQ are more wide (see Table 2).

TABLE 3: Throughputs values t of both conceptions of new courteous algorithm.

Parameter	Results									
Simulation time (s)	1	2	3	4	5	6	7	8	9	10
T_0 -sc2 $\times 1.0e + 05$ (Packets/s)	0.15562	0.30020	0.42965	0.54743	0.65404	0.74951	0.83115	0.9004	0.95833	1.00644
T_0 -sc3 $\times 1.0e + 05$ (packets/s)	0.15530	0.30094	0.43102	0.54864	0.65475	0.74942	0.83109	0.90037	0.95833	1.00644
T_1 -SC2 $\times 1.0e + 05$ (packets/s)	0.49796	0.96064	1.37927	1.75176	2.09294	2.39845	2.65968	2.88130	3.0666	3.22062
T_1 -sc3 $\times 1.0e + 05$ (packets/s)	0.49699	0.96300	1.37927	1.75564	2.09520	2.39814	2.65950	2.88120	3.0666	3.22059
T_2 -sc2 $\times 1.0e + 05$ (packets/s)	0.33198	0.64044	0.91660	1.16787	1.39532	1.59900	1.77316	1.92091	2.0449	2.14712
T_2 -sc3 $\times 1.0e + 05$ (packets/s)	0.33133	0.64201	0.91953	1.17045	1.39683	1.59879	1.77303	1.92084	2.04444	2.14710

In the remaining simulation time, the throughput achieved in the second scenario is slightly higher than the throughput of the third scenario because of the T_{r4} interval which became closer.

The packets delay of each class is shown in Figures 11–13, the variation of delays W_{qp} depends mainly on the queue lengths and the throughput values obtained since the delays are defined as

$$W_{qp} = \frac{L_{qp}}{\lambda_p (1 - \pi^*)}. \quad (27)$$

In Figure 11, W_{q0} -sc3, the waiting time of handoff traffic for the third scenario is below W_{q0} -sc2, which is the waiting time of handoff traffic for the second scenario. In both scenarios, the numbers of handoff interarrival packets are almost similar with a slight decrease in T_0 -sc3, the handoff throughput in scenario 3. Then, W_{q0} -sc3 becomes greater than W_{q0} -sc1 and W_{q0} -sc2 for $t > 2$ s to $t = 6$ s because of L_{q0} -sc3 which experienced a sizeable increase. However, the increase of T_0 -sc3 is less significant. From $t = 6$ s, W_{q0} -sc3 and W_{q0} -sc2 become almost similar since the values of the queues lengths and throughputs are almost similar in both scenarios.

Regarding the delays in the RT queue, the third scenario has the highest value among all values for t between 1 s and 2 s because of L_{q1} -sc3 which has increased and the throughput T_1 -sc3 which has slightly decreased. Then, the delay W_{q1} -sc3 decreases and becomes less than W_{q1} -sc1 and W_{q1} -sc2 in the period from 4 s to 6 s because L_{q1} -sc3 decreased in this period, and the throughput T_1 -sc3 increased from $t = 6$ s and the throughputs in the second and third scenario are almost similar.

Besides, when t is between 1 s and 4 s, W_{q2} -sc3 is the shortest, thanks to L_{q2} -sc3 which decreased and T_2 -sc3 which increased. Then, when $t = 2$ s, W_{q2} -sc3 approaches to W_{q2} -sc2, from $t = 4$ s, the two delays become approximately similar because the queues lengths and the throughputs are roughly equivalent.

Figure 14 shows that the blocking probability in the second and in the third scenario is less than the blocking probability in the first scenario. In other words, the number of users rejected by the network is lower. The tracings graphs of the blocking probability sc2 and sc3 in scenarios 2 and 3 are equivalent. However, the values are slightly different in the two scenarios. As shown in Table 4, the third scenario gives the best results of π^* for t between 2 s and 5 s.

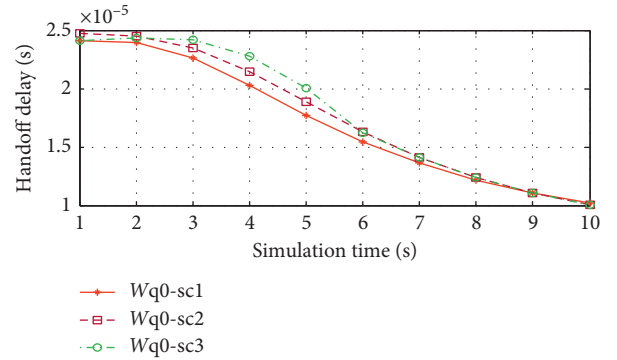


FIGURE 11: Average handoff delays vs. simulation time.

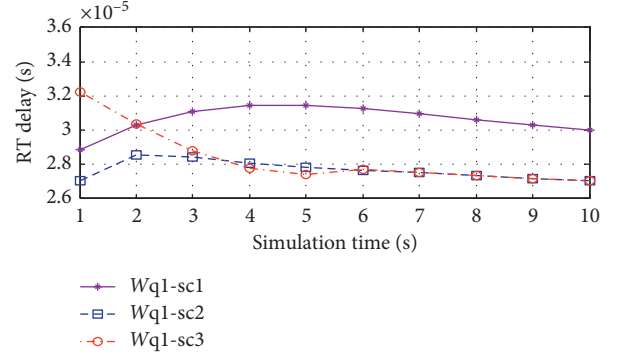


FIGURE 12: Average RT delays vs. simulation time.

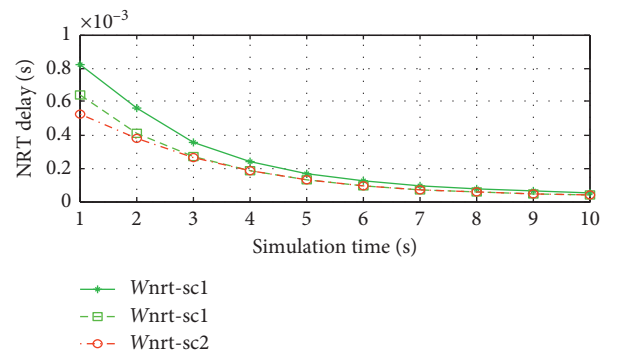


FIGURE 13: Average NRT delays vs. simulation time.

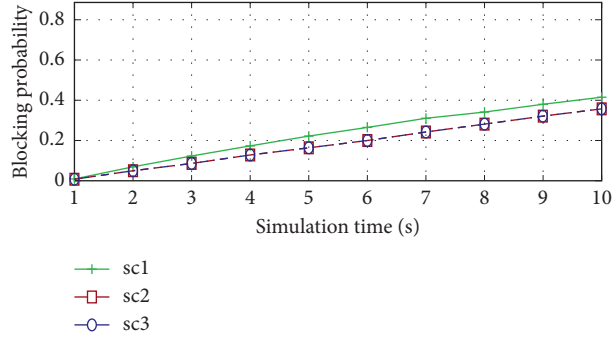


FIGURE 14: Blocking probability vs. simulation time.

TABLE 4: Blocking probability of both conceptions of new courteous algorithm.

Parameter	Results									
Simulation times (s)	1	2	3	4	5	6	7	8	9	10
Blocking probability	0.0041	0.3935	0.08342	0.12412	0.16282	0.20051	0.24009	0.27697	0.31852	0.35587
Blocking probability	0.00602	0.03699	0.08048	0.12218	0.16192	0.20062	0.24014	0.27969	0.31853	0.35588

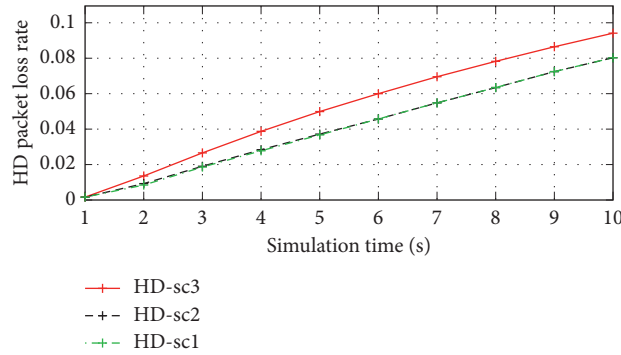


FIGURE 15: Handoff packet loss rates vs. simulation time.

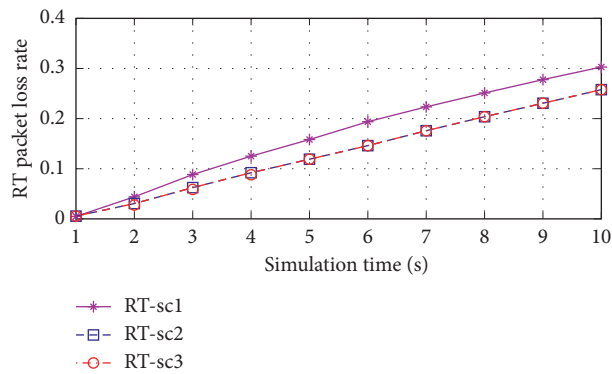


FIGURE 16: RT packet loss rates vs. simulation time.

Furthermore, the packet loss rate per classes is calculated as follows:

$$\pi_l p = \frac{\lambda p}{\lambda_0 + \lambda_1 + \lambda_2} \pi^*, \quad p = 0, 1, 2. \quad (28)$$

The packet loss rate related to the three queues is illustrated in Figures 15–17. According to the resulting

graphs, the application of two conceptions of the new courteous algorithm has downplayed rejecting the handoff, RT, and NRT packets.

The results obtained by Figure 15 are the related to the one obtained in Figures 5 and 11 for waiting time and queuing length of handoff traffic in the three scenarios. We observe that the waiting time and the queue length give best results

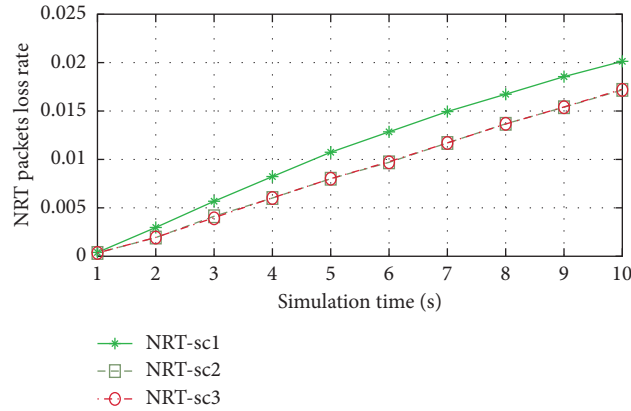


FIGURE 17: NRT packet loss rates vs. simulation time.

TABLE 5: Packet loss rates of both conceptions of new courteous algorithm.

Parameter	Results									
Simulation time (s)	1	2	3	4	5	6	7	8	9	10
HD-sc2	0.00092	0.00891	0.01890	0.02812	0.03689	0.04543	0.05440	0.06337	0.07217	0.08064
HD-sc3	0.00136	0.00838	0.01824	0.02768	0.03668	0.04546	0.05441	0.06337	0.07217	0.08063
RT-sc2	0.00296	0.02853	0.06048	0.08999	0.11805	0.14538	0.17408	0.20278	0.23095	0.25804
RT-sc3	0.00436	0.02682	0.05836	0.08859	0.11740	0.14546	0.17412	0.20280	0.23095	0.2580
NRT-sc2	0.00019	0.00190	0.00403	0.00599	0.00787	0.00969	0.01160	0.01352	0.01539	0.01720
NRT-sc3	0.00029	0.00179	0.00389	0.00590	0.00783	0.00970	0.01161	0.01352	0.01540	0.01720

when the basic scenario is considered, that is, when the Priority Queuing scheme is applied. Besides, when new courteous algorithm is executed, a part of handoff resources will be leave to RT and NRT traffics depending on their needs Table 5 indicates that in the period from 2 s and 5 s, the number of packets lost in the system in the second conception is the lowest. However, the results obtained in the second and third scenario are almost similar for t between 6 s and 10 s.

The transmission of NRT packets following PQ has delayed the service of the less tolerant packet delay (handoff and RT). Nevertheless, the sudden increase of packet loss rate was avoided because each increase in the queue length for a class was offset by a decrease in the length of queue of another class.

4. Conclusion

This paper deals with the scheduling system in the uplink over LTE-A and 5G networks. In this work, we propose a new courteous algorithm to improve the radio resource blocks (RBs) allocation over LTE-A and 5G networks. The low-priority traffic classes, using lower priority bearers, are favoured during a specific time interval based on the average waiting time for each class. This one is calculated using the mathematical study and integrated to the courteous algorithm, which is implemented in MATLAB and simulated to discuss the different results. Like the average waiting time, the average queue length is calculated using the mathematical model modeling in this paper and integrated to the courteous algorithm proposed in this study. Therefore, the analytical model was developed in order to measure the different delays and queues length in order to define the

different thresholds. These ones are directly related to the decisions made by the courteous scheme when changing the traffics priorities. Our scheme implements two conceptions of the new courteous algorithm, by firstly applying the courteous to the RT traffic only, which is called the first conception, and secondly applying the same scheme to the RT and NRT traffics consecutively, namely, the second conception. The first conception of the new courteous algorithm has reduced the delays and the queue lengths of RT class. It offers a lower blocking probability and packet loss rate of the three classes and optimizes the throughputs. Furthermore, it increases the length of the handoff, though the QoS of the handoff class was not degraded. The second conception provides more efficient results than the first design in terms of throughput, blocking probability, and packet loss rate when the transmission intervals dedicated to the NRT class served following PQ increase.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the NSERC grant.

References

- [1] F.-C. Kuo, K.-C. Ting, T. H.-C. Wang, C.-C. Tseng, and M.-W. Chen, "Differentiating and scheduling LTE uplink traffic based on exponentially weighted moving average of data rate," *Mobile Networks and Applications*, vol. 22, no. 1, pp. 113–124, 2017.
- [2] N. Abu-Ali, A.-E. M. Taha, M. Salah, and H. Hassanein, "Uplink scheduling in LTE and LTE-advanced: tutorial, survey and evaluation framework," *IEEE Communications Surveys and Tutorials*, vol. 16, no. 3, pp. 1239–1265, 2014.
- [3] A. Ragaleux, S. Baey, and M. Karaca, "Standard-compliant LTE-A uplink scheduling scheme with quality of service," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 8, pp. 7207–7222, 2017.
- [4] R. Kausar, Y. Chen, and K. K. Chai, "QoS aware Packet Scheduling with adaptive resource allocation for OFDMA based LTE-advanced networks," in *IET International Conference Communication Technology and Application, ICCTA 2011*, pp. 207–212, Beijing, China, October 2011.
- [5] S. Henrique da Mata and P. R. Guardieiro, "Resource allocation for the LTE uplink based on genetic algorithms in mixed traffic environments," *Computer Communications*, vol. 107, pp. 125–137, 2017.
- [6] B. A. Khitem, F. Zarai, and L. Kamoun, "Reducing handoff dropping probability in 3GPP LTE network," in *Proceedings of the Second International Conference on Communications and Networking (ComNet)*, Tozeur, Tunisia, November 2010.
- [7] V. G. Vassilakis, I. D. Moscholios, A. Bontozoglou, and M. D. Logothetis, "Mobility-aware QoS assurance in software-defined radio access networks: an analytical study," in *Proceedings of the 2015 1st IEEE Conference on Network Softwarization (NetSoft)*, pp. 747–750, London, UK, April 2015.
- [8] M. A. Safwat, H. M. E. Badawy, A. Yehya, and H. E. Motaafy, "Performance analysis for new call bounding scheme with SFR in LTE-advanced networks," in *Proceedings of the IEEE 6th Intl Symp on Cyberspace Safety and Security and 2014 IEEE 11th Intl Conf on Embedded Software and Syst (HPCC, CSS, ICSSS), High Performance Computing and Communications*, pp. 442–451, Paris, France, August 2014.
- [9] Y. W. Chen, P. Y. Hsu, J. H. Tsai, P. Y. Liao, and M. H. Lin, "A heuristic design of uplink scheduler in LTE-A networks," in *Proceedings of the Tenth International Conference Intelligent Information Hiding and Multimedia Signal Processing, IIH-MSP Kitakyushu*, pp. 884–888, Kitakyushu, Japan, August 2014.
- [10] J. Huang, C.-C. Xing, J. Zygmunt, and Haas, "Resource allocation for multicell device-to-device communications underlaying 5G networks: a game-theoretic mechanism with incomplete information," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 3, pp. 2557–2570, 2018.
- [11] M. Kadoch and C. Tata, "Courteous algorithm: performance optimization in WiMAX networks," in *Proceedings of the ICT'10 (WSEAS)*, 2010.
- [12] A. Alsawaai, "Performance modelling and analysis of weighted fair queueing for scheduling in communication networks," in *Proceedings of the 24th IEEE International Conference on Advanced Information Networking and Applications*, pp. 50–57, Perth, Australia, April, 2010.
- [13] K. Intharawijitr, K. Iida, and H. Koga, "Analysis of fog model considering computing and communication latency in 5G cellular networks," in *Proceedings of the 2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, pp. 1–4, Sydney, Australia, March 2016.
- [14] N. Li and D. A. Stanford, "Multi-server accumulating priority queues with heterogeneous servers," *European Journal of Operational Research*, vol. 252, no. 3, pp. 866–878, 2016.
- [15] T. Han, G. Mao, Q. Li, L. Wang, and J. Zhang, "Interference minimization in 5G heterogeneous networks," *Mobile Networks and Applications*, vol. 20, no. 6, pp. 756–762, 2015.
- [16] G. Natarajan, *Analysis of Queues: Methods and Applications*, CRC Press, Boca Raton, FL, USA, 1st edition, 2012.
- [17] Q. M. He, *Fundamentals of Matrix-Analytic Methods*, Springer, New York, NY, USA, 1st edition, 2014.
- [18] E. H. Elhafsi and M. Molle, *On the Solution to QBD Processes with Finite State Space*, Department of Computer Science and Engineering, University of California, Oakland, CA, USA, 2007.
- [19] L. Kleinrock, *Queueing Systems: Computer Applications*, Vol. 2, Wiley, New York, NY, USA, 1st edition, 1975.
- [20] S. Bergida and Y. Shavitt, *Analysis of Shared Memory Priority Queues with Two Discard Levels*, Tel Aviv University, Tel Aviv, Israel, 2005.