

Article

# An Intelligent Fuzzy Logic-based Content and Channel Aware Downlink Scheduler for Scalable Video over OFDMA Wireless Systems

Peter E. Omiyi <sup>1</sup>, Moustafa M. Nasralla <sup>2\*</sup>, Ikram Ur Rehman <sup>3</sup>, Nabeel Khan <sup>4</sup> and Maria G. Martini <sup>5</sup>

<sup>1</sup> Faculty of Engineering, Holon Institute of Technology

<sup>2</sup> Department of Communications and Networks Engineering, Prince Sultan University, Riyadh, Saudi Arabia;

<sup>3</sup> School of Computing and Engineering, University of West London, UK;

<sup>4,5</sup> Faculty of Science, Engineering and Computing, Kingston University;

\* Correspondence: mnasralla@psu.edu.sa

Academic Editor: name

Version June 22, 2020 submitted to Electronics; Typeset by L<sup>A</sup>T<sub>E</sub>X using class file mdpi.cls

**Abstract:** The recent advancements of wireless technology and applications make downlink scheduling and resource allocations an important research topic. In this paper, we consider the problem of downlink scheduling for multi-user scalable video streaming over Orthogonal Frequency Division Multiple Access (OFDMA) channels. The video streams are precoded using an scalable video coding (SVC) scheme. We propose a Fuzzy logic-based scheduling algorithm, which prioritizes the transmission to different users by considering video content, and channel conditions. Furthermore, a novel analytical model and a new performance metric have been developed for the performance analysis of the proposed scheduling algorithm. The obtained results show that the proposed algorithm outperforms the content-blind/channel aware scheduling algorithms with a gain of as much as 19% in terms of the number of supported users. The proposed algorithm allows for a fairer allocation of resources among users across the entire sector coverage, allowing for the enhancement of video quality at edges of the cell while minimizing the degradation of users closer to the base station.

**Keywords:** Content-Aware, Cross-Layer, Fuzzy Inference System, OFDMA, Scheduling, SVC, Video Streaming

## 1. Introduction

Supporting multimedia applications and services over wireless networks is challenging due to constraints and heterogeneities such as limited bandwidth, limited battery power, random time-varying channel conditions, different protocols and standards, and varying quality of service (QoS) requirements. Two main classifications can be performed as far as scheduling algorithms are concerned: channel-aware schedulers and content-aware schedulers. A comprehensive survey on downlink channel-aware and content-aware scheduling algorithms can be found in [1,2].

It is worth mentioning that channel-unaware schedulers make no use of channel state conditions such as power level, channel error, and loss rates. These basically focus on fulfilling delay and throughput constraints. Examples of the traditional channel-unaware schedulers are Round-Robin, weighted fair queuing (WFQ), and priority-based algorithms. Such algorithms assume perfect channel conditions, no loss, and unlimited power source. However, due to the nature of wireless

28 medium and the user mobility, these assumptions are not valid. The Base Station (BS) downlink  
29 scheduler could rather use channel information (e.g. channel state information (CSI), including the  
30 Carrier to Interference and Noise Ratio (CINR)) which is reported back from the mobile receiver. Most  
31 of the channel-aware algorithms assume that channel conditions do not change within the frame  
32 period. It is also assumed that the channel information is known at both the transmitter and the  
33 receiver. In general, schedulers favour the users with better channel quality to exploit the multi-user  
34 diversity and channel fading. However, to meet fairness requirements, the scheduler also needs to  
35 consider other users' requirements and should introduce some compensation mechanisms.

36 In content-unaware scheduling strategies, the QoS of the received video is measured in generic  
37 terms of packet delay, packet loss rate or data rate. In general, these methods exploit the variability  
38 of the wireless channel over time and across users, allocating a majority of the available resources  
39 to users with good channel quality. Ultimately, these scheduling strategies support higher data rates,  
40 while maintaining fairness across multiple users. In this context, these strategies attempt to maximize  
41 a utility function, which is defined as either a function of each user's current average throughput, or  
42 of each user's queue length or delay of the head-of-line packet [2].

43 In contrast, content-aware scheduling strategy is not a simple function of the data rate, delay or  
44 data loss but it is rather affected differently by the impact of losses and errors in different segments of  
45 the video stream. This is highlighted in an SVC bitstream, which consists of one base layer and  
46 multiple enhancement layers. As long as the base layer is received, the receiver can decode the  
47 video stream. As more enhancement layers are received, the decoded video quality is improved.  
48 In multi-user video transmission, this introduces a type of multi-user content diversity that can be  
49 exploited by content-aware scheduling policies in optimizing the utilization of the network resources.  
50 Examples of content-aware methods and current SVC studies are found in [2–9].

51 Unlike state-of-the-art content aware strategies, the proposed scheduling rule considers SVC  
52 layer priority index. The higher the layer priority, the higher the probability of the layer to be  
53 scheduled. The layers can be marked outside the eNodeB, for instance at the P-GW or video  
54 server, whereas the scheduler at the eNodeB exploits the layer priority marking and schedule layers  
55 contributing maximum to the overall video quality. There exist several QoE layer marking strategies,  
56 such as in [10,11], where SVC layers are marked based on their contribution to the overall QoE.  
57 Therefore, the proposed scheduling rule requires only layer priority index at the eNodeB, whereas the  
58 complex processing of SVC layer marking is performed outside the eNodeB. On the other hand, the  
59 state-of-the-art content aware scheduling requires complex video content processing at the eNodeB.  
60 However, the transfer of video content related information to the eNodeB is not practical thus  
61 restricting the usage of such strategies.

62 Several content aware scheduling strategies [12–17] evaluate the value of the content and  
63 maximize the video quality of the streaming users subject to the channel constraints. However, such  
64 strategies suffer from high computation complexity at the MAC layer of the eNodeB. In order to  
65 address the complexity issue, we proposed a scheduling strategy, where complexity in terms of the  
66 number of iterations varies linearly *w.r.t* the number of users and resources. In other words, the  
67 proposed fuzzy-based scheduling priority function is a linear function of the users (competing for  
68 resources) and the number of resources. This is in contrast to the content-aware scheduling strategies  
69 where scheduling complexity varies exponentially *w.r.t* the number of users and resources.

70 Furthermore, the literature lacks proposals on content-aware priority-based scheduling  
71 algorithm which utilizes an Fuzzy Inference System (FIS). An FIS considers the concept of vagueness  
72 and uses probability-based mathematical models to represent the vagueness. Words/estimates are  
73 potentially less precise than numbers or Boolean representation; however, words are closer to human  
74 intuition. Hence, FIS would be a good approach to explore the tolerance for imprecisions and  
75 hence gain a better understanding of the application. There are two common inference methods:  
76 the Mamdani's fuzzy inference method and the Takagi-Sugeno-Kang method of fuzzy inference. In  
77 several studies related to real time scheduling, as in [18,19], it was proven that Mamdani-type FIS and

78 Sugeno-type FIS perform similarly, except that using Sugeno-type FIS model allows the scheduling  
79 system to work at its full capacity. In addition, it was proven that Sugeno-type FIS has the advantage  
80 that it can be integrated with neural networks and genetic algorithm or other optimization techniques,  
81 so that the controller can adapt to individual user and variable channel conditions [18,19]. FIS is an  
82 effective tool to establish relationships between input and output variables. It is particularly useful for  
83 relatively small dataset and limited number of input variables. Utilising FIS, we propose a downlink  
84 scheduling algorithm and a user utility function, which complements our study. Furthermore,  
85 this method provides computational efficiency and is well-suited for optimization and adaption of  
86 algorithms, which makes it a potential candidate for scheduling problems, in particular for dynamic  
87 wireless systems. Hence, for our study the popular Sugeno's FIS method is chosen.

88 This paper provides four main contributions, highlighted below:

- 89 1. Proposing a multi-user content-aware priority-based scheduling algorithm, where packet  
90 priorities are selected based on Sugeno FIS.
- 91 2. Proposing a framework for quantitatively classifying the video content, in order to apply the  
92 proposed FIS.
- 93 3. Proposing a performance metric called significance throughput. This metric gives a better  
94 indication of the scheduler performance for content sensitive traffic than throughput.
- 95 4. Lastly, proposing a novel analytical model of the FIS-based scheduling algorithm, and  
96 providing analysis of it.

97 The rest of the paper is organized as follows. Section 2 provides a background on fuzzy inference  
98 system, OFDMA systems, and scalable video coding. In Section 3, we present the related work on the  
99 existing downlink scheduling algorithms. Section 4 describes the methodology which consists of an  
100 FIS-based downlink scheduling algorithm, a wireless system model, a novel key performance metric  
101 (i.e. significance throughput), and lastly, the analytical model to analyse the proposed scheduling  
102 algorithm. Results of the analytical model are reported in Section 5. Finally, Section 6 concludes the  
103 paper.

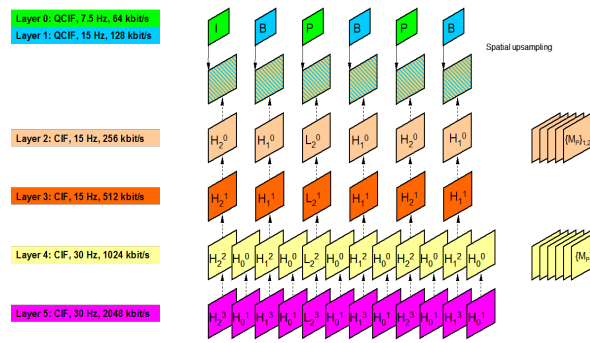
## 104 2. Background

105 In this section, we provide background on the core aspects of this paper which are fuzzy  
106 inference system, OFDMA systems, and scalable video coding.

107 As mentioned earlier in Section 1, for our study the Sugeno fuzzy inference method is chosen.  
108 The underlying concept of FIS is that of a linguistic variable which makes it closer to human intuition.  
109 Hence, fuzzy logic is a good approach to explore the tolerance for imprecisions and hence gain a better  
110 understanding of the application. An FIS performs the mapping of a given input to an output using  
111 the Fuzzy Logic and by employing components such as membership functions, fuzzy logic operators  
112 and If-Then rules. After the input and output variables are defined for the Fuzzy system, the next  
113 step is to assign linguistic labels in order to provide quantification of the values, which are defined  
114 through membership functions. More details on the functionality of FIS can be found in [20] [21].

115 The underlying wireless technology considered in this paper is 4G system, which is based  
116 on OFDMA. The OFDMA systems allow multiple users to share the spectrum at the same time.  
117 The subcarriers in OFDMA are shared between multiple users; to enable better utilisation of radio  
118 resources. This technique helps wireless technologies improve the system capability to achieve  
119 the following: 1) support high data rates, 2) provide multi-user diversity, 3) compact/eliminate  
120 the Inter-Symbol-Interference (ISI) caused by multipath fading, and 4) to be immune to frequency  
121 selective fading [2].

122 The video streams used in this paper are precoded using an SVC scheme. SVC [22] represents  
123 a video sequence via multiple layers with different quality, resolution, and/or frame rate as shown  
124 in Figure 1. SVC enables graceful degradation of video quality when resources are limited, hence it  
125 is particularly suitable for the case of multi-user video scheduling. In other words, an SVC stream



**Figure 1.** Temporal, Spatial and Quality scalability of SVC.

126 has a base layer and several enhancement layers. As long as the base layer is received, the receiver  
 127 can decode the video stream. As more enhancement layers are received, the decoded video quality  
 128 is improved. The scalability of SVC consists of temporal scalability, spatial scalability, and quality  
 129 scalability. In this work, we consider SVC with temporal scalability, however our approach is also  
 130 applicable to other scalability models which are defined in [22]. For example, in temporal scalability  
 131 model, we consider a Hierarchical B frame GOP structure as follows  $\{K_0 B_2 B_1 B_2 K_0 \dots\}$ , where  $K_0$  is an I  
 132 or P key picture. The number of coding layers  $N_L = 3$ . Each layer is composed of one or more frames.  
 133 The layers in order of importance are the key picture  $\{K_0\}$  with index  $l = 0$ ,  $\{B_1\}$  with index  $l = 1$ ,  
 134 and  $\{B_2 B_2\}$  with index  $l = 2$ . The significance values are  $v = 1$ ,  $v = 2/3$  and  $v = 1/3$ , respectively.

### 135 3. Related Work

136 Over the years, various packet scheduling algorithms have been developed to support Real  
 137 Time (RT) and non-Real Time (NRT) services, comprising the most commonly used ones, namely:  
 138 Proportional Fair (PF), Modified Largest Weighted Delay First (M-LWDF) and Exponential-PF  
 139 (EXP-PF) schedulers [2,3]. In the aforementioned schedulers, each flow is assigned a priority value  
 140 and the radio bearer which carries the flow with the highest priority value will be scheduled  
 141 first at the corresponding Transmission Time Interval (TTI). When transmitting multimedia  
 142 services to multiple users over wireless systems, a scheduling strategy should address the trade-off  
 143 between resource utilization and fairness among users. Network operators are mostly interested  
 144 in maximizing the exploitation of the resources, e.g., assigning more resources to the user(s)  
 145 experiencing better channel conditions. However, this approach of theirs can result in unsatisfied  
 146 users, which in turn would result in users' experiencing worse channel conditions and hence, leading  
 147 towards not meeting their QoS and Quality of Experience (QoE) requirements.

148 In our previous studies [2,23], we carried out a comprehensive review on the existing  
 149 content-aware strategies. In addition, we classified content-aware strategies into the following three  
 150 classes: 1) Quality driven scheduling approach, 2) Proxy driven radio resource allocation approach,  
 151 and 3) Client driven approach. In this paper, we take a step forward on proposing a content-aware  
 152 scheduling strategy that would fall under the first class i.e., Quality driven scheduling approach as  
 153 this approach consists of scheduling strategies specifically designed for video streaming traffic. In  
 154 this approach, the information on the content of different video traffic flows is provided through  
 155 cross-layer signaling to the Radio Access Network (RAN). These types of schedulers consider in their  
 156 scheduling decision different objective functions (e.g., mean square error (MSE), peak signal-to-noise  
 157 ratio (PSNR), and Structural Similarity (SSIM)) based on the video quality. The main goal of this  
 158 scheduler is to maximize the video quality of the streaming users under channel and bandwidth  
 159 constraints.

160 Content-aware downlink packet scheduling schemes for multi-user scalable video delivery  
 161 over wireless networks are proposed in [24–26]. Their schedulers use a gradient-based scheduling

162 framework along with SVC schemes. Similarly, a content-aware and fair downlink packet scheduling  
163 algorithm for scalable video transmission over Long-Term Evolution (LTE) systems is proposed in  
164 [27]. The authors proposed a Nash bargaining based on fair downlink scheduling strategy for scalable  
165 video transmission to multiple users. A novel utility metric based on the importance of the video  
166 contents obtained from a Group of Pictures (GOP) is used in conjunction with the decoding deadline  
167 of the GOP. The system capacity in terms of satisfied users can be increased by 20% with the proposed  
168 content-based utility in comparison with advanced, state-of-the-art throughput based strategies. The  
169 authors in [28] improve the work in [27] by exploiting multi-user time-averaged diversity. The reason  
170 for using SVC is to provide multiple high quality video streams over different prevailing channel  
171 conditions for multiple users. The schedulers proposed outperform the traditional content-blind  
172 scheduling approaches. Furthermore, a significant improvement was observed in terms of objective  
173 video quality metrics (e.g. Throughput, PSNR and SSIM etc.) when the proposed scheduling schemes  
174 were compared with the content-blind scheduling schemes in the presence of network congestion.  
175 Hence, it was established that the video content should be given utmost importance after QoS, when  
176 determining the quality of video sequences [3]. However, the proposed content-aware schedulers  
177 did not explicitly consider channel conditions in its allocation process. In a wireless environment,  
178 this could lead to poor video quality, with a few users with very poor channel conditions, using  
179 almost all the available channel resources to satisfy their video quality requirements.

180 It is worth mentioning that video quality is subjective, and while it is relatively straightforward  
181 to distinguish between the importance of different segments of the video stream, based on their  
182 relative impact on video quality, it is difficult to quantify these differences. In [3,29–31], priority-based  
183 scheduling algorithms are proposed, with the priority function taking into account the importance of  
184 different frame types, channel conditions, buffer state and the relative start time of the video streams  
185 of the users. At the beginning of a time slot the scheduler computes the priorities of all users and  
186 schedules the one with the highest priority to transmit. This scheme when compared to non-content  
187 aware scheduling ensures that the higher priority frames have a lower frame loss rate. However,  
188 it is not clear how to set the priorities assigned to the different frame types, in order to optimize  
189 performance. This is particularly an issue when SVC is considered and a larger set of possible  
190 priorities exist.

191 To elaborate on the significance of priority-based content-aware scheduling, a QoE-based packet  
192 marking strategy scheduling model is presented in Figure 2. According to the figure, the marking  
193 algorithm at the Packet Data Network Gateway (P-GW) provides packet prioritization for video  
194 streams having different number of quality enhancement layers. The algorithm at the P-GW exploits  
195 the utility functions (based on MOS vs. Bit-rate) of the video streams and mark layers according  
196 to their bit-rates and contribution towards the overall perceived video quality. The main goal of  
197 the marking is to achieve the maximum video quality under the constraint of the available network  
198 resources. Thus, the packets of video layers contributing less towards MOS at the expense of higher  
199 bit-rates are marked to be served with lower priority. The higher the priority class, the lower the  
200 importance of the marked packets, which is exploited by the scheduler at the eNodeB by dropping  
201 such packets when the system becomes highly congested as given in [32]. According to [10,33],  
202 priority-based optimized packet marking reduces congestion at the base station and provides timely  
203 video rate adaptation at the RAN. However, the approach is limited only to scalable video traffic  
204 without considering video traffic types which do not have scalable properties.

205 Furthermore, to demonstrate the significance of Fuzzy Logic in resource allocation and  
206 scheduling, the authors in [34] proposed a novel fuzzy scheduler for cell-edge users in LTE-advanced  
207 networks using Voronoi algorithm. In this study, the authors focused on proposing an energy efficient  
208 and QoS-aware downlink scheduler for real-time services. Moreover, fuzzy rules were used to  
209 optimize the resource allocation for the downlink scheduling of the cell-edge users. The results  
210 showed that the proposed scheduler is energy efficient, QoS-aware, and beneficial to the cell-edge  
211 users.

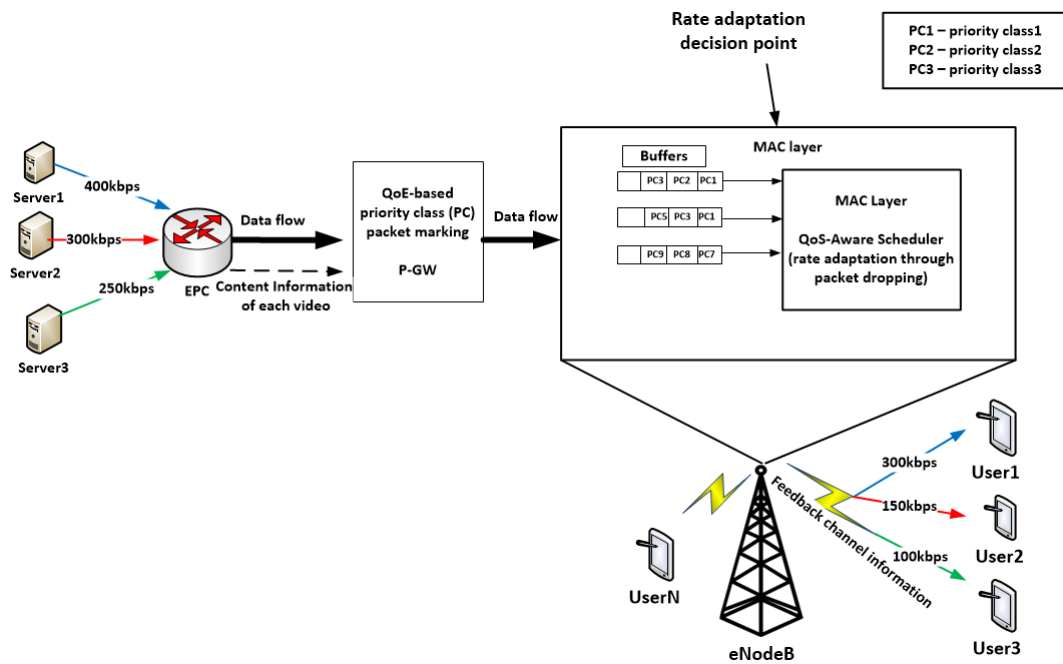


Figure 2. QoE-aware packet priority marking based scheduling.

212 Moreover, the authors in [35] proposed a joint downlink radio resource allocation and scheduling  
 213 algorithm for LTE networks using fuzzy-based adaptive priority and effective bandwidth estimation.  
 214 The resource allocation is based on estimating the utilized bandwidths of traffic flows, whereas the  
 215 downlink scheduling algorithm is designed to compute the adaptive priorities for the different users  
 216 by using fuzzy logic. This study focused on ensuring that the QoS parameters are compliant with  
 217 the requirements of the LTE network. Similarly, the authors in [36] opted for a fuzzy logic approach  
 218 and proposed a joint scheduling and link adaptation scheme. Furthermore, the proposed scheduler  
 219 was a priority-based scheme, which optimally allocates radio resources to multiple users based on  
 220 their QoS requirements for a given application. In addition, the authors went a step further and  
 221 included power optimisation feature, which would adapt to user's power supply constraints. The  
 222 results were obtained through numerical evaluations over FIS. The authors concluded that their  
 223 proposed fuzzy-based scheduler performed similar to the benchmark analytical approach, which  
 224 utilised Lagrange multipliers scheme, however, with less computational complexity.

225 In [37], the authors proposed an intelligent fuzzy logic-based channel-aware resource allocation  
 226 and scheduling scheme over LTE-A networks in the uplink direction. The proposed system was  
 227 designed to optimally accommodate multi-traffic classes (i.e. Real-time and Non-Real-time). Their  
 228 channel-aware framework employed Kalman filter controller for channel estimation as well as to  
 229 meet QoS requirements of the end-users. The performance analysis was carried out in terms of QoS  
 230 indices (e.g. bandwidth, throughput, fairness, jitter and delay), which indicated that the proposed  
 231 fuzzy-based scheduler delivered reliable scheduling for real-time services without comprising the  
 232 non-real-time traffic.

233 Nevertheless, as mentioned in Section 1, the literature lacks proposals on content-aware  
 234 priority-based scheduling algorithm utilizing FIS. FIS is beneficial for research and analysis because it  
 235 provides a trade-off between significance and precision and it relies on concepts of human reasoning  
 236 that is considered to be most reliable. In addition, the aforementioned studies focus more on  
 237 content-blind schedulers and lack addressing content-aware scheduling strategies using Fuzzy logic.  
 238 It is important to note that content-blind schedulers do not produce accurate results as video contents  
 239 contain different spatio-temporal features, which are their unique signatures, which this paper aims

240 to address by proposing an intelligent fuzzy logic-based content-aware and channel-aware downlink  
 241 scheduling algorithm for scalable videos over LTE Networks.

#### 242 4. Methodology

243 The methodology followed in this paper can be outlined as proposing a scheduling algorithm  
 244 based on fuzzy logic, which prioritizes the transmission to multiple users by considering video  
 245 content, and channel conditions. We start with proposing wireless system model, followed by  
 246 proposing a fuzzy logic-based content and channel aware downlink scheduler. Next, a novel key  
 247 performance metric (i.e. significance throughput) is proposed for measuring the performance of the  
 248 content-aware scheduling algorithm. Lastly, the analytical model is developed in order to analyse the  
 249 proposed scheduling algorithm.

##### 250 4.1. The Proposed Wireless System Model

251 To design the system model, we consider a single  $120^\circ$  sector of a tri-sectorized hexagonal cellular  
 252 downlink orthogonal frequency division multiple access (OFDMA) network, where each cell is served  
 253 by a base station (BS) with three collocated directional antennas, each serving its respective sector.  
 254 The tagged sector is serving  $N$  active wireless users, uniformly distributed across its area. The system  
 255 bandwidth for each allocation duration is divided into  $M$  physical resource blocks (PRBs), where  
 256 a PRB is a multi-dimensional resource unit spanning a fixed number of OFDMA subcarriers with  
 257 bandwidth  $B$  and symbol durations.

258 All the sectors of the same BS can use the same PRBs simultaneously without interference.  
 259 Adjacent sectors from neighboring cells form a cluster. interference from neighboring clusters is  
 260 substantially mitigated by the sectorized architecture and the propagation pathloss and fading. A  
 261 simple inter-cell coordination algorithm is assumed that avoids interference between sectors of the  
 262 same cluster for all users, by ensuring that neighboring sectors of the same cluster never use the same  
 263 PRBs simultaneously. By ensuring zero ICI for all users, the number of PRBs required by each user is  
 264 minimized, thus making available more PRBs for other users in the sector and in the cluster.

265 The number  $m \leq M$  of PRBs allocated to a sector is a function of the mean expected traffic load  
 266 (in bits/s) requirement of the sector relative to the other sectors in the same cluster. So if  $L_{\text{cluster}}$  and  
 267  $L_{\text{sector}}$  denote the total expected load of all sectors in the cluster and the load of the tagged sector  
 268 respectively, then  $m \approx L_{\text{sector}}/L_{\text{cluster}}$ . The base station can allocate  $m$  PRBs to a set of  $N$  users at each  
 269 allocation instance. At each allocation instance multiple PRBs can be assigned to a single user, each  
 270 PRB however can be assigned to at most one user.

We assume that the channel conditions vary across different PRBs and for different users. The  
 channel conditions vary with time, frequency (e.g. frequency selective multipath fading) and user  
 location. Therefore, each PRB has a corresponding user-dependent and time-varying channel quality  
 that is represented by the maximum possible transmission rate for that user over that PRB. Let  $r_i(t, \phi)$   
 denote the maximum possible transmission rate (bits/s) for user  $i$  over PRB  $\phi$  at time  $t$ . Then,

$$r_i(t, \phi) = G_{\text{mux}} B \log_2(1 + \epsilon_i(\phi) \gamma_i(t, \phi)), \quad (1)$$

271 where  $\gamma_i(t, \phi)$  is the estimated received signal-to-noise ratio (SNR) after diversity combining  
 272 (including MIMO antenna diversity and shadow fading),  $G_{\text{mux}}$  is a MIMO spatial multiplexing gain  
 273 and  $\epsilon_i(\phi)$  is the estimation error margin for  $\gamma_i(t, \phi)$ . We assume that the channel quality feedback from  
 274 the user to the BS, comprising of  $\gamma_i(t, \phi)$  and  $\epsilon_i(\phi)$ , are provided to scheduler within the channel's  
 275 coherence time. Taking bounds of the channel estimation error into account, minimizes the risk of  
 276 errors during transmission.

## 277 4.2. The Proposed Fuzzy Logic-based Content and Channel Aware Downlink Scheduler

278 Utilising FIS, we propose a downlink scheduling algorithm and a user utility function. In this  
279 subsection, we elaborate further on them, respectively.

### 280 4.2.1. The downlink scheduling algorithm

281 The proposed scheduler considers an initial buffer delay or maximum delay constraint  $T_D$ . Each  
282 user must receive one or more GOP, depending on its GOP rate, in this time duration. When a user  
283 needs to receive a number  $g$  GOPs in this time period, then there are  $g$  layers in total with the same  
284 index  $l$ . The proposed scheduler treats these as a single layer with index  $l$  and ensures that they  
285 are sent before any of the  $g$  layers in total with the same index  $l + 1$ . At the receiver, the layers are  
286 re-ordered and reconstituted into frames according to their playback order.

287 The proposed scheduler, at any time instant, allocates PRBs to users iteratively. Let  $\Phi_{ARB}(t, k)$   
288 and  $\Phi_{URB}(t, k)$  denote the set of allocated and unallocated PRBs, respectively by iteration  $k$  of time  
289 slot  $t$ . Let  $\Phi_i(t, k)$  denote the set of PRBs allocated to user  $i$  by iteration  $k$  of time slot  $t$ , and  $r_i(t, \varphi)$  is  
290 the attainable bit rate of the user on PRB  $\varphi \in \Phi_{URB}(t, k)$ . Therefore, each PRB has a corresponding  
291 user-dependent and time-varying channel quality that is represented by the maximum possible  
292 transmission rate for that user over that PRB. Let  $r_i(t, \varphi)$  denote the maximum possible transmission  
293 rate (bits/s) for user  $i$  over PRB  $\varphi$  at time  $t$ . The expression for  $r_i(t, \varphi)$  is given in (1).

For each user  $i$ , an antecedent layer is sent before any of its descendants. Let  $v_i(t, k)$  denote the  
significance of the layer with the highest significance to be sent to user  $i$  by iteration  $k$  of time slot  $t$ .  
The user-priority of user  $i$  by iteration  $k$  of time slot  $t$  on PRB  $\varphi \in \Phi_{URB}(t, k)$  is

$$u_i(t, k, \varphi) = F_{\text{fuzzy}}(v_i(t, k), r_i(t, \varphi)), \quad (2)$$

294 where the function  $F_{\text{fuzzy}}(v_i(t, k), r_i(t, \varphi))$  is determined by zero-order Sugeno fuzzy inference.

295 The iterative algorithm operates as follows at any iteration  $k$  at time  $t$ :

- 296 1. For  $|\Phi_{URB}(t, k)| > 0$ , find a PRB-user pair which has the highest user utility among all available  
297 PRBs and users.
- 298 2.  $\{i^*, \varphi^*\} = \arg \max_{i, \varphi \in \Phi_{URB}(t, k)} u_i(t, k, \varphi)$
- 299 3. Allocate PRB  $\varphi^*$  to user  $i^*$ :

$$\Phi_{PRB, i^*}(t, k + 1) = \Phi_{PRB, i^*}(t, k) + \{\varphi^*\}$$

4. Delete the PRB from the set of available PRBs:

$$\Phi_{URB}(t, k + 1) = \Phi_{URB}(t, k) - \{\varphi^*\}$$

- 300 5. Repeat above until all PRBs are allocated, i.e. until  $|\Phi_{URB}(t, k)| = 0$
- 301 6. Repeat above steps for new time-slot  $t = t + T_{PRB}$ , where  $T_{PRB}$  is the time duration of a single  
302 PRB.

### 303 4.2.2. User utility function based on fuzzy logic

304 The function defining the user utility  $u_i(t, k, \varphi) = F_{\text{fuzzy}}(v_i(t, k), r_i(t, \varphi))$  is derived by applying  
305 the following fuzzy rule base.

- 306 1. Rule1: If significance  $v_i(t, k)$  is high then user\_utility is high
- 307 2. Rule2: If significance  $v_i(t, k)$  is low then user\_utility is low
- 308 3. Rule3: If rate  $r_i(t, \varphi)$  is high then user\_utility is high
- 309 4. Rule4: If rate  $r_i(t, \varphi)$  is low then user\_utility is low



310 The fuzzy inference is applied to every user  $i$ /PRB  $\varphi$  pair for each iteration  $k$  of time slot  $t$ , where PRB  
 311  $\varphi \in \Phi_{\text{URB}}(t, k)$ .

312 Let  $V_{\text{high}}$  and  $V_{\text{low}}$  denote fuzzy significance sets over the universe of discourse of *significance*,  
 313 representing *high* and *low significance* set, respectively. Let  $U_V(V)$  denote a fuzzy singleton  
 314 consequent over the universe of discourse of *significance*, where  $U_V(V_{\text{high}})$  and  $U_V(V_{\text{low}})$  represent  
 315 the *high* and *low* consequents of *Rule1* and *Rule2*, respectively. Let  $\mu_V(v_i(t, k))$  denote the degree of  
 316 membership or membership function of *significance* with a crisp value  $v_i(t, k)$  in the fuzzy set  $V \in$   
 317  $\{V_{\text{high}}, V_{\text{low}}\}$ . Let  $R_{\text{high}}$  and  $R_{\text{low}}$  denote fuzzy rate sets over the universe of discourse of *significance*,  
 318 representing *high* and *low rate* set, respectively. Let  $U_R(R)$  denote a fuzzy singleton consequent  
 319 over the universe of discourse of *rate*, where  $U_R(R_{\text{high}})$  and  $U_R(R_{\text{low}})$  represent the *high* and *low*  
 320 consequents of *Rule3* and *Rule4*, respectively. Let  $\mu_R(r_i(t, \varphi))$  denote the degree of membership or  
 321 membership function of *rate* with a crisp value  $r_i(t, \varphi)$  in the fuzzy set  $R \in \{R_{\text{high}}, R_{\text{low}}\}$ .

Applying a zero-order Sugeno fuzzy inference results in a crisp value for *user\_utility*  $u$  expressed  
 as

$$u_i(t, k, \varphi) = F_{\text{fuzzy}}(v_i(t, k), r_i(t, \varphi)) \equiv \frac{\sum_V \mu_V(v_i(t, k, \varphi)) U_{r_m V}(V) + \sum_R \mu_R(r_i(t, k, \varphi)) U_R(R)}{\sum_V \mu_V(v_i(t, k, \varphi)) + \sum_R \mu_R(r_i(t, k, \varphi))} \quad (3)$$

The above expression is simplified by selecting membership functions such that  
 $\sum_V \mu_V(v_i(t, k, \varphi)) = 1$  and  $\sum_R \mu_R(r_i(t, k, \varphi)) = 1$  then

$$u_i(t, k, \varphi) \equiv 0.5 \left[ \sum_V \mu_V(v_i(t, k, \varphi)) U_V(V) + \sum_R \mu_R(r_i(t, k, \varphi)) U_R(R) \right] \quad (4)$$

The expression is further simplified by setting  $U_V(V) = U_R(R) = 0$  for  $V = V_{\text{low}}$  and  $R = R_{\text{low}}$ .

$$u_i(t, k, \varphi) \equiv 0.5 [\mu_V(v_i(t, k, \varphi)) U_V(V) + \mu_R(r_i(t, k, \varphi)) U_R(R)], \text{ for } V = V_{\text{high}} \text{ and } R = R_{\text{high}} \quad (5)$$

Finally, let  $U_V(V) = \alpha$  and  $U_R(R) = 1 - \alpha$ , where  $V = V_{\text{high}}$  and  $R = R_{\text{high}}$ , then:

$$u_i(t, k, \varphi) \equiv 0.5 [\mu_V(v_i(t, k, \varphi)) \alpha + \mu_R(r_i(t, k, \varphi)) (1 - \alpha)], \text{ for } V = V_{\text{high}} \text{ and } R = R_{\text{high}}, \quad (6)$$

322 where  $\alpha$  is referred to as the *utility coefficient* and determines the trade-off between content and channel  
 323 awareness.

324 Linear membership functions are used. The membership functions  $\mu_V(v) = v$  and  $\mu_R(r) =$   
 325  $r/r_{\text{max}}$  for  $V = V_{\text{high}}$  and  $R = R_{\text{high}}$ , respectively, where  $r_{\text{max}}$  is the maximum rate in bits/s that can  
 326 be supported over a single PRB when using the highest order modulation.

### 327 4.3. Key System Parameters and Key Performance Metrics

328 There are two key system parameters for the joint multi-user content and channel aware  
 329 scheduling, namely the utility coefficient  $\alpha$  and the number of users  $N$ . Specifically, we consider  
 330 a single tagged user for observation and  $N - 1$  competing users. The tagged user is representative  
 331 of all users within a limited area of the sector. The utility coefficient  $\alpha$  determines to what extent the  
 332 scheduler prioritizes according to channel quality or content importance.

333 A novel key performance metric is proposed in this paper for evaluating the performance of  
 334 content-aware scheduling. This metric is the significance throughput  $Z_{\text{sig}}(p)$ . Other metrics used are  
 335 the bit throughput  $Z_{\text{bit}}(p)$  in bits/s and average PSNR  $Q_P(p)$ , respectively. The metrics are computed  
 336 for a tagged user occupying a limited area of the sector containing  $p\%$  of the closest users to the BS.  
 337 More elaborations and mathematical expressions on the aforementioned metrics are provided next in  
 338 Section 4.4.

#### 339 4.4. The Proposed Analytical Model

340 There are  $m$  PRBs per time-slot, where a time-slot is the allocation period for the scheduler. We  
 341 consider the allocation over a period of  $N_{TS}$  time slots, where the duration of a time-slot is  $T_{PRB}$   
 342 seconds. The time  $T_{PRB}N_{TS}$  denotes the maximum delay constraint  $T_D$  for all layers belonging to a  
 343 one or more GOPs of a user to be received. The frame rate  $R_{frame}$  is related to delay constraint as  
 344  $R_{frame} = gN_{frame}/T_D$ , where  $N_{frame}$  is the number of frames per GOP and  $g$  is the number of GOPs  
 345 sent in  $T_D$ . The GOP rate is  $g/T_D$  and is determined by how the video has been coded. We consider  
 346 a period of operation of the scheduling algorithm over the duration  $T_D$ . Figure 3 shows the time and  
 347 frequency distribution of PRBs over a single allocation period.

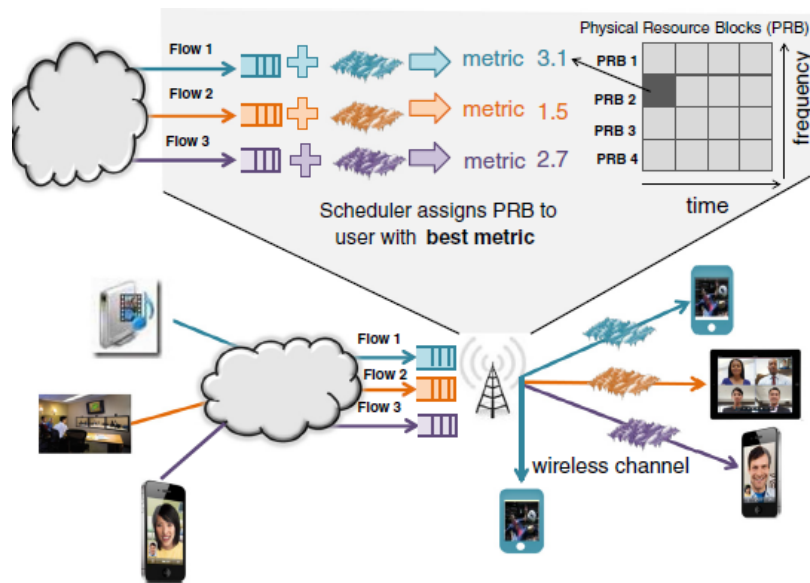


Figure 3. Time/frequency distribution of the PRBs over one allocation period.

We consider a given layer  $l$  of a tagged user  $i$  competing with the layers of  $N - 1$  other users. Associated with each layer, after the last iteration of the last slot in this period, is a set of PRBs from the pool of  $M_{PRB} = mN_{TS}$  PRBs. The set of PRBs associated with each layer represents the minimum number of an arbitrary grouping of PRBs required to send the layer to the user within the  $T_D$  time period, where feasible, or the total number of PRBs, otherwise. Each PRB from this set falls into two categories. Either the PRB was allocated to the user and used to send bits of the layer or it was not allocated due to competition. Associated with the set of PRBs is an average rate per PRB  $R_i(d)$  which is a function of the user's distance  $d$  from the BS, and the long-term shadow fading. It is assumed that short-term frequency selectivity across the set of PRBs is effectively mitigated using diversity, such as frequency domain equalization, interleaving/coding and distributed subcarrier allocation for each PRB. This assumption simplifies the analysis and though it increases the complexity of the physical layer, these suggested diversity schemes are features of advanced OFDMA standards such as LTE. We assume that MIMO is used in antenna diversity mode and not in spatial multiplexing mode, such that  $G_{mux} = 1$ .

$$R_i(d) = B \log_2(1 + e_{fade} \gamma_i(d)), \quad (7)$$

348 where  $\gamma_i(d)$  is the mean SNR and  $e_{fade}$  is the shadow fading random variable with pdf  $f_{fade}(e)$ . The  
 349 pdf of  $R_i(d)$ ,  $f_R(r, d)$ , is obtained as a transformation of  $f_{fade}(e)$ .

350 Users are assumed to be uniformly distributed in the sector, where the radius of the sector is  $d_{\text{rad}}$   
 351 and its area is  $A_{\text{sec}} = \pi d_{\text{rad}}^2 / 3$ . The area of the  $120^\circ$  sector centered at the BS occupied by  $p\%$  of the  
 352 users is  $0.01pA_{\text{sec}}$  and has a radius  $d_{\text{sec}}(p) = (0.03pA_{\text{sec}}/\pi)^{0.5}$ . The parameter  $\delta_{\text{user}} = N/A_{\text{sec}}$  is the  
 353 user density, where  $N$  is the total number of users. The distance  $d$  of a single user within the sector  
 354 defined by radius  $d_{\text{sec}}(p)$  is a random variable determined by the uniform user distribution with a  
 355 pdf  $f_d(d, p)$ , where  $d \leq d_{\text{sec}}(p) \leq d_{\text{rad}}$  and  $d_{\text{sec}}(100\%) = d_{\text{rad}}$ .

Let  $N_{\text{bits}}(l, i)$  denote the number of bits of layer  $l$  of user  $i$ , which has a probability mass function  $P_{\text{LB}}(k, l, i)$ . Let  $N_{\text{PRB}}(l, i, d)$  denote the number of PRBs required to send layer  $l$  of user  $i$ , then:

$$N_{\text{PRB}}(l, i, d) = \left\lceil \frac{N_{\text{bits}}(l, i)}{R_i(d)T_{\text{PRB}}} \right\rceil \quad (8)$$

356 Let  $P_{\text{PRB}}(k, l, i, d)$  denote the probability mass function of  $N_{\text{PRB}}(l, i, d)$ , which is derived from  
 357  $P_{\text{LB}}(k, l, i)$  and  $f_R(r, d)$ .

Let  $v_{i,l}$  denote the significance of layer  $l$  of user  $i$ , and let  $u_{i,l}(d)$  denote the utility of user  $i$  when sending bits of layer  $l$ .

$$u_{i,l}(d) = F_{\text{fuzzy}}(v_{i,l}, R_i(d)), \quad (9)$$

Let  $f_{\text{sig}}(l, i, d)$  denote the pdf of  $u_{i,l}(d)$ , which is derived from  $f_R(r, d)$ . Let  $S_{i,l}(d)$  the number of bits of layer  $l$  of user  $i$  that are transmitted before the delay constraint. Let  $i$  denote the index of the tagged user and  $l$  the index of the layer under consideration. Let  $\hat{i}$  denote the index of a competing user and  $\hat{l}$  the index of a layer of this user. Let  $S_{\text{PRB}}(l, i, d)$  denote the sum of PRBs required to send the layers of the competing users that have a higher utility than the tagged user and the layers of the tagged user up to layer  $l$ . For  $l > 0$

$$S_{\text{PRB}}(l, i, d) = \sum_{\hat{i}=0, \hat{i} \neq i}^{N-1} \sum_{\hat{l}=0}^{N_L(\hat{i})-1} N_{\text{PRB}}(\hat{l}, \hat{i}, \hat{d}) I(u_{i,\hat{l}}(\hat{d}) > u_{i,l}(d)) - \sum_{\lambda=0}^{l-1} N_{\text{PRB}}(\lambda, i, d), \quad (10)$$

where the function  $I(\text{condition})$  equals 0, if *condition* is *false* and equals 1, otherwise. For  $l = 0$

$$S_{\text{PRB}}(l, i, d) = \sum_{\hat{i}=0, \hat{i} \neq i}^{N-1} \sum_{\hat{l}=0}^{N_L(\hat{i})-1} N_{\text{PRB}}(\hat{l}, \hat{i}, \hat{d}) I(u_{i,\hat{l}}(\hat{d}) > u_{i,l}(d)). \quad (11)$$

The difference  $M_{\text{PRB}} - S_{\text{PRB}}(l, i, d)$  determines the number of PRBs available to send layer  $l$  of the tagged user. If this difference is zero or negative, then no bits of the layer are sent. If it is non-zero, positive, and less than  $N_{\text{PRB}}(l, i, d)$ , then some but not all bits of the layer are sent. If it is non-zero, positive, and equal to or more than  $N_{\text{PRB}}(l, i, d)$ , then all bits of the layer are sent. Therefore,

$$S_{i,l}(d) = \begin{cases} 0 & \text{if } M_{\text{PRB}} - S_{\text{PRB}}(l, i, d) \leq 0 \\ R_i(d) E \left\{ M_{\text{PRB}} - S_{\text{PRB}}(l, i, d) \right\} & \text{if } 0 < M_{\text{PRB}} - S_{\text{PRB}}(l, i, d) \leq N_{\text{PRB}}(l, i, d) \\ R_i(d) E \left\{ N_{\text{PRB}}(l, i, d) \right\} & \text{if } M_{\text{PRB}} - S_{\text{PRB}}(l, i, d) > N_{\text{PRB}}(l, i, d) \end{cases} \quad (12)$$

The significance throughput  $S_{\text{sig}}(d)$  of a user at distance  $d$  from its BS

$$S_{\text{sig}}(d) = \frac{1}{N_L(i)} \sum_{l=0}^{N_L(i)-1} \frac{S_{i,l}(d)}{R_i(d) E \{ N_{\text{PRB}}(l, i, d) \}}, \quad (13)$$

The bit throughput  $S_{\text{bit}}(d)$  of a user at distance  $d$  from its BS

$$S_{\text{bit}}(d) = \sum_{l=0}^{N_L(i)-1} \frac{1}{T_D} S_{i,l}(d), \quad (14)$$

The average PSNR  $Q(d)$  of a user at distance  $d$  from its BS

$$Q(d) = \sum_{l=0}^{N_L(i)-1} (q_l - q_{l-1}) \frac{S_{i,l}(d)}{R_i(d) \mathbb{E}\{N_{\text{PRB}}(l, i, d)\}}, \quad (15)$$

358 where  $q_l$  is the average PSNR if all  $l + 1$  layers have been received without error and equals zero for  
359  $l < 0$ .

The significance throughput  $Z_{\text{sig}}(p)$  of a user among the  $p\%$  of users closest to the BS

$$Z_{\text{sig}}(p) = \int_0^{d_{\text{sec}}(p)} f_d(d, p) S_{\text{sig}}(d) dd, \quad (16)$$

The bit throughput  $Z_{\text{bit}}(p)$  of a user among the  $p\%$  of users closest to the BS

$$Z_{\text{bit}}(p) = \int_0^{d_{\text{sec}}(p)} f_d(d, p) S_{\text{bit}}(d) dd, \quad (17)$$

The average PSNR  $Q_P(p)$  of a user among the  $p\%$  of users closest to the BS

$$Q_P(p) = \int_0^{d_{\text{sec}}(p)} f_d(d, p) Q(d) dd. \quad (18)$$

## 360 5. Numerical Results and Analysis

361 All users are assumed to be streaming video sequences with identical traffic and quality statistics.  
362 Specifically, statistics of the first hour of the *Tokyo Olympics* video (133 128 frames at 30 frames/sec)  
363 [38] are used. Its traffic statistics, quality statistics, and trace are publicly available at [39]. The video  
364 sequence is in the Common Intermediate Format (CIF, 352 x 288 pixels). We consider the temporal  
365 layers embedded in the video stream encoded with H.264 SVC, with a GOP structure  $\{K_0 B_2 B_1 B_2 K_0 \dots\}$ ,  
366 where  $K_0$  is an I or P key picture. Thus,  $N_L = 3$  and  $l \in \{0, 1, 2\}$ . The probability distribution  
367  $P_{\text{LB}}(k, l, i)$  of layer  $l$  and the set of average PSNR values  $q_l$  are obtained from [39]. In [39], two values  
368 are given for the average PSNR of the key picture, namely, 27.31 dB and 26.94 dB for the I frame and  
369 P frame, respectively. In this study, we use the lower value for both types of key pictures and set  
370  $q_0 = 26.94$  dB, while  $q_1$  and  $q_2$  equal 28.43 dB and 29.32 dB, respectively.

371 For the wireless system, parameter values are taken mostly from [40,41]. The channel is assumed  
372 to be flat in time and frequency due to the OFDM modulation and the effective exploitation of  
373 diversity in the time and frequency domains. Independent lognormal shadow fading with pdf  
374  $f_{\text{fade}}(e)$  and a standard deviation of 8 dB has been assumed. The values for the BS antenna gain,  
375 UE antenna gain, UE noise figure and total sector TX power are 14 dBi, 0 dBi, 7 dB and 46 dBm,  
376 respectively. The time-slot duration  $T_{\text{PRB}}$  is 0.5 ms. The coverage of the sector has a radius of 250m.  
377 The maximum number of PRBs per sector  $m$  is 34 per slot, with each PRB having a bandwidth of  
378 180 kHz. Furthermore, we assume a maximum spectral efficiency of 6 bits/s/Hz in each PRB, for  
379 64QAM modulation without MIMO spatial multiplexing. Therefore, the maximum bit rate per PRB  
380  $r_{\text{max}}$  is 1080 kbits/s. A 2x2 MIMO antenna diversity gain of 6 dB is assumed. The distance-dependent  
381 path gain is given by  $-128.1 - 37.6 \log_{10}(d)$ .

382 Of all the statistical distributions used in the analytical model, the distributions  $P_{\text{LB}}(k, l, i)$  and  
383  $f_{\text{fade}}(e)$  are all given, the former is obtained empirically [39], while the latter is assumed to be

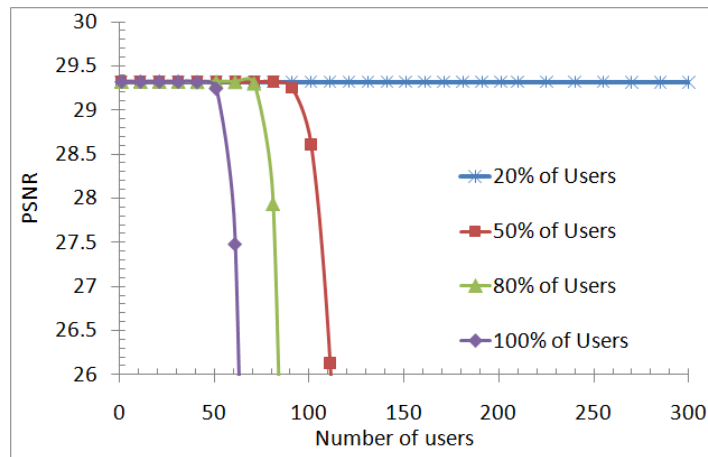
**Table 1.** Table of symbols.

$Q(d), S_{\text{sig}}(d), S_{\text{bit}}(d)$	Average PSNR, Significance throughput and Bit throughput, respectively, for a user at distance $d$ from the BS
$Q_P(p), Z_{\text{sig}}(p)$ and $Z_{\text{bit}}(p)$	Average PSNR, Significance throughput and Bit throughput, respectively, of a user among the $p\%$ of users closest to the BS
$q_l$	Average PSNR if all $l + 1$ layers have been received without error
$S_{i,l}(d)$	Number of bits of layer $l$ of user $i$ that are transmitted before the delay constraint.
$v_{i,l}$	Significance of layer $l$ of user $i$
$R_i(d), f_R(r, d)$	Average rate per PRB $R_i(d)$ and its pdf as a function of the the user's distance $d$ from the BS
$u_{i,l}(d)$	Utility of user $i$ when sending bits of layer $l$ .
$f_{\text{sig}}(l, i, d)$	The pdf of $u_{i,l}(d)$
$N_{\text{PRB}}(l, i)(d)$	Number of PRBs required to send layer $l$ of user $i$
$N_{\text{bits}}(l, i)$	Number of bits of layer $l$ of user $i$ ,
$d_{\text{rad}}$	Radius of cell sector
$A_{\text{sec}}$	Area of cell sector
$d_{\text{sec}}(p)$	Radius of a sector centered at the BS occupied by $p\%$ of the users
$f_d(d, p)$	The pdf of the distance $d$ of a single user within the sector defined by radius $d_{\text{sec}}(p)$ .
$T_{\text{PRB}}$	The time duration of a single PRB or time-slot
$R_{\text{frame}}$	Frame rate
$N_7$	Number of frames per GOP
$N_{\text{TS}}, M_{\text{PRB}}, T_{\text{D}}$	Maximum number of time-slots, maximum number of PRBs and maximum delay constraint or duration, respectively, to send all layers of a GOP
$m$	Number of PRBs per time-slot
$i, l$	Indexes of the tagged user and tagged layer, respectively.
$\hat{i}, \hat{l}$	Indexes of a competing user and layer, respectively.
$\gamma_i(d)$	Mean SNR of a user at distance $d$ from the BS.
$\alpha$	Utility coefficient
$N_{\text{PRB}}(l, i, d), P_{\text{PRB}}(k, l, i, d)$	Number of PRBs required to send layer $l$ of user $i$ , and its probability mass function, respectively
$N_{\text{bits}}(l, i), P_{\text{LB}}(k, l, i)$	Number of bits of layer $l$ of user $i$ and its probability mass function, respectively
$N$	Number of users
$e_{\text{fade}}, f_{\text{fade}}(e)$	Shadow fading random variable and its pdf, respectively.

384 lognormal. The pdf of  $f_d(d, p)$  is derived from a transformation, given that  $d^2$  is uniformly distributed  
385 over the range  $(0, d_{\text{sec}}(p)]$ . The other distributions mentioned above are derived from one or more of  
386 these three distributions, and are obtained numerically from Monte-Carlo simulations.

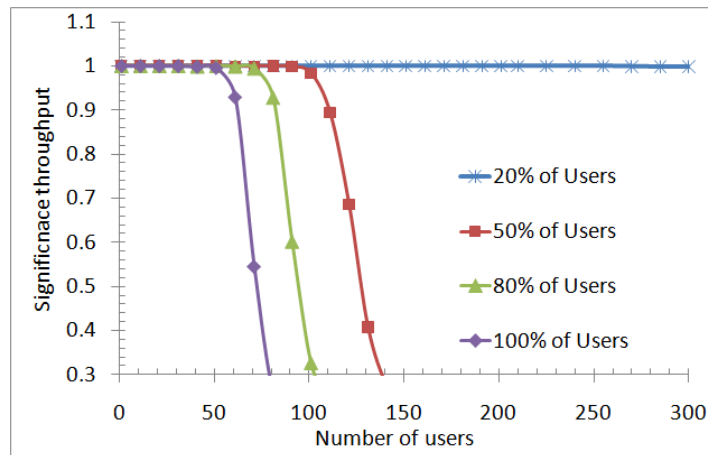
387 Figure 4 shows a plot of PSNR versus number of users for different classes of users and for  
388  $\alpha = 0$ , where users are classified according to the region they occupy in the sector. This scenario  
389 corresponds to channel-aware only scheduling. The results show that the closest 20% of users achieve  
390 the maximum PSNR performance over the entire observed range. The PSNR deteriorates at a rapid  
391 rate the further the range of users considered.

392 Figure 5 shows the corresponding plot using the significance throughput metric for different  
393 classes of users and for  $\alpha = 0$ . The results show that the closest 20% of users achieve the maximum  
394 significance throughput of unity, hence the maximum PSNR performance, over the entire observed  
395 range. A frame rate of 30fps is considered, which maps to a maximum delay constraint of 0.1333 secs  
396 to deliver all the layers comprising the the 4 frames of the GOP. The significance throughput follows  
397 the same trend as the PSNR and deteriorates at a rapid rate the further the range of users considered.



**Figure 4.** PSNR versus number of users for different classes of users and  $\alpha = 0$ .

Figure 6 and Figure 7 shows corresponding plots using the significance throughput for  $\alpha = 0.25$  and



**Figure 5.** Significance throughput versus number of users for different classes of users and  $\alpha = 0$ .

398

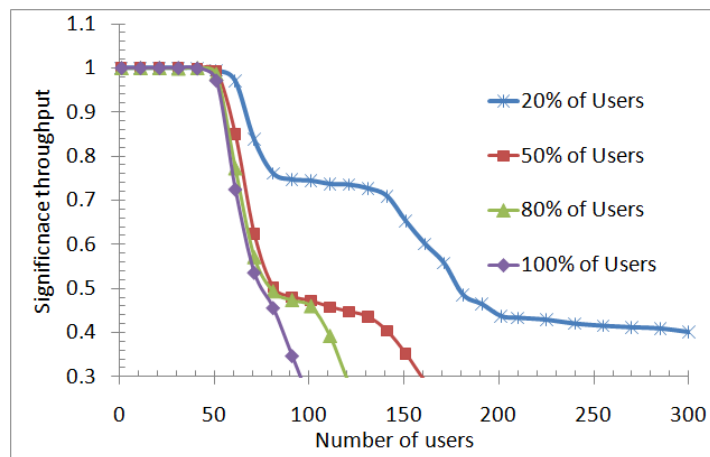
399  $\alpha = 0.5$ , respectively. These plots show the impact of introducing content-awareness, while reducing  
 400 proportionately the extent of channel awareness. The results show that the performance of the closest  
 401 20% of users declines at a rapid rate as  $\alpha$  is increased from zero to 0.5, while the performance of the  
 402 users measured over larger distance ranges improves at a slower rate. For  $\alpha = 0.5$ , the performance  
 403 across all distance ranges have converged significantly.

404

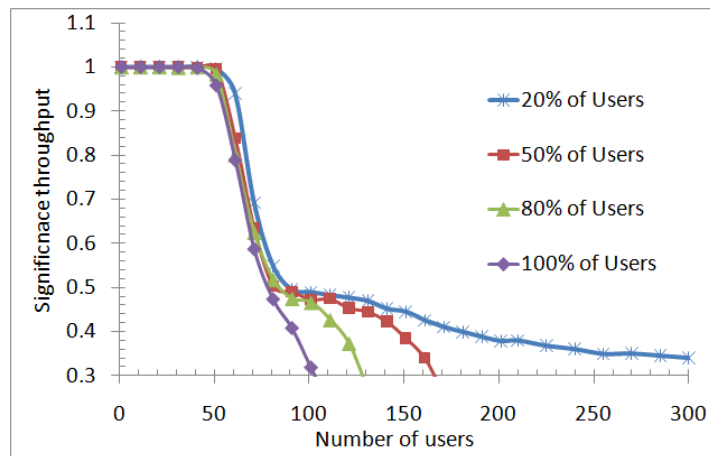
The objective of introducing content awareness is to improve the fairness of the proposed  
 405 scheduling algorithm, in general, and particularly to enhance the performance of distant users at  
 406 a minimum penalty to users close to the BS. For illustration, we consider minimum significance  
 407 throughput target for the closest 20% of users to be  $\frac{2}{3}$ , which corresponds to receiving the most  
 408 important two out of the three layers. For the closest 100% of users, that is all users, we consider  
 409 minimum significance to be  $\frac{1}{3}$ , which corresponds to receiving the most important one out of the  
 410 three layers. With these constraints, the maximum number of users that can be supported increases  
 411 from 78% to 90% (a 15% enhancement), as  $\alpha$  increases from 0 to 0.25. It declines to 72% as  $\alpha$   
 412 increases from 0.25 to 0.5.

413

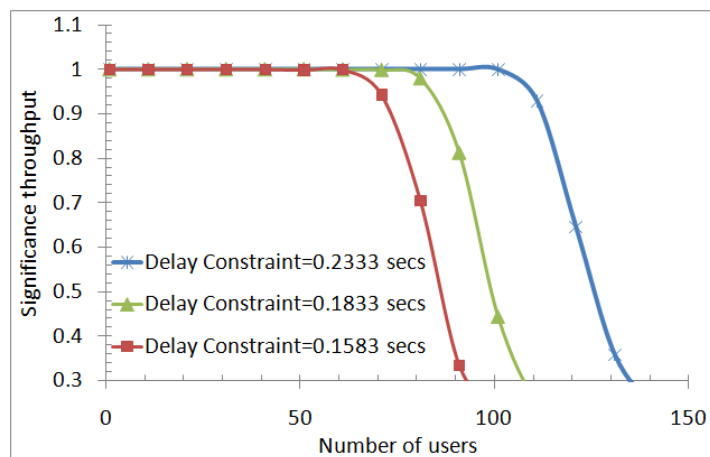
Figure 8, Figure 9 and Figure 10 show plots of significance throughput versus number of  
 414 users for different maximum delay constraints, corresponding to slightly reduce frame rates. Since  
 415 the playback rate is constant at 30fps, reducing the frame-rate at scheduler implies that the video  
 416 sequence will experience short pauses. the shorter the pauses the less perceptible to the user. Figure 8



**Figure 6.** Significance throughput versus number of users for different classes of users and  $\alpha = 0.25$ .



**Figure 7.** Significance throughput versus number of users for different classes of users and  $\alpha = 0.5$ .



**Figure 8.** Significance throughput versus number of users for different maximum delay constraints, 100% of users and  $\alpha = 0$ .

417 and Figure 10 show the case for 100% of the users with  $\alpha$  equal to zero and 0.25, respectively, while  
 418 Figure 9 shows the case for 20% of the users and  $\alpha = 0.25$ .

419 The results show that increasing the delay constraint increases the significance throughput, and  
 420 hence PSNR, for all cases. The results show that significant performance improvement is possible  
 421 for small increases in delay, or equivalently small reductions in frame rate. Consider an increase in  
 422 the delay constraint from 0.1333 secs to 0.1583 secs, corresponding to a reduction in frame rate at  
 423 the scheduler from 30fps to 25.27fps. Given the constraints on the minimum significance throughput  
 424 for the closest 20% and 100% users mentioned above, the maximum number of users that can be  
 425 supported increases from 91% to 109% (a 19.8% enhancement), as  $\alpha$  increases from 0 to 0.25.

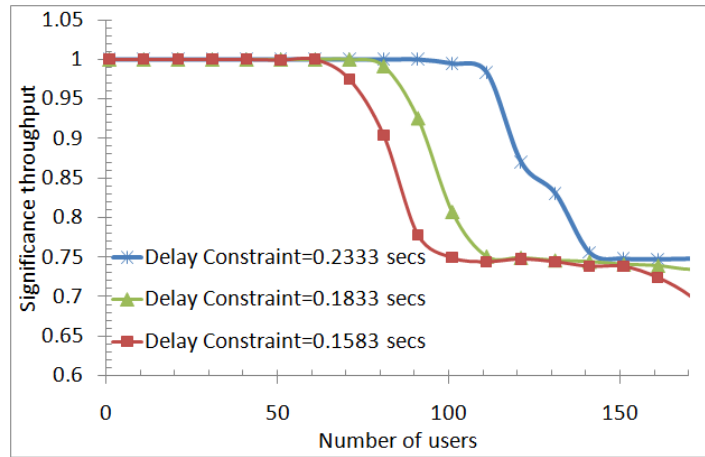


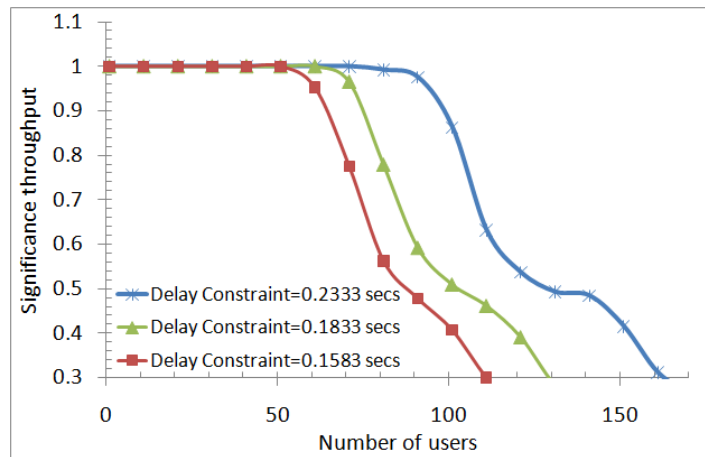
Figure 9. Significance throughput versus number of users for different maximum delay constraints, 20% of users and  $\alpha = 0.25$ .

Table 2. Simulation results for content-aware and content-unaware scheduling strategies.

Parameters	Fuzzy-based Scheduler					PF-Scheduler					M-LWDF-Scheduler				
Channel Aware	x					x					x				
Delay Aware											x				
Content Aware	x														
Results															
Metrics	Network Load (Users)					Network Load (Users)					Network Load (Users)				
	8	12	16	20	24	8	12	16	20	24	8	12	16	20	24
PSNR (dB)	35.8	34.9	32	30.7	29	31.8	25.3	21.5	20.8	20.8	37.1	30.8	21.5	20.8	20.8

426 Table 2 shows the simulation results for content-aware and content-unaware scheduling  
 427 strategies. A comparison between the proposed fuzzy-based scheduler and the standard schedulers  
 428 has been carried out, as these schedulers provide fairly good performance in terms of PSNR and  
 429 number of users. The simulation framework and the channel model parameters are the same as in  
 430 our previous study [2]. Furthermore, we also select the same SVC videos as in our previous study  
 431 [2]. The SVC layers of different video contents are marked with a priority index according to the QoE  
 432 based marking algorithm in [11]. As a benchmark strategy, we utilize the Proportional Fair (PF) and  
 433 M-LWDF schedulers. The simulation results of the proposed and benchmark strategies are reported  
 434 in Table 2. According to the table, the proposed fuzzy based scheduler achieves a cumulative video  
 435 quality of 35.8 dB when the total number of video streaming users is 8. On the other hand, PF and  
 436 M-LWdF schedulers achieves a video quality of 31.8 dB and 37.1 dB respectively. The increase in  
 437 load (in terms of the total number of streaming users) decreases the cumulative video quality of all  
 438 the strategies. However, the degradation in video quality of the proposed fuzzy-based scheduler is  
 439 lower as compared to the benchmark strategies. This is mainly because the fuzzy-based scheduler





**Figure 10.** Significance throughput versus number of users for different maximum delay constraints, 100% of users and  $\alpha = 0.25$ .

440 prioritizes the most important SVC layers. Therefore, layers contributing highest to the QoE are  
 441 scheduled before their deadline. The increase in load prioritizes the most important SVC layer of  
 442 each user. On the other hand, PF and M-LWDF strategies assign radio resources to the SVC layers  
 443 irrespective of their quality contributions, which results in a steep fall in the quality, as shown in Table  
 444 2, when the number of users is increased from 8 to 24.

## 445 6. Conclusions

446 A novel intelligent fuzzy logic-based content and channel aware downlink scheduler for scalable  
 447 video streaming has been proposed. Using novel content-aware and standard performance metrics,  
 448 the performance of the proposed scheduling algorithm through the analytical model is evaluated.  
 449 The fuzzy logic controller allows a single utility parameter to be defined and use the trade-off  
 450 between content and channel-awareness in order to enhance the overall user experience throughout  
 451 the coverage area. The results show that the number of supported users can be enhanced by as  
 452 much as 15%, for playback without pauses and as much as 19% if short imperceptible pauses are  
 453 acceptable. Significantly, the results demonstrate that channel-aware only and content-aware only  
 454 schemes are inadequate for supporting video services in a cellular environment. The former delivers  
 455 disproportionately good quality to users close to the BS, while users at the sector edge are unable  
 456 to meet a minimum quality. The latter significantly penalises users with good channels, while the  
 457 performance of edge users, though improved, remains minimal. The proposed algorithm allows  
 458 for a fairer allocation of resources among users across the entire sector coverage, allowing for the  
 459 enhancement of video quality at edges of the cell while minimizing the degradation to users closer  
 460 to the BS. Future work will consider heterogeneous video traffic and the sensitivity to different fuzzy  
 461 rule bases and membership functions for the fuzzy-controller. In addition, a performance analysis  
 462 between our proposed scheduling algorithm and other content and channel aware scheduling  
 463 algorithms will be considered.

464 **Acknowledgments:** Dr. Peter Omiyi would like to acknowledge the Holon Institute for its support. Also,  
 465 Dr. Moustafa Nasralla would like to acknowledge the management of Prince Sultan University (PSU) and  
 466 the Renewable Energy Lab for the valued support and research environmental provision which have led to  
 467 completing this work. Moreover, Dr. Ikram Ur Rehman would like to acknowledge the West London University  
 468 for its support. Finally, Prof. Maria Martini would like to acknowledge Kingston University for its support.

469 **Conflicts of Interest:** The authors declare no conflict of interest.

470

## 471 Bibliography

- 472 1. Nasralla, M.M. A Hybrid Downlink Scheduling Approach for Multi-Traffic Classes in LTE Wireless  
473 Systems. *IEEE Access* **2020**, *8*, 82173–82186.
- 474 2. Nasralla, M.M.; Khan, N.; Martini, M.G. Content-aware downlink scheduling for LTE wireless systems:  
475 A survey and performance comparison of key approaches. *Computer Communications* **2018**, *130*, 78 – 100.
- 476 3. Nasralla, M.M.; Razaak, M.; Rehman, I.U.; Martini, M.G. Content-aware packet scheduling strategy for  
477 medical ultrasound videos over LTE wireless networks. *Computer Networks* **2018**, *140*, 126–137.
- 478 4. Rehman, I.U.; Nasralla, M.M.; Philip, N.Y. Multilayer perceptron neural network-based QoS-aware,  
479 content-aware and device-aware QoE prediction model: a proposed prediction model for medical  
480 ultrasound streaming over small cell networks. *Electronics* **2019**, *8*, 194.
- 481 5. Xu, Z.; Cao, Y.; Wang, W.; Jiang, T.; Zhang, Q. Incentive Mechanism for Cooperative Scalable Video  
482 Coding (SVC) Multicast Based on Contract Theory. *IEEE Transactions on Multimedia* **2019**.
- 483 6. Ghermezcheshmeh, M.; Shah-Mansouri, V.; Ghanbari, M. Analysis and performance evaluation of  
484 scalable video coding over heterogeneous cellular networks. *Computer Networks* **2019**, *148*, 151–163.
- 485 7. van der Schaar, M.; Andreopoulos, Y.; Hu, Z. Optimized Scalable Video Streaming over IEEE 802.11a/e  
486 HCCA Wireless Networks under Delay Constraints. *IEEE Trans. on Mobile Computing* **2006**, *5*, 755–768.
- 487 8. Pahalawatta, P.V.; Berry, R.; Pappas, T.N.; Katsaggelos, A.K. Content-aware resource allocation and packet  
488 scheduling for video transmission over wireless networks. *IEEE Journal on Sel. Areas on Commun.* **2007**,  
489 *25*, 749–759.
- 490 9. Martini, M.G.; Tralli, V. Video quality based adaptive wireless video streaming to multiple users. Proc.  
491 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting; , 2008; pp. 1–4.
- 492 10. Fu, B.; Staehle, D.; Kunzmann, G.; Steinbach, E.; Kellerer, W. QoE-aware Priority Marking and Traffic  
493 Management for H.264/SVC-based Mobile Video Delivery. 8th ACM workshop on Performance  
494 monitoring and measurement of heterogeneous wireless and wired networks; , 2013; pp. 173–180.
- 495 11. Fu, B.; staehle, D.; Kunzmann, G.; Steinbach, E.; Kellerer, W. QoE-based SVC layer dropping  
496 in LTE networks using content-aware layer priorities. *ACM Transactions on Multimedia Computing,*  
497 *Communications, and Applications (TOMM)* **2015**, *12*, 23.
- 498 12. Zhang, Y.; Liu, G. Fine granularity resource allocation algorithm for video transmission in orthogonal  
499 frequency division multiple access system. *IEEE IET (Institution of Engineering and Technology)*  
500 *Communications* **2013**, *7*, 1383–1393.
- 501 13. Li, F.; Ren, P.; Du, Q. Joint Packet Scheduling and Subcarrier Assignment for Video Communications  
502 Over Downlink OFDMA Systems. *IEEE Transactions on Vehicular Technology* **July. 2012**, *61*, 2753–2767.
- 503 14. Li, F.; Zhang, D.; Wang, M. Multiuser multimedia communication over orthogonal frequency-division  
504 multiple access downlink systems. *Concurrency and Computation: Practice and Experience* **2013**,  
505 *25*, 1081–1090.
- 506 15. Li, P.; Chang, Y.; Feng, N.; Yang, F. A Cross-Layer Algorithm of Packet Scheduling and Resource  
507 Allocation for Multi-User Wireless Video Transmission. *IEEE Transactions on Consumer Electronics* **Sept.**  
508 **2011**, *57*, 1128–1134.
- 509 16. Ji, X.; Huang, J.; Chiang, M.; Lafruit, G.; Catthoor, F. Scheduling and Resource Allocation for SVC  
510 Streaming over OFDM Downlink Systems. *IEEE Trans. on Circuits and Systems* **2009**, *19*, 1549–1555.
- 511 17. Cicalò, S.; Tralli, V. Distortion-Fair Cross-Layer Resource Allocation for Scalable Video Transmission in  
512 OFDMA Wireless Networks. *IEEE Transactions on Multimedia* **April. 2014**, *16*, 848–863.
- 513 18. Blej, M.; Azizi, M. Comparison of Mamdani-type and Sugeno-type fuzzy inference systems for fuzzy real  
514 time scheduling. *International Journal of Applied Engineering Research* **2016**, *11*, 11071–11075.
- 515 19. Kaur, A.; Kaur, A. Comparison of Mamdani-type and Sugeno-type fuzzy inference systems for air  
516 conditioning system. *International journal of soft computing and engineering* **2012**, *2*, 323–325.
- 517 20. Sivanandam, S.; Sumathi, S.; Deepa, S.; others. *Introduction to fuzzy logic using MATLAB*; Vol. 1, Springer,  
518 2007.
- 519 21. Jang, J.S. ANFIS: adaptive-network-based fuzzy inference system. *IEEE transactions on systems, man, and*  
520 *cybernetics* **1993**, *23*, 665–685.
- 521 22. Schwarz, H.; Marpe, D.; Wiegand, T. Overview of the scalable video coding extension of the H.264/AVC  
522 standard. *IEEE Transactions on Circuits and Systems for Video Technology* **2007**, *17*, 1103–1120.

- 523 23. Khan, N.; Nasralla, M.M.; Martini, M. Network and User Centric Performance Analysis of Scheduling  
524 Strategies for Video Streaming over LTE. *IEEE International Conference on Communications (ICC) -*  
525 *Workshop on Quality of Experience-based Management for Future Internet Applications and Services*  
526 *(QoE-FI)*; , 2015.
- 527 24. Khan, N.; Martini, M.G. QoE-driven multi-user scheduling and rate adaptation with reduced cross-layer  
528 signaling for scalable video streaming over LTE wireless systems. *EURASIP Journal on Wireless*  
529 *Communications and Networking* **2016**, *2016*, 93.
- 530 25. Maani, E.; Pahalawatta, P.V.; Berry, R.; Katsaggelos, A.K. Content-aware packet scheduling for multiuser  
531 scalable video delivery over wireless networks. *SPIE Optical Engineering and Applications*; , 2009.
- 532 26. Pahalawatta, P.; Berry, R.; Pappas, T.; Katsaggelos, A. Content-aware resource allocation and  
533 packet scheduling for video transmission over wireless networks. *IEEE Journal on Selected Areas in*  
534 *Communications* **2007**, *25*, 749–759.
- 535 27. Khan, N.; Martini, M.G.; Bharucha, Z. Quality-aware fair downlink scheduling for scalable video  
536 transmission over LTE systems. *IEEE International Workshop on Signal Processing Advances in Wireless*  
537 *Communications (SPAWC)*; , 2012; pp. 334–338.
- 538 28. Khan, N.; Martini, M.G.; Staehle, D. Opportunistic Proportional Fair Downlink Scheduling for Scalable  
539 Video Transmission over LTE Systems. *IEEE Vehicular Technology Conference (VTC)*; , 2013.
- 540 29. Mostafa, A.E.; Gadallah, Y. A statistical priority-based scheduling metric for M2M communications in  
541 LTE networks. *IEEE Access* **2017**, *5*, 8106–8117.
- 542 30. Zhang, W.; Ye, S.; Li, B.; Zhao, H.; Zheng, Q. A priority-based adaptive scheme for multi-view live  
543 streaming over HTTP. *Computer Communications* **2016**, *85*, 89–97.
- 544 31. Edelman, B.A.; Gay, J.; Lozben, S.; Shetty, P. Real-time priority-based media communication, 2018. US  
545 Patent 9,900,361.
- 546 32. Khan, N.; Martini, M. Hysteresis based Rate Adaptation for Scalable Video Traffic over an LTE Downlink.  
547 *IEEE International Conference on Communications (ICC) - Workshop on Smart Communication Protocols*  
548 *and Algorithms*; , 2015.
- 549 33. Khan, N. Quality-Driven Multi-User Resource Allocation and Scheduling Over LTE for Delay Sensitive  
550 Multimedia Applications. In *Ph.D. Dissertation*; Kingston University London, UK, 2014.
- 551 34. Radhakrishnan, S.; Neduncheliyan, S.; Thyagarajan, K. A novel fuzzy scheduler for cell-edge users in  
552 LTE-advanced networks using Voronoi algorithm. *Cluster Computing* **2017**, pp. 1–11.
- 553 35. Abrahão, D.C.; Vieira, F.H.T. Resource allocation algorithm for LTE networks using fuzzy based adaptive  
554 priority and effective bandwidth estimation. *Wireless Networks* **2018**, *24*, 423–437.
- 555 36. Taki, M.; Heshmati, M.; Omid, Y. Fuzzy-based optimized QoS-constrained resource allocation in a  
556 heterogeneous wireless network. *International Journal of Fuzzy Systems* **2016**, *18*, 1131–1140.
- 557 37. Mardani, M.R.; Ghanbari, M. Robust resource allocation scheme under channel uncertainties for LTE-A  
558 systems. *Wireless Networks* **2019**, *25*, 1313–1325.
- 559 38. der Auwera, G.V.; David, P.T.; Reisslein, M.; Karam, L.J. Traffic and Quality Characterization of the  
560 H.264/AVC Scalable Video Coding Extension. *Eurasip Journal on Advances in Multimedia* **2008**, *25*.
- 561 39. <http://trace.eas.asu.edu/>. H.264/AVC and SVC Video Trace Library.
- 562 40. 3GPP TR 25.814 V7.1.0. Physical Layer Aspects for Evolved Universal Terrestrial Radio Access (UTRA)  
563 (Release 7), 2006.
- 564 41. 3GPP TS 36.211 V8.2.0. EUTRA Physical Channels and Modulation (Release 8), 2008.