Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 6584845, 13 pages
https://doi.org/10.1155/2018/6584845

Research Article

Developing a Video Buffer Framework for Video Streaming in Cellular Networks

Saba Qasim Jabbar,[1,2] **Dheyaa Jasim Kadhim**,[2] **and Yu Li**[1]

[1] *Wuhan National Laboratory for Optoelectronics, Division of Communication and Intelligent Networks, School of Electronics and Information, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China*
[2] *Department of Computer Engineering, Baghdad University, Baghdad, Iraq*

Correspondence should be addressed to Saba Qasim Jabbar; shura2007515@yahoo.com

Received 21 September 2017; Revised 7 February 2018; Accepted 18 March 2018; Published 27 June 2018

Academic Editor: Mauro Femminella

This work proposes a new video buffer framework (VBF) to acquire a favorable quality of experience (QoE) for video streaming in cellular networks. The proposed framework consists of three main parts: client selection algorithm, categorization method, and distribution mechanism. The client selection algorithm was named independent client selection algorithm (ICSA), which is proposed to select the best clients who have less interfering effects on video quality and recognize the clients' urgency based on buffer occupancy level. In the categorization method, each frame in the video buffer is given a specific number for better estimation of the playout outage probability, so it can efficiently handle so many frames from different videos at different bitrates. Meanwhile, at the proposed distribution mechanism, a predetermined threshold value is selected for lower and upper levels of playout outage probability. Then, the control unit at the base station will distribute the radio resources and decide the minimum rate requirement based on clients' urgency categories. Simulation results showed that the VBF grantees fairness of resources distribution among different clients within the same cellular network while minimizing the interruption duration and controlling the video buffer at an acceptable level. Also, the results showed that the system throughput of the proposed framework outperforms other existing algorithms such as playout buffer and discontinuous reception aware scheduling (PBDAS), maximum carrier-to-interface ratio (MAX-CIR), and proportional fair (PF) due to enhancing the quality of experience for video streaming by increasing the radio resources in fairness manner.

1. Introduction

The wealth of online videos and widespread access points make people easily get Internet anywhere for any content they want. According to Cisco visual networking index for world Internet traffic, the most of traffic on the Internet is caused by video streaming. In cellular networks, a video streaming is considered as a major research problem nowadays because of the increased demand for high video quality [1, 2]. Since the wireless channel has different time features, the video streams playing process may be interrupted, which causes frames not to complete their flow into the buffer on client's device [3, 4].

However, the services of video streaming will make up 70% of the Internet movement in the near future [5], so that several techniques will be used to provide a high data throughput to support high quality of Service (QoS) for video streaming over wireless networks, such as 3GPP (long-term evolution) LTE, mmWave MIMO, and massive MIMO. These cellular technologies can offer extra grades of freedom to customers, which can be utilized to ensure reducing the noise, fade, and hardware impairments when signals from a large number of antennas are collected in common air. Consequently, it can rise the capacity by many times and it can improve energy efficiency radiating by many times, which can ensure the quality of received signal and achieve a high reliable link [6, 7].

In the cellular networks, a high video quality with playback continuity can be achieved together through (1) increasing network throughput, (2) minimizing the interruption duration, and (3) mitigating the variation of buffer occupancy

level. However, if the bandwidth allocation cannot meet the bitrate requirement, a playout outage problem will occur which will impact the video quality. To solve this problem, a lot of algorithms had been suggested and the works of these algorithms were approached into main two directions. The first direction is utilizing the potential of mobile network techniques for improving video quality (i.e., LTE, LTE-A, massive MIMO, etc.). The main advantage of this approach is that massive MIMO can enhance the network performance through allocating more antennas mobile users and hence achieving higher transmission rates, low latency, and high energy efficiency. Even that massive MIMO has the problem called pilot contamination because of the signals that can be sent from mobile users to the base station for channel estimation. However, massive MIMO can achieve better performance without increasing training overhead especially when the number of antennas increases. The second direction is adapting the bitrates for next downloaded segments through estimating network bandwidth or video buffer status. If the network bandwidth is high, the client can select the video format with high quality; otherwise the algorithm must switch to low quality of video format for avoiding video playout outage. The advantage of adaptive algorithm is to rise the video quality through meeting conflicting objectives in a way enhancing the user's viewing experience, such as choosing a group of video bitrates which are the highest feasible, a voiding unnecessary bitrate switches, and keeping the buffer content to avoid interruption of playback. The main challenge for this approach is the large bandwidth fluctuation for the video streaming, so the adaptive algorithm must respond to that fluctuation by quickly adapting the bitrate accordingly.

Authors of [8] proposed a rate adaptation algorithm based on calculating the average segment downloaded rate. The algorithm switches up or down the bitrate with the network bandwidth in aggressive way to maintain the video quality in acceptable level. The author proposed a rate adaptation algorithm which follows a conservative way for increasing the video bitrate, but when the available bandwidth drops suddenly, the algorithm aggressively decreases the video bitrate and, hence, causes a sudden decrease in the video quality. Authors of [9] proposed a playout buffer and discontinuous reception (DRX) aware scheduling scheme (PBDAS) to enhance video streaming over long-term evolution (LTE) network. This scheduler scheme can distinguish the urgency among the clients based on a metric called remaining playout time (RPT) which is proposed for estimating playout buffer status. Then, two levels of resource allocation are presented: the upper level is used for specifying the scheduling set according to RPT, while the lower level is used for allocating the resources to the clients' equipment's in the set. The proposed scheme could shorten the interruption duration and maintain the consumed power at an acceptable level. However, this scheme faced the problem of inability to allocate the most urgent clients with the delay aware scheduler like TLS (transport layer security) since the delay of longer packet transmission does not mean worse video playback continuity but makes it difficult for TLS to estimate the user urgency accurately. Also when the DRX cycle is longer than the threshold, there would be a lot of interruptions.

Authors of [10] proposed a stream switching algorithm based on encoding the raw video chain for multiple video formats with different qualities and bitrates where two control mechanisms are implemented: one for controlling the video buffer and the other for selecting the suitable video representation level under time varying condition; however what impairs this proposed algorithm is that it did not well consider the fairness among clients. In [11], an adaptation algorithm for adaptive streaming over HTTP (AAASH) is proposed for rate adaptation with multiple parameters and conditions. This algorithm consists of two phases: fast phase for increasing the buffer level to predefined threshold value and steady phase for controlling the buffer level from going back to underflow state, so the algorithm could limit the number of video quality switches. However, this approach has the following disadvantage: when the network condition is bad, the video buffer would be filled with low quality segments; when the network condition is good, the client would still stream the lower quality video since the buffer is filled with low quality segments; and only after these segments have been played, the client could stream the higher quality segments.

Hence, it can be seen from the above literature and other works such as [12, 13] that these schemes are trying to improve the video quality either through allocating resources to clients that have good channel quality or through keeping video playback continuity with low bitrates, so it may fail to improve video quality and achieve the fairness among clients, while our proposed framework comes with a new distributing mechanism for controlling the playout outage probability level relying on current buffer occupancy. The proposed framework achieves accepted balance of fairness among different clients urgency, minimizes the interruption duration, and maintains the buffer level away from underflow/overflow, hence improving the quality of experience.

This work aims to reduce the time of playout outage which is defined here as the average time in which the client does not run a video due to the emptiness of video buffer, that is, reducing the average time in which the client does not run a video because the video buffer is empty. Therefore, a new proposed video buffer framework (VBF) is suggested to improve the continuity of video playing services over cellular networks. This framework aims to keep the video buffer as not empty as possible, so we can guarantee the video playback continuity at client's device. The proposed framework is constructed from three main approaches, namely, client selection algorithm, categorization method, and distribution mechanism. This proposed framework comes with a new resource distributing mechanism which will distribute the radio resources among clients upon the level of playout outage probability and then decide the minimum rate requirement based on clients' urgency categories.

This paper is organized as follows: Section 2 will describe our system design for video streaming over cellular network. Section 3 will present our proposed video buffer framework in detail. Section 4 will show the simulation results of applying our proposed framework in video streaming over

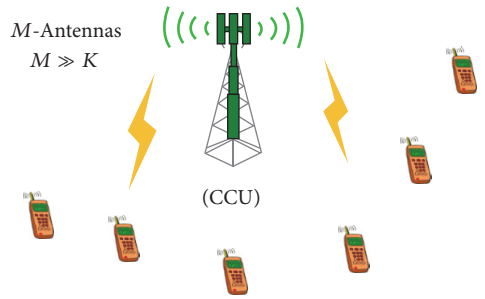

Figure 1: Modern cellular network.

cellular network. Finally, Section 5 will give the research conclusions that can be drawn from this work.

2. System Design

The network topology implemented in this work consists of one base station (i.e., central control unit (CCU)) and several clients. All the clients request the video streaming with different types of terminals, such as laptop, tablet, and smart phone. The videos were pre-encoded in H.264/AVC and stored in the video server. After being requested, the video data will be transmitted to CCU through a reliable link. At CCU, the data packets will be buffered in a specific queue. Then, all these packets are forwarded to the requesting clients using a given scheduling technique. The CCU at the base station will have a radio resource scheme and packet scheduling strategy and these two schemes consider some influencing parameters such as packet size, current channel status, client playout buffer occupancy, and the total interruption time. This work mainly focused on the resource allocation part, while the packet scheduling strategy is set to be FIFO (first-in, first-out).

2.1. Cellular Network Model. The cellular network consists of one central controller unit (CCU) equipped with large number of M antennas (i.e., massive MIMO) and K single-antenna clients assuming ($M \gg K$) as shown in Figure 1.

Let the download power (during transmission) at the base station be p_{dk} and the upload power (during transmission) at each client be p_{uk}. At the receiver, the base band signal for kth client is given as

$$y_k = h_k^H x_k + n_k \quad \text{for } k = 1, 2, 3, \ldots, K, \tag{1}$$

where $h_k \in C^{M\times 1}$ is the kth channel vector which is assumed to be a quasi-static independent and identically distributed (iid) Rayleigh fading channel and the block length of the channel is denoted by T_c, while $x \in C^{K\times 1}$ is the transmitted symbol vector with $E[xx^H] = \text{diag}(P_d) = (p_{d1}, p_{d2}, \ldots, p_{dk})$ subjected to normalized power $\|P_d\| = 1$, and each element in P is not less than 0. n_k is the noise term according to an independent complex Gaussian distribution with zero mean and unit variance $(0, 1)$ respectively. For a low complexity, we consider the use of the most common linear precoding scheme called zero-forcing (ZF) beamforming. In this case, the transmitted signal is a summation of the products formed by the desired signal and the associated precoding vector. Therefore, the received signal at the kth client can be written as

$$y_k = \sqrt{p_{dk}} h_k^H o_k s_k + \sum_{j=1, j\neq k}^{K} \sqrt{p_{dj}} h_k^H o_j s_j + i_k, \tag{2}$$

where the first term of the right side of (2) contains the desired signal for the kth client, the second term represents the interference caused by the other clients, and the last term is the background noise. $o_k \in C^{1\times K}$ denotes the kth column of the pseudoinverse channel matrix formed by the clients of the set U, and s_k represents the data symbol signal of the kth client. We assume that the CCU performs antenna selection by choosing N antennas from a set of A_n antennas among the M antennas where $N \ll M$. Since the ZF beamforming is used, supposing the instantaneous values of path gains are known at the receiver and the transmit power is uniformly allocated for the N transmit antennas, the received signal to interference plus noise ratio (SINR) for client $k \in U$ when antennas in A_n are serving the clients in U is given by

$$\Upsilon_k (k, A_n, t) = \frac{p_{dk} \left| h_k^H o_k \right|^2}{\sum_{j=1, j\neq k}^{K} p_{dj} \left| h_k^H o_j \right|^2 + \sigma^2}, \tag{3}$$

where p_{dk} is the transmit power of the kth client which must satisfy the following constraint for the selected antennas of the set A_n. h_k and o_k are the kth column in H and O matrices, respectively.

$$\sum_{k=1}^{K} p_{dk} \leq P_d, \tag{4}$$

where P_d represents the total transmit power. The achievable sum rate at the broadcast channel can be obtained by

$$\begin{aligned} R(A_n, t) &= \sum_{k=1}^{K} R_k (k, A_n, t) \\ &= \sum_{k=1}^{K} \log_2 (1 + \Upsilon_k (k, A_n, t)), \end{aligned} \tag{5}$$

where R_k is the achievable rate for kth client which is defined as a successful transmission of information bits per unit area under required connection outage or as the size of segments (in bits) divided by the segments delivery duration. The control unit for scheduling is responsible for controlling and differentiating between clients' urgency. Also, there is a FIFO cache in central control unit to request many popular videos on the web.

2.2. Client Buffer Model. The client video rate can be chosen from a set of ϖ discrete video rates $\{V_{R1}, \ldots, V_{R\varpi}\}$, where $V_{R1} < V_{R2} < \cdots < V_{R\varpi}$. We also refer to V_{R1} as $V_{R\min}$ (the minimum video rate) and $V_{R\varpi}$ as $V_{R\max}$ (the maximum video rate). Each segment contains T_{seg} seconds of video and the client can only change its chosen video rate on a segment-by-segment basis. In the proposed model, the streaming buffer in

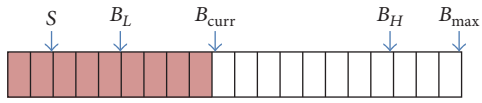

Figure 2: Client playout buffer model.

the client is typically measured in seconds of running time, the buffer may contain segments with many different video rates, and the output bitrate of the buffer will rely on the video rate of the segment currently being played. By measuring the buffer in the time domain, the client keeps a record of the number of video seconds from playing the video that is currently in the buffer without having to track the video rate associated with each video clip [14]. The client playout buffer model is shown in Figure 2.

Initially, three thresholds in the buffer level are defined for the media player at the client, namely, S, B_L, and B_H measured in seconds with $0 < B_L < B_H < B_{\max}$, where $B_{\max}$ denotes the maximum value of the buffer level and B_{curr} is the current buffer level. While $R(t)$ denotes the network capacity for download side at time t, let $B_{\text{curr}}(t)$ be the playback buffer occupancy at time t and let $V_R(t)$ denote the video rate selected at time t. Note that if the buffer is full, then no segment can be downloaded at time t (i.e., if $B_{\text{curr}}(t) = B_{\max}$, then $R(t) = 0$). Every time one second of video is removed from the buffer, another one second is played at the client player. The client seeks segments of video from the server; each segment contains a fixed duration of video (e.g., four seconds per segment). The higher video rate means that the segment size in bytes is large [15]. Now, if the adaptive video rate $V_R(t)$ is greater than system capacity $R(t)$, then new data is put into the buffer at rate $R(t)/V_R(t) < 1$ (i.e., depleting rate), so the buffer decreases, which means more than one segment is played before the next segment arrives, and then the buffer is depleted. If an adaptive bitrate algorithm is used, it may keep requesting segments with large sizes in order to sustain the network continuity (i.e., the video rate is too high); eventually the buffer will be empty quickly (depleted), so the playback stops and a message of rebuffering will appear. Other algorithms rely in their works on adjusting the capacity estimation based on the buffer occupancy [8, 10], where the client computes how fast segments reach estimate capacity $\widehat{R}(t)$ with a modification function of playback buffer $f(B_{\text{curr}}(t))$. The estimate is optionally supplemented with knowledge of the buffer occupancy and the selected video rate is $V_R(t) = f(B_{\text{curr}}(t))\widehat{R}(t)$. When the buffer contains many segments, $V_R(t)$ can safely deviate from $R(t)$ without trigger a buffering event. The client can try to maximize video quality by choosing $V_R(t) = \widehat{R}(t)$, but when the buffer is low, the client should choose a lower video rate and rapidly renew the buffer. The unnecessary probability of playout outage happens when an adaptive bitrate algorithm chooses a video rate that is higher than what the system capacity can maintain; however these interruptions could be avoidable if the algorithm chooses a lower video rate. In this case, designing the modification function is much harder as it will be shown in the following analysis. Consider the case when there is only one segment in the buffer and the requested segment (T_{seg} seconds) should reach before the current segment is played or else the buffer will run dryly. To prevent playout outage or interruption through video playing, we have

$$V_R(t) < R(t). \tag{6}$$

In terms of buffer occupancy, we have

$$\frac{T_{\text{seg}} V_R(t)}{R(t)} < B_{\text{curr}}(t), \tag{7}$$

where $T_{\text{seg}}V_R(t)$ is a segment size in bytes; expression (7) is the time needed to download the segment.

To avoid playout outage, $V_R(t)$ is replaced with $f(B_{\text{curr}}(t))\widehat{R}(t)$; then

$$f(B_{\text{curr}}(t)) < \left(\frac{B_{\text{curr}}(t)}{T_{\text{seg}}}\right)\left(\frac{R(t)}{\widehat{R}(t)}\right). \tag{8}$$

Thus, $fB_{\text{curr}}(t)$ must be smaller than the ratio $(R(t)/\widehat{R}(t))$ to prevent buffering event. A bad situation results if $f(B_{\text{curr}}(t))$ makes user pick a rate lower than the minimum video rate available, as the constraint becomes impossible to meet. From the above analysis, a video quality faces an interrupted duration which is indicated by a fraction of time when users experience buffer stalling (i.e., video paused and played again) while watching a video.

Let the playout outage probability during playout time be P_p^{out}, since the system status is expressed by the current buffer occupancy at any given time (n = number of segments in the client buffer) left behind the just served segment. We denote the steady state probabilities of the system by $P_p^{\text{out}0}, P_p^{\text{out}1}, \ldots, P_p^{\text{out}\,n}$ where $P_p^{\text{out}0}$ is directly related to the outage events, because it corresponds to the situations when there is no segment for playing out, and $P_p^{\text{out}\,n}$ is equal to zero because we observe the system just after a segment has finished service. Our analysis follows the imbedded Markov chain approach, where $P_p^{\text{out}}(B_{\text{curr}} = a)$ is the probability that, during playout time, new segments (a) exactly arrive to the client's buffer, where the probabilities of particular system can be formally written as the set of equations presented below:

$$\begin{aligned}
P_p^{\text{out}0} &= \left(P_p^{\text{out}0} + P_p^{\text{out}1}\right) \times P_p^{\text{out}}\left(B_{\text{curr}} < S = 0\right),\\
&\vdots\\
P_p^{\text{out}\,n} &= P_p^{\text{out}0} \times P_p^{\text{out}}\left(B_{\text{curr}} \le B_H = n\right) + \cdots\\
&\quad + \sum_{a=0}^{n} P_p^{\text{out}\,n+1-a} \times P_p^{\text{out}}\left(B_{\text{curr}} = a\right),\\
&\qquad\qquad \text{for } n = 1, \ldots, B_H - 2\\
\sum_{q=0}^{B_H} P_p^{\text{out}\,q} &= 1.
\end{aligned} \tag{9}$$

The first expression above corresponds to the situation where no segment is left in the client's buffer just after the

completion of the service of the previous segment. This happens when the buffer has been empty or has held just one segment and no segment has arrived during T_{seg} service time. The second expression above row describes a set of B_H expressions corresponding to nth segments remaining in the client's buffer for $n = 1, \ldots, B_H$. The last expression above is a normalization condition. The solution of the expression (9) allows calculating the rebuffering probability for a given size of the client's buffer.

In conclusion, from above, the video quality Q_{Vk} of kth client is affected by the delay due to playout outage and errors due to channel fading and interferences, so this work proposed a new scheme for video buffering to solve the following optimization problem:

$$\begin{aligned} \max \quad & \sum_{k=1}^{K} F\left(Q_{Vk}\right) Q_{Vk} \\ \text{s.t:} \quad & \lambda \overline{Q}_{Vk} \le Q_{Vk} \\ & P_{p-L}^{\text{out}} \le P_{p}^{\text{out}}\left(B_{\text{curr}}\left(k,t\right)\right) \le P_{p-U}^{\text{out}} \\ & B_L \le B_{\text{curr}} < B_H, \end{aligned} \tag{10}$$

where λ is a quality guarantee factor determined by the system. The expected level of video quality for users will be denoted as a discrete set:

$$\overline{Q}_V = \left\{\overline{Q}_{V1}, \overline{Q}_{V2}, \ldots, \overline{Q}_{Vk}\right\}, \tag{11}$$

where $F(Q_{Vk})$ is denoted as the fairness coefficient and its definition is the ratio of user's expected QoS to the logarithm of its current QoS, which could be expressed as

$$F\left(Q_{Vk}\right) = \frac{\overline{Q}_{Vk}}{\log_2 Q_{Vk}}. \tag{12}$$

We assign a lower bound and upper bound of the required playout outage probability to be P_{p-L}^{out} and P_{p-U}^{out}, respectively. Since video quality is impacted by the delay due to number of interruptions in playback video, we define a metric for video quality in terms of probability of playout outage and PSNR (i.e., PSNR is a measurement metric for objective video quality) as follows:

$$Q_{Vk} = \frac{\text{PSNR}}{P_p^{\text{out}}}, \tag{13}$$

where PSNR is the peak signal to noise ratio and its related to achievable data rate over wireless channels which is computed according to receiving SINR (Υ_k); PSNR is calculated as in [16]:

$$\text{PSNR} = \Omega \log_{10}\left(R_k\left(k, A_n, t\right)\right), \tag{14}$$

where Ω is a predefined parameter that is related to video contents. The video content-related parameter Ω is set to be between 10 and 12.

3. Proposed Video Buffer Framework

The proposed video buffer framework (VBF) consists of three main parts: client selection algorithm, categorization method, and distribution mechanism. These three approaches are described in detail in the following points.

3.1. Client Selection Algorithm. The control unit is responsible for selecting the best clients when they enter the cellular network. Then a reported buffered frames number would be sent to CCU after the client receives video frames. CCU would classify client's urgency depending on existing buffer content level as explained below.

When K clients join the cellular network at time t, they simultaneously transmit pilot sequences ψ of length τ symbols. The CCU will perform a specific algorithm for choosing the best clients who have less interfering effects through implementing this algorithm which is called independent client selection algorithm (ICSA), which is described in detail below. Let U be selected clients according to (ICSA); the pilot sequences can be represented as in [17] by $\tau \times K$ matrix, so the receiving pilot signal of K clients is given by

$$\begin{aligned} Y &= \sqrt{\tau p_u} H \Psi^T + Z \\ y_k &= \sqrt{\tau p_u} h_k \psi_k^T + \sum_{j=1, j\neq k}^{K} \sqrt{\tau p_u} h_j \psi_j^T + z_k, \end{aligned} \tag{15}$$

where $\sqrt{\tau p_u}\Psi^T$ represents the pilot sequence matrix with $\tau \times K$. The steps of ICSA algorithm are described below; for every client in the set K it calculates I_{k,A_n} which is the component of the channel matrix corresponding to the kth client, and it is orthogonal to the subspace spanned by $\{\widetilde{I}_1, \widetilde{I}_2, \ldots, \widetilde{I}_{a-1}\}$; h_{k,A_n} is used as the channel vector between kth client and antenna set A_n. After that, it finds client that maximizes norm of this subspace. As a result, the algorithm ensures that the selected client has a good level of orthogonality with another client. Therefore, the associated precoding vector of the selected client has a low value that leads to better power resource allocation. The ICSA repeats this phase till having the required number of best clients. The control unit calculates the received signal to noise ratio (SNR_k) for each client and compares it with a predetermined threshold value d_{th} for client's classification.

$$\text{SNR}_k = \tau p_u \left|h_k \psi^T\right|^2 \quad k = 1, 2, \ldots, U. \tag{16}$$

The Steps for Selecting Independent Clients Using ICSA Algorithm

Input:

Set of available antennas: $A_n \leftarrow N$

No. of joining clients: K

Initialization:

$a \leftarrow 1$; no. of iteration

$K \leftarrow \{1, 2, \ldots, k\}; U \leftarrow 0;$

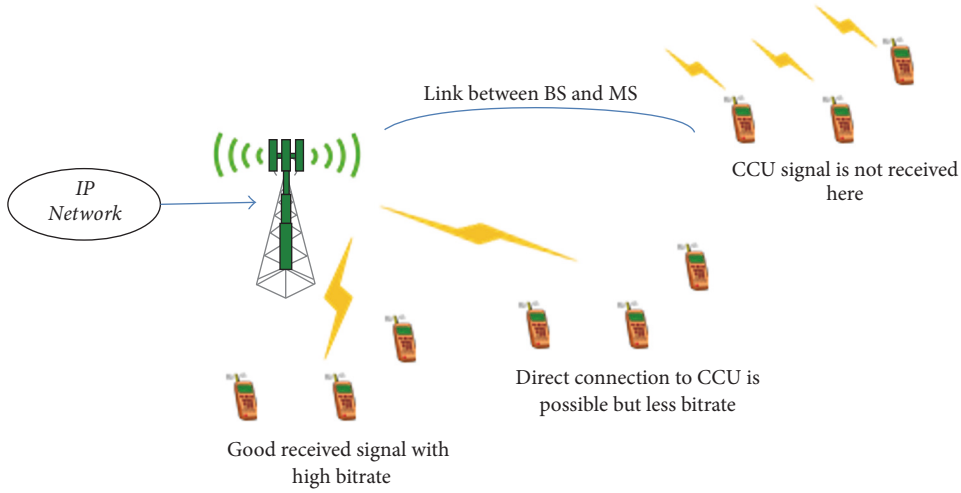

FIGURE 3: Clients' locations inside cellular network with respect to their places from CCU.

While $(a < N)$

{

For each kth in K do {

$$I_{k,A_n} = h_{k,A_n} - \sum_{j=1}^{a-1}((h_{k,A_n} - \tilde{I}_j^H \tilde{I}_j)/\|\tilde{I}_j\|_2);$$

}

end

$a_{\text{opt}} = \arg\max_{k \in K} \|I_{k,A_n}\|_2;$

$U \leftarrow U \cup \{a_{\text{opt}}\};$

$\tilde{I}_a = \tilde{I}_{a_{\text{opt}}};$

$a = a + 1;$

} **end while loop**

output: new clients set: U

3.2. Client Categorization Method. Here, we suppose that a video frame can be run only if all the packets related to it have been successfully received. Each client sends a reported buffered frames number $B_{\text{RE}}(k,t)$ for client kth at t time. The buffer occupancy information is assumed to be reported along with the channel quality indicator (CQI) messages through feedback to the CCU [18]. $B_{\text{curr}}(k,t)$ is the current buffer occupancy, which is initialized as $B_{\text{curr}}(k,t) = B_{\text{RE}}(k,t)$ and would be updated according to the resource distribution results of kth client. To accurately distinguish the client urgency, CCU will classify the clients into three groups according to their received $B_{\text{curr}}(k,t)$ and SNR_k, these groups are as follows:

$$\begin{aligned} g_a &= \left\{k \mid \text{SNR}_k < d_{\text{th}} \Longrightarrow \left(\frac{R_k(k,t)}{V_R(k,t)}\right) < 1,\ B_{\text{curr}}(k,t) < B_L\right\} \\ g_b &= \left\{k \mid \text{SNR}_k \geq d_{\text{th}} \Longrightarrow \left(\frac{R_k(k,t)}{V_R(k,t)}\right) \geq 1,\ B_L < B_{\text{curr}}(k,t) < B_H\right\} \\ g_c &= \left\{k \mid \text{SNR}_k > d_{\text{th}} \Longrightarrow \left(\frac{R_k(k,t)}{V_R(k,t)}\right) > 1,\ B_{\text{curr}}(k,t) < B_{\max}\right\}. \end{aligned} \tag{17}$$

The clients in different set have different playout status according to their place inside the cell as shown in Figure 3. These groups are described as follows:

(i) $\mathbf{g_a}$: the clients in this group reside near the edge cell and cannot keep the continuity of the playout process when there are no enough resources allocated to the clients in time. Therefore, client's buffer will suffer from time interruption due to poor received signal; that is, $R_k(t) < V_R(t)$, which means if a video rate is greater than the network throughput, then new data is stored into the buffer at rate $R_k(t)/V_R(t) < 1$ and so the buffer decreases. In other words, if many segments are played before the next one arrives, then the buffer is consumed.

(ii) $\mathbf{g_b}$: the clients in this group reside far away from edge cell, so they have lower interruption probability than the ones in $\mathbf{g_a}$.

(iii) $\mathbf{g_c}$: the clients in this group reside near the base station "CCU"; they have enough buffered video frames to

overcome the wireless bandwidth fluctuation and no rebuffering probability due to the fact that if the network throughput $R_k(t)$ is always higher than the lowest video rate $V_{R\min}(t)$, that is, $R(t) > V_{R\min}(t)$, $t > 0$, there will be no need to a rebuffering event.

3.3. Distributing Mechanism. To help clients in reducing the interruption time, a resource distribution scheme is adopted to support clients with high download rates by utilizing the potential of multiple antennas that are offered by modern cellular networks. The distributing mechanism for resource scheme works as in the following steps:

1. The K receiver terminals use a fixed number of antennas, say K, to communicate with the base station and the control unit only selects N antennas among M large antennas for data transmission at the start of communications. The value of N can change with time and it is written as a function of time, $N(t)$.
2. After selecting the best clients, that is, "U", a video frame is sent; it will be split into several packets to adapt the physical transmission rate. A video frame can be run only if all the packets belonging to it have been successfully received.
3. The receiver client observes its current buffer value ($B_{\text{curr}}(t)$) during a time interval, where the time index $t = 1, 2, \ldots, T$ and $T \le Tc$ denotes the length of the monitoring time. The instantaneous value ($B_{\text{curr}}(t)$) varies due to the randomness nature of the wireless channel.
4. Using the playout outage probability in (6), the control unit can measure client's buffer performance of playout probability by sensing its value of ($B_{\text{curr}}(k,t)$).
5. To ensure delivering video streaming among clients with playout continuity through the cellular network, we assign a lower and an upper bound for the suitable playout outage probability to be P_{p-L}^{out} and P_{p-U}^{out}, respectively. With instantaneous playout outage probability $P_p^{\text{out}}(B_{\text{curr}}(k,t))$ for clients at t. The CCU monitors the level of playout outage probability and decides how to distribute radio resources during video transmission interval by the following steps:
   (a) When playout outage probability level (P_p^{out}) is between the lower and upper bounds for all t time, the cellular network performance in terms of playout outage probability behaves with good quality for delivering video as well as link quality being outperformed. Hence, if $N(t)$ is the number of radio resources (i.e., propagation channels) at t, then $N(t) + \Delta = N(t) + 0$ where Δ is a nature number which means no need for additional antennas.
   (b) If playout outage probability is more than the upper level (P_{p-U}^{out}), that means the network will deliver video streaming among clients with poor quality since clients suffer from interruption times during playout video streams. The CCU needs to provide additional radio resources for data transmission. The purpose of these additional resources is to supply higher spatial diversity gains to increase the link reliability. CCU will monitor the link quality and add a number of propagation channels to enhance system functionality in terms of download rate. Thus, number of antennas is $N(t) = N(t) + \Delta$, Δ is a number of added radio resources.
   (c) When playout outage probability is less than lower level (P_{p-L}^{out}) which means that the link quality is improved, CCU may need to reduce the number of antennas, so $N(t) = N(t) - \Delta$, Δ is a number of radio resources that go to be off.
6. The CCU periodically performs the steps from (3) to (5).

In the cases described in step (5), which can be defined as $\{C_0, C_1, C_2\}$, the observed buffer B_{curr} is set to be random vector for K clients and it can be given as

$$B_{\text{curr}} = \left\{B_{\text{curr}}(1,t), B_{\text{curr}}(2,t), \ldots, B_{\text{curr}}(k,t)\right\}. \tag{18}$$

Hence, the decision rules of the distributing mechanism are given by

$$\begin{aligned}
C_0 &: \left\{B_{\text{curr}} \mid 1 \ge P_p^{\text{out}} B_{\text{curr}}(k,t) > P_{p-U}^{\text{out}},\ t = 1, 2, \ldots, T,\ k \in g_a\right\}, \quad \Delta = +1;\\
C_1 &: \left\{B_{\text{curr}} \mid P_{p-L}^{\text{out}} \le P_p^{\text{out}} B_{\text{curr}}(k,t) \le P_{p-U}^{\text{out}},\ t = 1, 2, \ldots, T,\ k \in g_b\right\}, \quad \Delta = +0;\\
C_2 &: \left\{B_{\text{curr}} \mid P_{p-L}^{\text{out}} > P_p^{\text{out}} B_{\text{curr}}(k,t) \ge 0,\ t = 1, 2, \ldots, T,\ k \in g_c\right\}, \quad \Delta = -1;\\
C_3 &: \left\{B_{\text{curr}} \mid \text{the others for } P_p^{\text{out}} B_{\text{curr}}(k,t),\ t = 1, 2, \ldots, T\right\}, \quad \Delta = +0,
\end{aligned} \tag{19}$$

where Δ denotes the requested incremental number of radio resources from the central control unit, that is, $N(t) + \Delta$. Based on the above distributing mechanism, two important goals are achieved: firstly, the client can always perform higher spatial multiplexing for data transmission; secondly, the required playout outage probability for client buffer can be guaranteed simultaneously during data transmission. The distributing priority of clients in these three groups should be related to their urgencies; for example, the g_b clients will not be scheduled if $g_a \ne \phi$ and g_c will not be scheduled if $g_a \ne \phi$ or $g_b \ne \phi$. Hence, the distributing metric of kth client in each group can be identified according to (19) as follows:

$$\text{Ds}(k, A_n, t) = \begin{bmatrix} R_k^{\min}(k, A_n, t) & k \in C_0 \\ R_k^{a}(k, A_n, t) & k \in C_1 \\ R_k^{a}(k, A_n, t) & k \in C_2 \end{bmatrix}, \tag{20}$$

where $R_k^a(k, A_n, t)$ is the achieving rate that kth client can have when it allocates more antennas from the set A_n as in [6] with the following expression:

$$R_k^a(k, A_n, t) = \frac{c_m}{\sum_{m=1, m\in A_n}^{N} c_m} \frac{\chi_k}{\sum_{K}^{k=1} \chi_k} R_k(k, A_n, t), \quad (21)$$

where c_m is an extra weight for radio resource which is an integer and its value follows the value of SNR_k; that is, if SNR_k has small value in the set A_n of antennas for kth client, then c_m can represent the number of feasible video quality options. χ_k is the quality level that a client can reach, and $R_k(k, A_n, t)$ is described in expression (5). Then, $R_k^{\min}(k, A_n, t)$ can be obtained as follows:

$$R_k^{\min}(k, A_n, t) = \frac{c_m}{\sum_{m=1, m\in A_n}^{N} c_m} \frac{\chi_k}{\sum_{K}^{k=1} \chi_k} R_k^{\min}(k, t), \quad (22)$$

where $R_k^{\min}(k, t)$ is the minimum data rate required by clients at g_a.

$F(k, t)$ is defined as the most urgent frame of frames that have not been fully transmitted of the kth client at the time t, and the total size of the residual packets of this frame which wait in the MAC queue is defined as $\mathrm{SZ}_T(k, t)$. Then, $R_k^{\min}(k, t)$ can be given as follows:

$$R_k^{\min}(k, t) = \frac{\mathrm{SZ}_T(k, t)}{T(k, t)} = \frac{\mathrm{SZ}_T(k, t)}{B_{\mathrm{curr}}(k, t)/V_R(k, t)}. \quad (23)$$

Then the resource will be distributed for kth client and associated with its specific group if

$$(k, n) = \arg\max \mathrm{Ds}(k, A_n, t) \quad K \in g_x,\ n \in A_n^k, \quad (24)$$

where g_x is the group that kth client belongs to and A_n^k is the available resources for kth client.

4. Simulation Results

In this section, the performance of the proposed video buffer framework (VBF) is evaluated and its performance is compared with other known schemes such as PBDAS [9], MAX-CIR [19], and PF [20]. The idea of proportional fair (PF) scheme is based on the past average throughput which can act as a weighting factor of the expected data rate for the user in order that the users with bad channel conditions would be served within a certain amount of time. While the maximum carrier-to-interface ratio (MAX-CIR) scheduler finds the maximum value for each resource block, it searches for users whose values of CQI feedback equal the maximum found per each resource block (a random user would be selected if there is more than one user per resource block that would have the maximum value). In terms of fairness, the principle of this scheme is not fair in all situations and could be very biased.

The simulation is considered a single cell massive cellular network with five video streaming clients who are allocated in different places in this cell, so they have different bitrates. Table 1 shows the main configurations for this massive cellular network.

Table 1: Massive cellular network configurations.

Parameter	Configuration setting
Number of antennas	300
Bandwidth	20 MHz
Number of clients	5
Predefined SNR threshold	40 dB
Noise power spectral density	10^{-6} Watt/Hz
Frame size	1500 bytes
P_{p-U}^{out}	10^{-4}
P_{p-L}^{out}	10^{-5}
B_H	10
B_L	5

Figure 4 explains the cumulative interruption duration of each client for the proposed framework (VBF) and reference algorithms (PBDAS, MAX-CIR, and PF). From this figure, we notice that the interruption duration rises smoothly for PBDAS, which means the resource is distributed with fairness manner among clients. MAX-CIR has great partiality for the clients that have the best channel status while other clients with bad channel status are left waiting. For PF, the interruption time of the low bitrate clients is nearly zero and does not rise as time passed, which means that these clients have high resource distribution priority along all transmission process. On the other side, our proposed framework can insure the fairness among clients limiting the interruption duration length at a low level. The reason behind this superiority is that when the number of resources is high, proposed framework (VBF) will limit the interruption times of each client to achieve continuity during video playback according to our distribution mechanism which is described in Section 3.3.

Figure 5 shows that the system throughput for the proposed framework is compared with other exiting schemes. This figure shows that the proposed framework outperforms PBDAS, MAX-CIR, and PF, since the work of the proposed framework is based on increasing radio resource to enhance video quality at client side in order to maintain the playout outage probability value at accepted level, and also the decision on selecting the active clients is reached according to our proposed client selection algorithm which is described in Section 3.1. Additionally, this figure proved that PBDAS performance is less than the proposed framework because when all the clients have enough frames at the video buffer, the channel status is chosen as the index in scheduling the resource allocation in order to effectively ensure the system throughput.

Figure 6 shows the comparison of the proposed framework with other schemes in terms of total interruption duration under clients' bitrate variations condition. From this figure, PBDAS can work better than PF and MAX-CIR for the clients that demand videos with similar bitrates, but its works become partiality for clients with low bitrate while the proposed framework can achieve the minimum average interruption duration and ensure the fairness among clients.

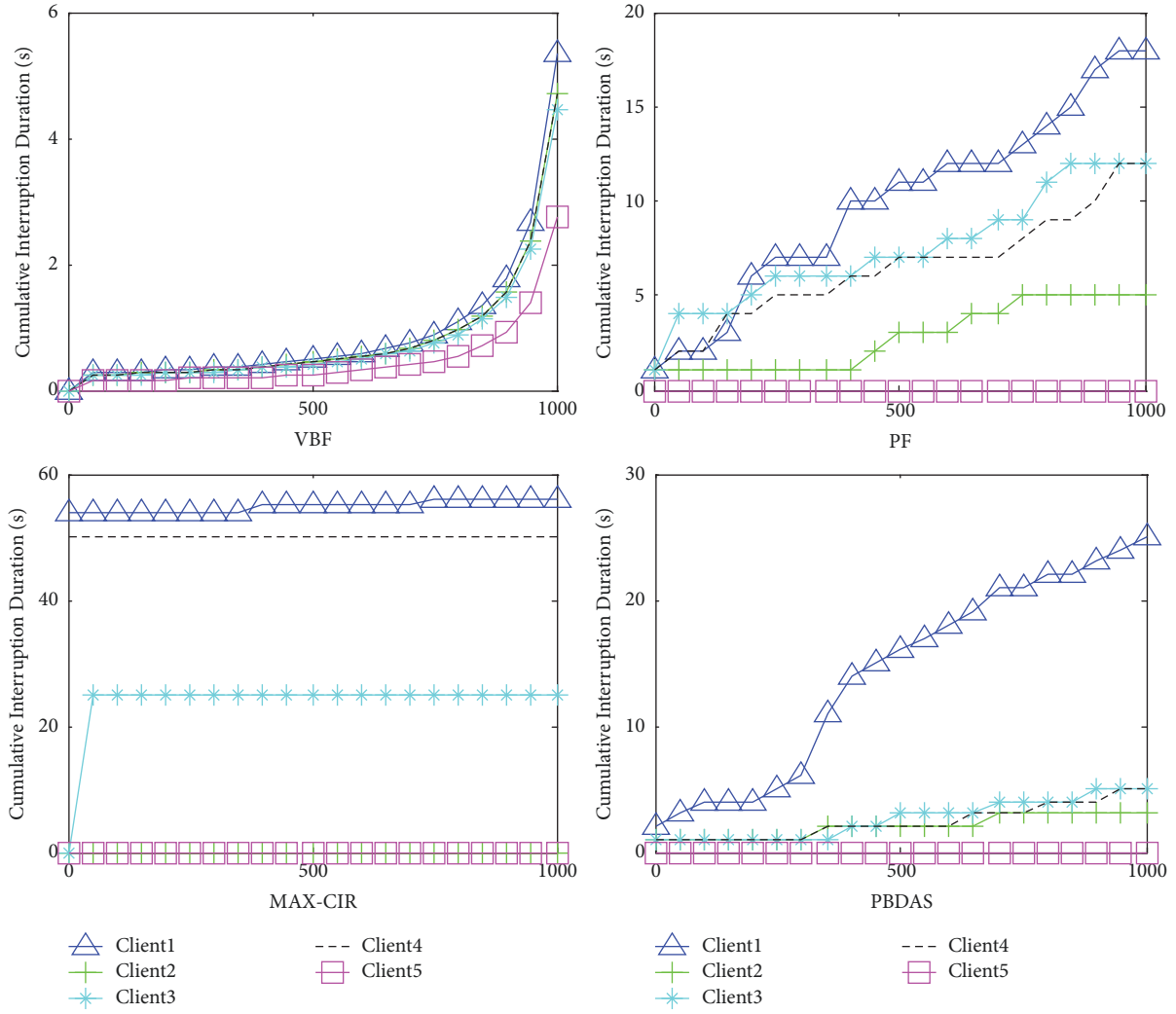

Figure 4: Total interruption duration for different clients in cellular network.

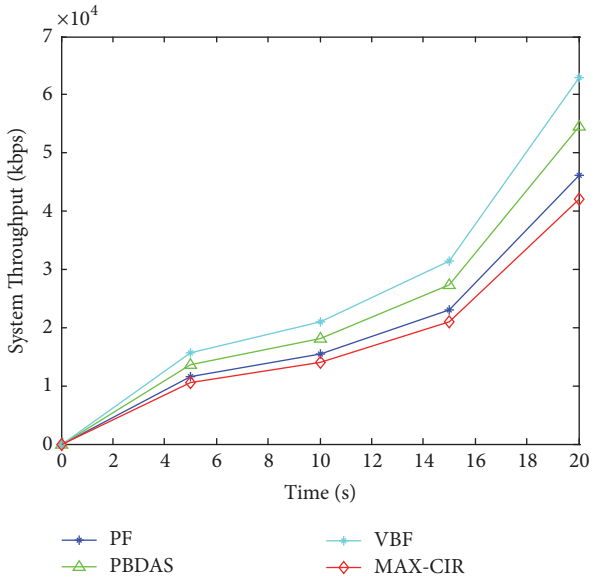

Figure 5: System throughput comparison for different schemes.

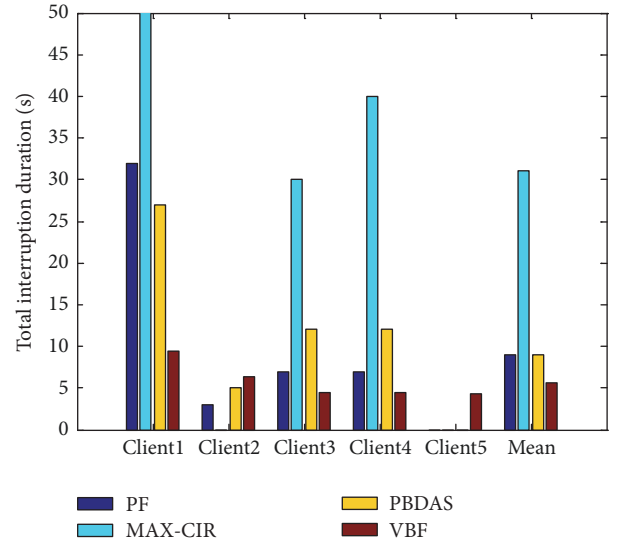

Figure 6: Total interruption duration of clients with different bitrates.

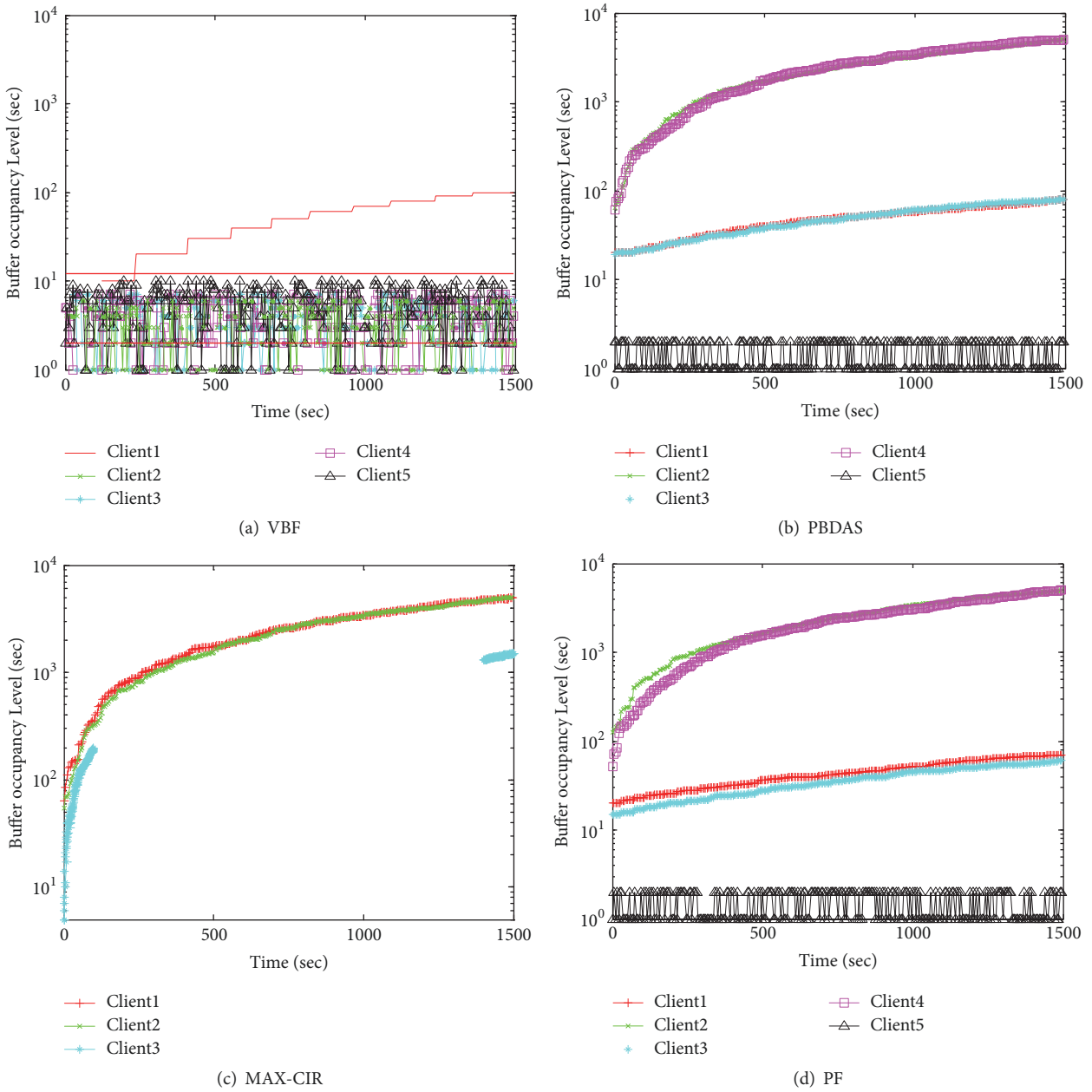

Figure 7: Buffer level occupancy with different algorithms.

Figure 7 shows the buffer occupancy level with time and the two red lines in this figure represent the B_H and B_L. At the starting of the simulation, all the clients have data more than B_H, then using our proposed client selection algorithm (described in Section 3.1) providing the client with a good channel quality is assigned the highest priority, which leads to buffer occupancy rising (the red curve). By using our proposed client categorization method which is described in Section 3.2, the clients in group g_a have the buffer occupancy lower than the B_L, so the playout outage probability in terms of buffer level is above P_{p-U}^{out}. Consequently, the CCU will use our proposed distributing mechanism (described in Section 3.3) to distribute the resources according to the clients' urgency in order to reduce the video playout outage and achieve continuity during video playback. Therefore, this figure proved that our proposed framework (VBF) is effectively controlled video buffer from going to below B_L.

In Figure 8, our proposed scheme is compared with conventional scheme in terms of efficiency, stability, and fairness based on the following metrics [21].

(a) Instability Metric. Clients are likely to be sensitive to frequent and important video bitrate switches as indicated by some studies. The instability metric is defined as

$$M_{\text{instability}} = \frac{\sum_{a=0}^{z-1} \left| V_{Ri,i-a} - V_{Ri,i-a-1} \right| \omega(a)}{\sum_{a=1}^{z} V_{Ri,i-a} \omega(a)}. \quad (25)$$

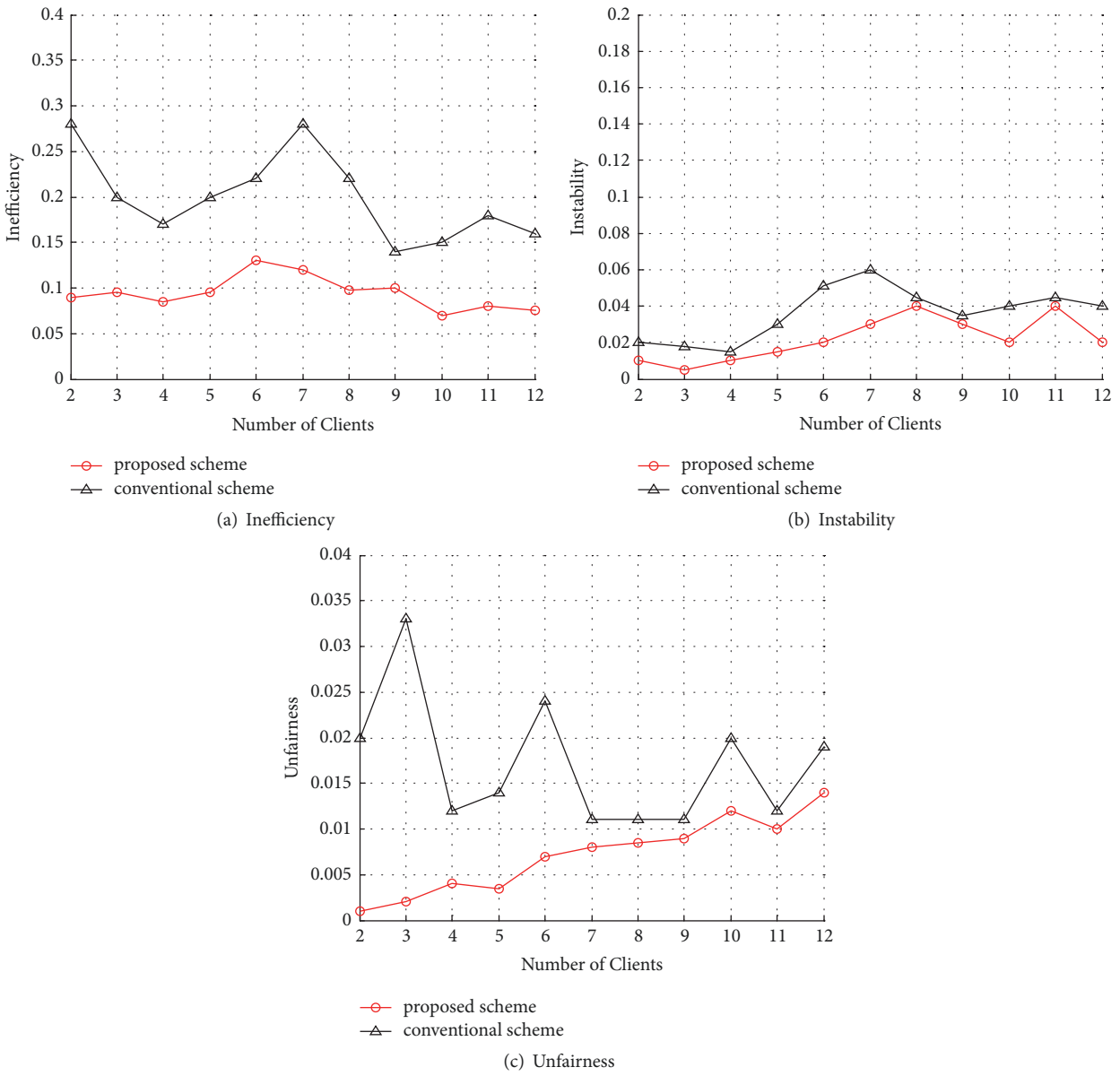

FIGURE 8: Performance evaluation when the available network bandwidth is fixed at 8 Mbps, and the number of competing clients varies from 2 to 12.

The instability metric equals the weighted sum of all bitrate switch steps monitored within the last 10 segments divided by the weighted sum of bitrates in the last $z = 10$ segments. V_{Ri} is the segment video bitrate at ith index and $\omega(a)$ is the weight function and it is equal to $i - a$.

(b) Unfairness Metric. At time t, the unfairness metric is defined as [21]

$$M_{\text{unfairness}} = \sqrt{1 - \text{jainFair}_t}, \tag{26}$$

where jainFair_t is the index of Jain fairness at time t and is calculated based on the bitrates V_{Ri} over all clients.

(c) Inefficiency Metric. This metric is calculated as

$$M_{\text{inefficiency}} = \left| \frac{\sum_j V_{Ri,j}}{T} \right|, \tag{27}$$

where $V_{Ri,j}$ is the video bitrate of ith segment for jth client and T is the available bandwidth. A value close to zero means that the clients in average are using as high average video bitrate as possible to improve video quality.

In Figure 8, the number of clients is changed from two to twelve with the available network bandwidth fixed at 8 Mbps in order to evaluate the performance of our scheme with another scheme. From this figure, the proposed scheme outperforms the conventional scheme based on bandwidth measurement of video buffer model under three operation

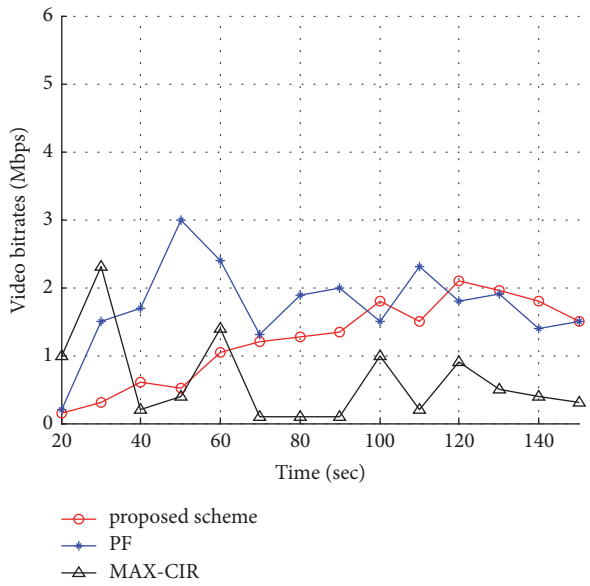

Figure 9: Video bitrates versus time under segment size variations.

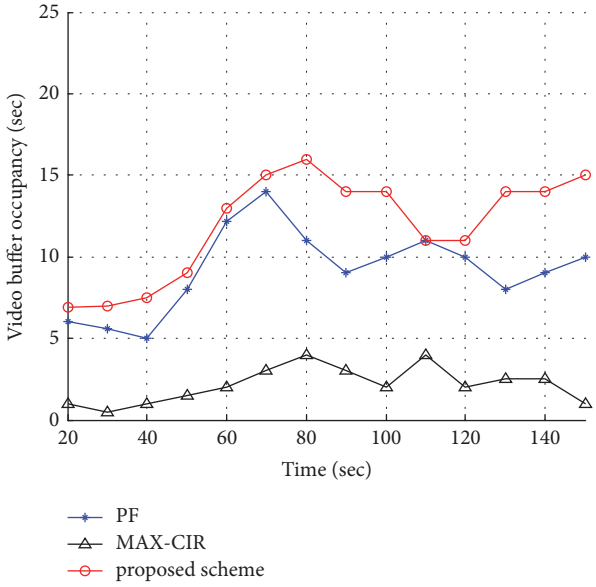

Figure 10: Video buffer occupancy versus time under segment size variations.

thresholds for preventing buffer underflow/overflow states. Also this figure shows the robustness of the adaptive algorithm for estimating the download time of next segment according to its size.

Figures 9 and 10 show that the conventional schemes change their bitrate very frequently whereas the proposed scheme adapts the video bitrate smoothly using the actual segment size and adapts its bitrate based on the future buffer occupancy. Maximum carrier-to-interface ratio (MAX-CIR) scheme is extremely oscillating in this scenario, since its estimation method does not consider VBR characteristics as we mentioned earlier. VBR characteristics also are affected by the buffer-based algorithm such as proportional fair (PF) scheme.

5. Conclusions

This work discusses the problem of video streaming playback over cellular network so that the client can have a continuity of video streams keeping the quality of video at acceptable viewing over the volatile nature of the wireless channel. Conventional schemes are trying to improve the video quality either through allocating resources to clients that have good channel quality, such as MAX-CIR scheme, or through allocating resources dynamically supporting the clients with low bitrates, such as PF scheme. Therefore, such schemes may fail to achieve fairness among clients. The proposed framework comes with a new distributing mechanism for controlling the playout outage probability level relying on the current buffer occupancy, which achieves accepted balance of fairness among different clients' urgency while minimizing the interruption duration and maintaining the buffer level between BL and BH. Moreover, the simulation results show that system throughput of the proposed approach outperforms other algorithms in enhancing the video quality within client experience because of increasing the radio resources in a fairness manner. Finally, the main limitations of this work are as follows: Our approach considers the client side buffered video time as feedback signal which means studying QoE for client instead of studying QoS for whole system. The other limitation is that our approach shows that the opportunities of buffer overflow/underflow are incurred by bandwidth estimation, but this approach does not consider the estimation error which can be affected by reserving a small positive/negative bandwidth margin. Finally, we suggest for future work to apply our proposed video buffer framework (VBF) with a multicell massive MIMO system to study the effects of intercell interference problem against video playback continuity.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective video quality assessment methods: a classification, review, and performance comparison," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 165–182, 2011.

[2] L. De Cicco, S. Mascolo, and V. Palmisano, "Feedback control for adaptive live video streaming," in *Proceedings of the 2nd Annual ACM Multimedia Systems Conference, MMSys'11*, pp. 145–156, USA, February 2011.

[3] S. R. Gulliver and G. Ghinea, "The perceptual and attentive impact of delay and jitter in multimedia delivery," *IEEE Transactions on Broadcasting*, vol. 53, no. 2, pp. 449–458, 2007.

[4] H. Chaari, K. Mnif, and L. Kamoun, "An overview of quality assessment methods of video transmission over wireless networks," in *Proceedings of the 2012 16th IEEE Mediterranean Electrotechnical Conference, MELECON 2012*, pp. 741–744, Tunisia, March 2012.

[5] I. Cisco, *Cisco visual networking index: Forecast and methodology, 20112016. White paper*, Cisco visual networking index, Forecast and methodology, 2012.

[6] B. Liu, H. Zhang, H. Ji, X. Li, and K. Wang, "Joint antenna allocation and rate adaption for video transmission in massive MIMO systems," in *Proceedings of the 2016 Digital Media Industry and Academic Forum, DMIAF 2016*, pp. 77–82, grc, July 2016.

[7] H. Fu, H. Yuan, M. Li, Z. Sun, and F. Li, "Models and analysis of video streaming end-to-end distortion over LTE network," in *Proceedings of the 11th IEEE Conference on Industrial Electronics and Applications, ICIEA 2016*, pp. 516–521, China, June 2016.

[8] C. Liu, I. Bouazizi, and M. Gabbouj, "Rate adaptation for adaptive HTTP streaming," in *Proceedings of the 2nd Annual ACM Multimedia Systems Conference (MMSys '11)*, pp. 169–174, San Jose, Calif, USA, February 2011.

[9] Y. Chen and G. Liu, "Playout buffer and DRX aware scheduling scheme for video streaming over LTE system," *IET Communications*, vol. 10, no. 15, pp. 1971–1978, 2016.

[10] L. De Cicco and S. Mascolo, "An adaptive video streaming control system: Modeling, validation, and performance evaluation," *IEEE/ACM Transactions on Networking*, vol. 22, no. 2, pp. 526–539, 2014.

[11] K. Miller, E. Quacchio, G. Gennari, and A. Wolisz, "Adaptation algorithm for adaptive streaming over HTTP," in *Proceedings of the 19th International Packet Video Workshop (PV '12)*, pp. 173–178, May 2012.

[12] L. Yu, T. Tillo, and J. Xiao, "QoE-Driven dynamic adaptive video streaming strategy with future information," *IEEE Transactions on Broadcasting*, vol. 63, no. 3, pp. 523–534, 2017.

[13] S. Kim, D. Yun, and K. Chung, "Video quality adaptation scheme for improving QoE in HTTP adaptive streaming," in *Proceedings of the 30th International Conference on Information Networking, ICOIN 2016*, pp. 201–205, Malaysia, January 2016.

[14] L. Arun Raj, D. Kumar, H. Iswarya, S. Aparna, and A. Srinivasan, "Adaptive video streaming over HTTP through 4G wireless networks based on buffer analysis," *Eurasip Journal on Image and Video Processing*, vol. 2017, no. 1, article no. 41, 2017.

[15] Z. Li, A. C. Begen, J. Gahm, Y. Shan, B. Osler, and D. Oran, "Streaming video over HTTP with consistent quality," in *Proceedings of the 5th ACM Multimedia Systems Conference, MMSys 2014*, pp. 248–258, Singapore, March 2014.

[16] M. Chen, M. Ponec, S. Sengupta, J. Li, and P. A. Chou, "Utility maximization in peer-to-peer systems with applications to video conferencing," *IEEE/ACM Transactions on Networking*, vol. 20, no. 6, pp. 1681–1694, 2012.

[17] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Uplink power efficiency of multiuser MIMO with very large antenna arrays," in *Proceedings of the 2011 49th Annual Allerton Conference on Communication, Control, and Computing, Allerton 2011*, pp. 1272–1279, USA, September 2011.

[18] T. Kim, K. Min, and S. Choi, "Multiuser CQI prediction based on quantization error feedback for massive MIMO systems," in *Proceedings of the 2015 International Conference on Computing, Networking and Communications, ICNC 2015*, pp. 32–36, USA, February 2015.

[19] G. Miao, J. Zander, K. W. Sung, and S. Ben Slimane, *Fundamentals of Mobile Data Networks*, Cambridge University Press, Cambridge, UK, 2016.

[20] F. Capozzi, G. Piro, L. A. Grieco, G. Boggia, and P. Camarda, "Downlink packet scheduling in LTE cellular networks: key design issues and a survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 2, pp. 678–700, 2013.

[21] J. Jiang, V. Sekar, and H. Zhang, "Improving fairness, efficiency, and stability in HTTP-based adaptive video streaming with FESTIVE," in *Proceedings of the 8th ACM International Conference on Emerging Networking EXperiments and Technologies (CoNEXT '12)*, pp. 97–108, ACM Press, New York, NY, USA, December 2012, http://dl.acm.org/citation.cfm?doid=2413176.2413189.