Thesis no: MSEE-2015:09

Developing a Fluid Flow Model for Mobile Video Transmission in the Presence of Play-Out Hysteresis

Atefeh Dehghannayyeri

Faculty of Computing
Blekinge Institute of Technology
SE-371 79 Karlskrona Sweden

This thesis is submitted to the Faculty of Computing at Blekinge Institute of Technology in partial fulfillment of the requirements for the degree of Master in Telecommunication Systems.

The thesis is equivalent to 20 weeks of full time studies.

Contact Information:
Author:
Atefeh Dehghannayyeri
E-mail: atde09@student.bth.se

University Advisor:
Hans-Jürgen Zepernick
Department of Communication Systems

Faculty of Computing
Blekinge Institute of Technology
SE-371 79 Karlskrona, Sweden

Internet : www.bth.se
Phone : +46 455 38 50 00
Fax : +46 455 38 50 57

ABSTRACT

This work focuses on improving video transmission quality over a mobile link. More specifically, the impact of buffering and link outages on the freeze probability of transmitted videos is studied. It introduces a new fluid flow model that provides an approximation of the freeze probability in the presence of play-out hysteresis. The proposed model is used to study the impact of two streaming buffer sizes over different possible combinations of outage parameters (data channel on/off times). The outcome of this thesis shows that outage parameters play a dominant role in freezing of streaming video content, and that an increase in these parameters cannot be easily compensated for by an increase in the size of the receiving buffer. Generally, in most cases when there is a variation in outage parameters, an increased buffer size has a negative impact on the freeze probability. To lower the probability of freeze during video playback over a weak mobile link, it is better to sacrifice resolution just to keep the video content playing. Similarly, shifting focus from off to on times brings better results than increasing buffer size.

Keywords: fluid flow models, video transmission, play-out hysteresis, freeze, buffering

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude and appreciation to my supervisor, Professor Hans-Jürgen Zepernick, for his continuous support, patience helps, motivation and immense knowledge during my thesis work. Beside my advisor, I would like to thank Professor Markus Fiedler, whom I adore unlimited for countless advices he gave me for improving the quality of this thesis and for introducing me to the topic as well as for his help in mathematical modeling. Finally yet importantly, I would like to thank my parents and my husband for their unconditional and unlimited support and love has never fainted.

CONTENTS

ABBREVIATIONS

CAGR	Compound Annual Growth Rate
CBR	Constant Bitrate
DSL	Digital Subscriber Loop
FV	Flash Video
HD	High Definition
HTTP	Hyper Text Transport Protocol
ISP	Internet Service Provider
JPEG	Joint Photographic Experts Group
MPEG	Moving Picture Experts Group
PSS	Packet-switched Streaming Service
QoD	Quality of Delivery
QoE	Quality of Experience
QoP	Quality of Presentation
QoS	Quality of Service
RTP	Real Time Protocol
VBR	Variable Bitrate
VoD	Video-on-Demand
3GPP	Third Generation Partnership Project

1 INTRODUCTION

Today, video streaming is one of the most popular services on the Internet. According to [1], in 2018, about 89% of mobile data traffic will be contributed by streaming/downloading audio and video services. Over the next five years, video streaming will be considered as an essential part of consumer's lives.

As the data networks improve in the world, it is expected that consumers will start using even higher levels of data. Mobile video will grow at a Compound Annual Growth Rate (CAGR) of 62% between 2015 and 2020, higher than the overall average mobile traffic CAGR of 53%. Of the 30.6 Exabyte's per month crossing the mobile network by 2020, 23.0 Exabyte's will be due to video [2].

If the transmitted frames of a video stream do not arrive in time, temporal artifacts may become visible during the video playback, which has a potentially adverse effect of user's experience. Video streams need a steady data rate to have smooth play-out, which put tighter constraints on the performance and throughput of the underlying network requiring more intelligent transmission models and buffer management algorithms to be put in place.

1.1 Motivation

With the multitude of video transmitted across the Internet infrastructure, video Quality of Experience (QoE) is of large interest for users, Internet service or content providers, and component manufacturers alike. Consequently, video-related services have received a lot of attention by researchers that lead to many development activities over the past years.

Current video streaming services are given guaranteed Quality of Service (QoS) from the Internet Service Provider (ISP) to the subscriber network, but QoS is not guaranteed from the subscriber network to each user because the network transmission policy is based on best effort [3]. To support stable video streaming to the end user, there is an eminent need to find a better model for analyzing video transmission behavior in the underlying network. This can be used to study the effect of different network configuration parameters and to define an optimal configuration suitable for video streaming.

Years have gone of efforts for developing efficient models for video transmission that offer acceptable quality to satisfy the user expectations. Among all of them, a few have received the attention of the research community. In this paper, we propose a new model for video transmission based on play-out hysteresis. The hysteresis is used to assess the buffer performance in networks with different scenarios. Also to solve the problem in having delay variations, hysteresis is used.

1.2 Research Questions

The following research questions express the need for finding a fluid flow model based on play-out hysteresis, which would be helpful in increasing the level of QoS in video transmission networks:

1. Can video transmission over a mobile link accurately be modelled, taking into account the play-out hysteresis of the jitter buffer output?

2. How can the dynamic behavior of the freezes during video transmission be described, in particular with respect to their frequency and durations?

3. How can freeze times during video transmissions be reduced?

The above research questions will be answered by designing a new model of a video transmission with focus on play-out hysteresis.

1.3 Related Work

In order to shed light on the capabilities of buffering, in this work, we study the impact of buffering through freeze probability of the video. We present an analytical model for video streaming systems in the presence of play-out hysteresis. In particular, we will see that the effect of buffering is affected heavily by the duration of outages (during which no content is flowing into the buffer), and that an increase in average outage times cannot be compensated for by an increase of buffer size. In this chapter, we briefly describe other research works that propose similar fluid flow model approaches for video transmission.

In [4], two general types of causes affecting video quality are distinguished. These can be categorized to be Quality of Delivery (QoD) and Quality of Presentation (QoP). Video codec, bit rate and loss recovery techniques are examples of application layer factors that may affect QoP whereas delay, jitter and packet loss are examples of network factors that could potentially affect QoD. To achieve a better understanding on the effect of different parameters on video quality, focus must be laid on the consequences of QoD.

The work presented in [5] shows that users prefer to wait for the video at the beginning of the video instead of frequent freezes during the playback. Earlier works on packet-switched networks [6] has already shown that buffering has a limited effect on the achievable load given the QoS constraints. Despite this, the buffers kept on growing during the last years that resulted in users complaining about the "bufferbloat" phenomena [7] that caused a set of non-optimal behaviors and experiences such a huge delays [8]. On the other hand, there are research works that show increasing the buffer size decreases the delay in the system. Thus, an infinite buffer size will result in an ideal condition for video streaming. These works focused on finding fluid flow models for video transmission based on having a buffer with infinite size. In [9], it is argued that increasing the size of the receiver buffer does not necessarily decrease the delay or risk of freezes in video transmission over a mobile link. In this thesis, we try to design a system based on [9]. Therefore, our research is focusing on finding a fluid flow model for mobile video transmission based on having a receiver buffer with finite size.

There exists a huge number of research works that concentrate on developing a fluid flow model where they analyze the effect of specific parameters. For instance, a study of a fluid model analysis of streaming media in the presence of time-varying bandwidth was given in [10]. They designed a model in which a constant bit-rate (CBR) media application is streamed over an unreliable network. The model consists of a tandem of two fluid queues. The first queue is a Markov-modulated fluid queue that models the network congestion, and the second queue represents the play-out buffer [11]. Unlike their work, we are interested in finding out a fluid flow model in the presence of play-out hysteresis.

In [12], a fluid flow model has been applied for describing the output rate process of a buffer. Even though the source model used is quite simple, a number of homogeneous outage sources with autocorrelation were assumed forming a one-dimensional Markov process. This resulted in the formulation of the output process to become very complicated resulting in a three-dimensional Markov process.

According to [13], in the last decade the literature on queueing theory has paid considerable attention to Markov-modulated fluid flow models. In these models, a fluid buffer is either filled, depleted or both at rates which are determined by the current state of a background Markov process, also called Markovian random environment. Partly, the work engaged in analysing such type of model. Scheinhardt encountered a new type of models and he called them “feedback fluid models” [13]. This thesis is close to the work presented in [8], but the main factor which makes our work different and unique in comparison to his work is that our work concentrates on finding a fluid flow model for the system in the presence of play-out hysteresis.

The focus of this thesis is toward designing a new fluid flow model for video transmission over a mobile link. The receiver of the system consists of a play-out buffer and a “freeze-indicator” reservoir. The play-out buffer is modeled as jitter buffer and exhibits hysteresis. The playback of the video is possible as long as the “freeze-indicator” reservoir is not empty.

1.4 Thesis Outline

The rest of the thesis is organized as follows. Chapter 2 provides background on video streaming and introduces the fluid flow modeling approach. Further, an overview of different factors that influence streamed video traffic is given. Finally, the concepts of jitter buffer and play-out hysteresis are introduced. Chapter 3 presents the proposed fluid flow model and describes calculation regarding the freeze probability. Chapter 4 explains the simulation of the proposed video transmission model. MATLAB is used as a simulation framework to study the behavior of the system. Results and analysis from three different simulation scenarios are presented. Chapter 5 concludes the work and suggests topics for future work.

2 BACKGROUND

In this chapter, the basic concepts used in mobile video transmission are introduced. An overview of video streaming is provided. Next, an explanation related to the jitter buffer in the video context is presented. Finally, the notion of jitter buffer in fluid flow model and play-out hysteresis is introduced.

2.1 Video Streaming

In the early days of video streaming, watching online content was not easy because users had to wait long time until the video content buffers and then plays. With the rapid introduction of high-bandwidth networking technologies like digital subscriber loop (DSL) and cable modems, the ability to stream high quality video has become more feasible than ever.

Video-on-Demand (VoD) services like Netflix, YouTube are becoming very popular among Internet users. These services have been successful in disseminating video content to any customer irrespective of geographical location. Additionally, video streaming over the Internet is being increasingly used in applications such as security monitoring, smart home, distance learning, collaboration and virtual environments [14, 15]. It is expected that consumers will start using even higher levels of video data growing at a CAGR of 62% between 2015 and 2020, which is higher than the overall average mobile traffic CAGR of 53% [2].

As more people start using web locations for video content, this leads to sustainability issues when clients raise above the upload capabilities of the streaming server. A large number of streaming protocols have been developed to address this issue. Both open-source and proprietary solutions are available. The Real Time Protocol (RTP) is one of the most popular open video streaming standards. The Third Generation Partnership Project (3GPP) Packet-switched Streaming Service (PSS) [16] recommends RTP as main transport component for media streams. RTP, however, does not provide any QoS guarantees. Each RTP stream maintains a queue, the jitter buffer, in which it buffers incoming data. The jitter buffer concept is discussed in more detail in Section 2.3. Adobe's Hyper Text Transport Protocol (HTTP) Dynamic Streaming System [17] is an example of a proprietary streaming solution and is currently being used in a variety of High Definition (HD) streaming networks. YouTube [18] encodes the videos in a Flash Video (FV) container for playback in Flash enabled browsers and it uses RTP for its mobile variant.

Besides the growing demand and advancements in streaming technologies, a key design challenge that still needs further attention is how video streams can be adapted to fit the limited bandwidth resources of the underlying network. For live video applications, the adaption occurs typically along changing the quality of video, adapting the frame rate or using a mixture of these two [19, 20]. For stored media stream delivery, adaption occurs with respect to three parameters: quality, frame rate and buffering. These parameters can be addressed in isolation or combined together. Efficient new techniques need to be develop to deliver highest quality of smooth video playback over the network.

2.2 Factors Impacting Streamed Video Traffic

Video being streamed is like water rushing through a pipe. When the pipe is bigger, more water can flow through it. An encoder fills a pipe with compressed video. Therefore, when the bit rate is larger, the encoder can compress more video information. This additional information makes a video with higher quality.

The key factors that affect the quality of streaming video through bitrate are encoding and compression schemes.

2.2.1 VBR and CBR Encoding

A video encoder produces a video stream at a bit rate. This bit rate is mainly dependent on the codec and content of the video [21]. There is a complex tradeoff between the achievable minimum encoding bit rate R and the distortion D of decoded images, which is described by the rate distortion function R(D) [22]. The entropy rate in bits/s of a source determines the maximum compression (or minimum bit rate) achieved for lossless (D = 0) coding. For a video frame it will be time varying, depending roughly on the instantaneous activity or motion. Higher activity sources will have large R(D) for the same D (see Figure 2.1 [23]).

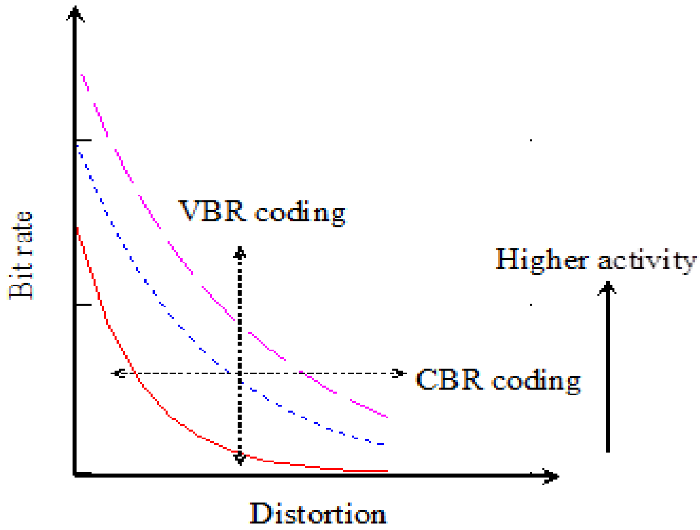


Figure 2.1: Rate-distortion function [23].

The rate distortion curves shown in Figure 2.1 define a region for video encoders operation. This region is bounded by two orthogonal lines:

1) Variable Bit Rate (VBR) coding (vertical line), which maintain constant quality throughout the video.
2) Constant Bit Rate (CBR) coding (horizontal line), in which a constant rate is maintained throughout the video.

From a network point of view, CBR coding is easier to handle. However, several real-life applications for high-quality video use VBR coding due to its flexibility.

2.2.2 Compression Schemes

Variable bit rate video streams depend strongly on the underlying compression scheme. In recent years, the respective standardization committees have been working on providing a set of generic compression standards that can be used for a variety of video applications. These include H.216 for video teleconferencing [23], Joint Photographic Experts Group (JPEG) format for still images [23], and Moving Picture Experts Group (MPEG) format for full-motion video [23]. Online compression is performed on the fly for real time video, and for services like VoD offline compression is performed.

2.3 Jitter Buffer

In streaming applications, jitter buffer, or play-out buffer, is a queue at the receiver’s end that temporarily buffers data frames before they are processed by the media decoder. The reason behind the deployment of a jitter buffer is to mask the effects of fluctuating network QoS on the delivery of media. The storage capacity of the jitter buffer varies depending upon the particular application. For example, in video conferencing, the queue length is minimized to reduce the communication latency. In VoD and other streaming applications in browsers such as YouTube, the jitter buffer length is equal to the size of the streaming media. For this, the necessary storage space should be available beforehand. In desktop computers and laptops, this is generally not a problem, but mobile devices are often constrained in memory.

Depending on the number of channels in streamed media, a media decoder can maintain several jitter buffers in parallel. A basic stream usually consists of one channel for audio and another one for video. Each jitter buffer can be modeled as a G/D/1 queue where data arrival is assumed to be a general process, and the output frames are consumed at deterministic time-intervals resembling frame-rate of the video or sample-rate of the audio. The state of the jitter buffer at a given time can be considered as an indicator of the network performance, which has an effect on the media decoder’s ability to display streamed media content in a timely manner. A jitter buffer can be distinguished with four different state types: full buffer, empty buffer, critical holes, and non-critical holes. An illustration of four different states of a jitter buffer is shown in Figure 2.2.

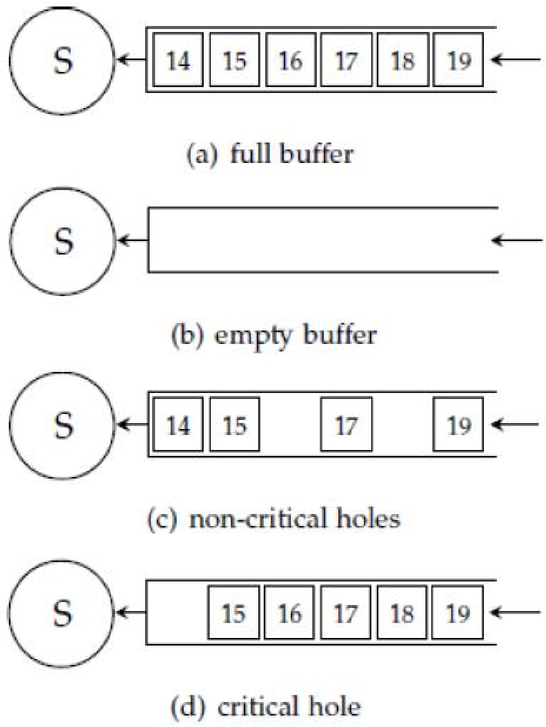


Figure 2.2: Four different states of jitter buffer with six places [24].

A full buffer (Figure 2.2(a)) can be defined as a state when all packets starting from the first schedule packet until the last received packet are present in the buffer. An empty buffer (Figure 2.2(b)), on the other hand, does not have any packet in the queue. Figure 2.2(c) shows non-critical holes. These missing packets are not first in the queue. Last, critical holes are shown in Figure 2.2(d). Critical holes are created when packets that are supposed to be first in the queue are missing.

The effect of these jitter buffer states on media playback is dependent on the decoder being used. An empty buffer will halt media playback temporarily. Similarly, a critical hole will affect the media quality, e.g. frame skipping, changes in color, or blurred pictures. An empty buffer is more critical for the timely playback than a critical hole. A non-critical hole can become a critical hole when it moves to the first place in the queue. A non-critical hole could also disappear into a full buffer when the missing packets arrive at a later point in time because of packet retransmission due to losses or corrupted communication.

When a jitter buffer remains empty for some time, the media playback is temporarily halted. During this time, the jitter buffer has the opportunity to populate again while no frames are retrieved, this is called re-buffering.

2.4 Jitter Buffer Fluid Flow Model

Fluid flow queues have demonstrated to be a powerful modeling paradigm in a wide range of applications and have therefore received much attention in literature. On one hand, they often capture the key characteristics that determine the performance of communication networks with complex packet-level dynamics while, on the other hand, they remain mathematically tractable [25].

A jitter buffer can be modeled as a fluid flow model where a media stream arriving at and departing from the jitter buffer is considered a bit stream. Let Y(t) be the amount of data in the jitter buffer at a given time t. In particular, Y(t) is the amount of downloaded data, minus the amount of data retrieved by the video decoder up to time t. The rate at which Y(t) changes over time is given by

$$\frac{dY(t)}{dt} = \mathrm{J(t)} - \mathrm{X(t)} \tag{2.1}$$

where $\mathrm{X(t)}$ is the data rate at which the video decoder fetches data from the jitter buffer and $\mathrm{J(t)}$ is the data arrival rate to the buffer. Specifically, $\mathrm{J(t)}$ represents a random process equal to the transmission rate and $\mathrm{X(t)}$ is a batch process proportional to the frame rate of the video. The data arrival rate $\mathrm{J(t)}$ can be affected by the network if it is unable to sustain the transmission. The following three scenarios can occur:

1) $\mathrm{J(t)} = \mathrm{X(t)}$: The amount of data in the jitter buffer remains stable.
2) $\mathrm{J(t)} > \mathrm{X(t)}$: The buffer size grows.
3) $\mathrm{J(t)} < \mathrm{X(t)}$: The buffer size shrinks.

Depending on the type of application, $J(t)$ is targeted to be larger than or equal to $X(t)$ by controlling the transmission rate. In both these cases $dY(t)/dt \geq 0$ to sustain a seamless media playback from the jitter buffer. When $dY(t)/dt < 0$ for a considerable time, this indicatives unreliable network conditions. This condition will eventually lead to buffer underflow or exhaustion if the network performance does not improve. Thus, $dY(t)/dt$ indicates the ability of the network to deliver a media stream over a network in a timely fashion.

The jitter buffer can starve because data is not arriving at the minimal rate required to sustain a certain population of the jitter buffer. The starvation of a jitter buffer could also be the consequence of packet losses within the network on the path to the jitter buffer. There is a small delay between the actual start of the streaming and the start of the media consumption from the jitter buffer. This is to allow the jitter buffer to populate before the actual playback of the media starts. It has been shown that for a longer start-up latency, the chance of jitter buffer starvation decreases [26, 27].

2.5 Play-Out Hysteresis

Hysteresis is a widely occurring phenomenon. It can be found in a wide variety of natural and constructed systems. Generally, it represents history dependence, of a system when a characteristic looping behavior of the input-output is formed.

Hysteresis occurs in video streaming in response to varying content arrival rate at the jitter buffer. Video play-out happens until the buffer becomes empty and then there will not be any play-out until the buffer becomes full again. It forms a loop with typically two branches. The first branch represents the filling of the buffer until the desired threshold is reached, during which the video stream is displayed. The second branch applies to decreasing buffer content, during which play-out takes place until the buffer becomes empty. This phenomenon is called “play-out hysteresis”.

3 PROPOSED VIDEO TRANSMISSION MODEL

One of the important key design issues in video transmission is how video streams can be adapted to fit within the available bandwidth of the network. Several video transmission models have been proposed in the literature, but none of these models take into account the effect of play-out hysteresis.

Therefore, an analytical model for video streaming systems in the presence of play-out hysteresis is proposed in this chapter. This model can be used to understand and quantify the effect of different parameters like mean off time, mean on time and buffer size on the process of freezing during video transmission.

3.1 Proposed Model

Normally, queueing models are used to assess the performance of data services at the packet level in mobile networks, for instance, model of multiplexers or buffers. In recent years, however, stochastic fluid flow models are being used as a simplification to describe competition, allocation and consumption processes of a resource. A fluid flow model of a queuing system gives an abstraction in which the queue with its discrete waiting times is replaced by a bucket or funnel offering a certain volume. The input flows are modulated by finite continuous-time Markov chains. Such a fluid flow model is particularly well suited for streaming video services that produce a certain amount of traffic to be transmitted within a time interval. By using this approach, the modeling of streams is reduced to modeling their intensities, and to describe issues with nominal reaching values due to disturbances such as varying capacities.

The proposed analytical model uses a similar approach to describe video transmission over a mobile data link where outages are non-deterministic. The receiver of the system consists of a play-out reservoir and a “freeze-indicator”. The play-out reservoir is modeled as a jitter buffer that exhibits hysteresis. The playback of the video is possible as long as the “freeze-indicator” reservoir is not empty.

The proposed model is based on the previous work reported in [8], where a fluid flow model consisting of two infinite size reservoirs is presented (see Figure 3.1). One reservoir with content D_t is located at the sender, and the second one is located at the receiver with content C_t. Both reservoirs are driven by a continuous-time Markov process M_t, which is characterized by its state space $\{0, 1\}$. When M_t=0, the content of the first reservoir decreases and when M_t=1, the content of the first reservoir increases.

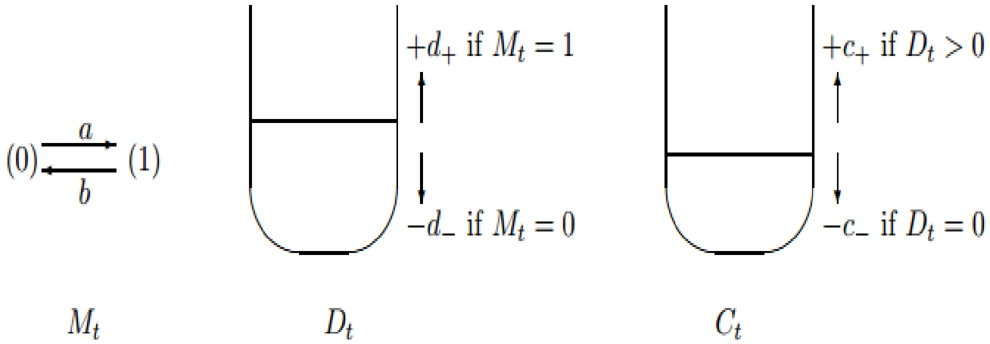


Figure 3.1: Interaction of buffer content processes [8].

As it illustrated in Figure 3.1, the first reservoir receives its input from an exponential on-off source and gets emptied at a constant rate. The second reservoir is driven by the first. When M_t is in state one, the content of the first reservoir increases with constant rate d_+ and when M_t is in state zero, its content decreases with rate d_-. When the first reservoir's content is not empty, the content of the second reservoir increases at rate c_+ and else decreases at rate c_-.

The proposed fluid flow model consists of one finite size play-out reservoir and a freeze indicator. The reservoir is filled with content C_t and the freeze indicator is filled with content F_t. The content of the play-out reservoir increases at rate c_+ and decreases at rate c_-. Figure 3.2 shows the proposed fluid flow model.

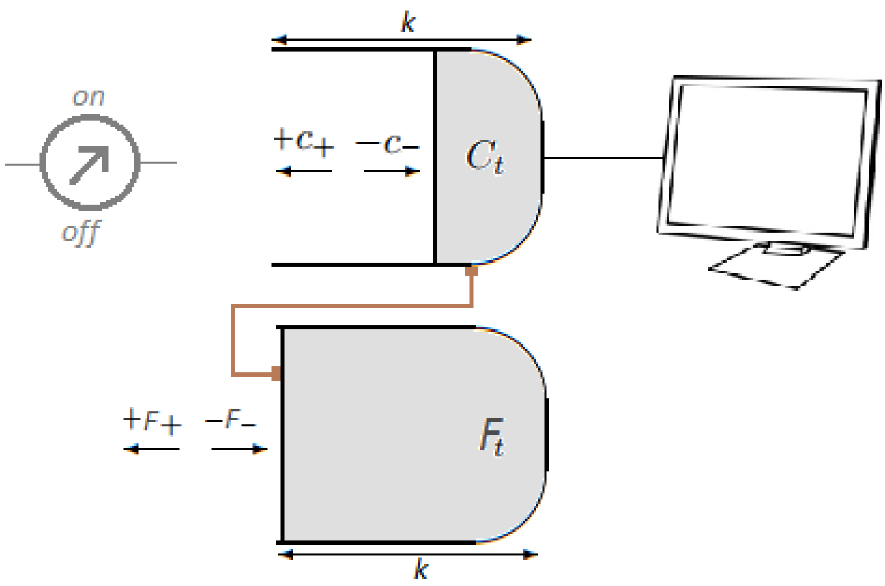


Figure 3.2: Interaction of buffer content processes in the proposed model.

The first reservoir (play-out buffer) receives its input at a constant rate C_{on} from an on-off source and gets emptied at a constant rate C_{off}. The content of the freeze indicator F_t depends on the content of play-out buffer C_t. When the play-out buffer becomes full, then the freeze indicator immediately becomes full and remains full as long as the play-out buffer does not become empty. At the moment the play-out buffer becomes empty, the freeze indicator immediately becomes empty and remains empty until the play-out buffer becomes full again.

The model toggles between the two following states:

- An "ON" state, during which the mobile channel is transparent to the application and the play-out buffer content increases at constant rate C_{on} $[Du/Tu]$. Here, Du and Tu denote \`Data Units´ and ´Time Units´, respectively. The ON state duration equals to T_{on} $[Tu]$.

- An "OFF" state, during which the mobile channel is blocked and the play-out buffer content decreases at constant rate C_{off} $[Du/Tu]$. The OFF state duration equals to T_{off} $[Tu]$.

The proposed model is based on the following assumptions:

- The size of the play-out buffer and freeze indicator are equal to k.
- C+, C-, F+, F- are positive rates.
- At startup, both play-out buffer and freeze indicator are full.
- There is no data coming into the play-out buffer at the beginning.
- $C_{on} \geq C_{off}$ to avoid indefinite freeze condition.

The freeze indicator can be empty in the following scenarios:

- When the system is in OFF state and the play-out buffer becomes empty, then the freeze indicator will immediately become empty.

- If the play-out buffer and freeze indicator became empty in the previous state, and the play-out buffer did not became full after time T has passed, then the freeze indicator remains empty.

Therefore, we can calculate the probability of empty freeze indicator $Pr[F = 0]$ as follows:

$$Pr[F = 0] = Pr[C = 0] + Pr[O < C < k]\, Pr[F = 0] \quad (3.1)$$

where $Pr[C = 0]$ is the probability that the play-out buffer is empty and $Pr[O < C < k]$ denotes the probability that the play-out buffer is neither full nor empty.

The probability $Pr[O < C < k]$ is calculated as

$$Pr[O < C < k] = 1 - Pr[C = 0] - Pr[C = k] \quad (3.2)$$

Substituting (3.2) into (3.1) gives

$$Pr[F = 0] = Pr[C = 0] + (1 - Pr[C = 0] - Pr[C = k]\,)Pr[F = 0] \quad (3.3)$$

Then, rewrite (3.3) as

$$Pr[F = 0] - (1 - Pr[C = 0] - Pr[C = k]\,)Pr[F = 0] = \mathrm{Pr}[C = 0] \quad (3.4)$$

and after simple manipulation, we have

$$Pr[F = 0](1 - 1 + Pr[C = 0] + Pr[C = k]\,) = \Pr[C = 0] \tag{3.5}$$

Therefore, the probability that the freeze indicator is empty is given by

$$Pr[F = 0] = \Pr[C = 0]/(\Pr[C = 0] + \Pr[C = k]) \tag{3.6}$$

Similarly, the freeze indicator is full for the following two scenarios:

- When the system is in ON state and the play-out buffer becomes full, then the freeze indicator will immediately become full.
- If the play-out buffer and freeze indicator became full in the previous state and reached the desire threshold k, and after time T has passed during which the play-out buffer did not became empty, then the freeze indicator remains full.

Therefore, we can express the probability that the freeze indicator is full, $Pr[F = k]$, as

$$Pr[F = k] = Pr[C > 0]Pr[F = k] + Pr[C = k]Pr[F = 0] \tag{3.7}$$

where $Pr[C > 0]$ is the probability that the play-out buffer is not empty and $Pr[C = k]$ is the probability that the play-out buffer is full.

Since, the freeze indicator can be only full or empty, $Pr[F = 0]$ can be expressed as

$$Pr[F = 0] = 1 - Pr[F = k] \tag{3.8}$$

Substituting (3.8) into (3.7) gives

$$Pr[F = k] = Pr[C > 0]Pr[F = k] + Pr[C = k](1 - Pr[F = k]) \tag{3.9}$$

and after simple manipulation, we have

$$Pr[F = k] - Pr[C > 0]Pr[F = k] + Pr[F = k]Pr[C = k] = Pr[C = k] \quad (3.10)$$

Then, we rewrite (3.10) as

$$Pr[F = k](1 - Pr[C > 0]) + Pr[F = k]Pr[C = k] = Pr[C = k] \quad (3.11)$$

Since $1 - Pr[C > 0]$=$Pr[C = 0]$, (3.11) can be written as

$$Pr[F = k](Pr[C = 0] + Pr[C = k]) = Pr[C = k] \quad (3.13)$$

Therefore, the probability of freeze indicator being full is given by

$$Pr[F = k] = \mathrm{Pr}[C = k]/(\mathrm{Pr}[C = 0] + \mathrm{Pr}[C = k]) \quad (3.14)$$

Then, the expressions (3.6) and (3.14) show the dependence of the freeze indicator on the contents of the play-out buffer. To verify these equations, the proposed model is simulated and the results are presented in Chapter 4.

The model exhibits hysteresis after time T has passed, during which the content of the play-out buffer reaches C such that 0<C<k. At that moment, it is not possible to know the state of the freeze indicator unless the previous states of the play-out buffer are known.

If in its previous state, the play-out buffer became full (reached threshold k), then the freeze indicator will remain in the full state and the video keeps playing. But if in its previous state, the play-out buffer became empty (reached threshold 0), then the freeze indicator remains in the empty state and video playback freezes. Figure 3.3 illustrates these two scenarios.

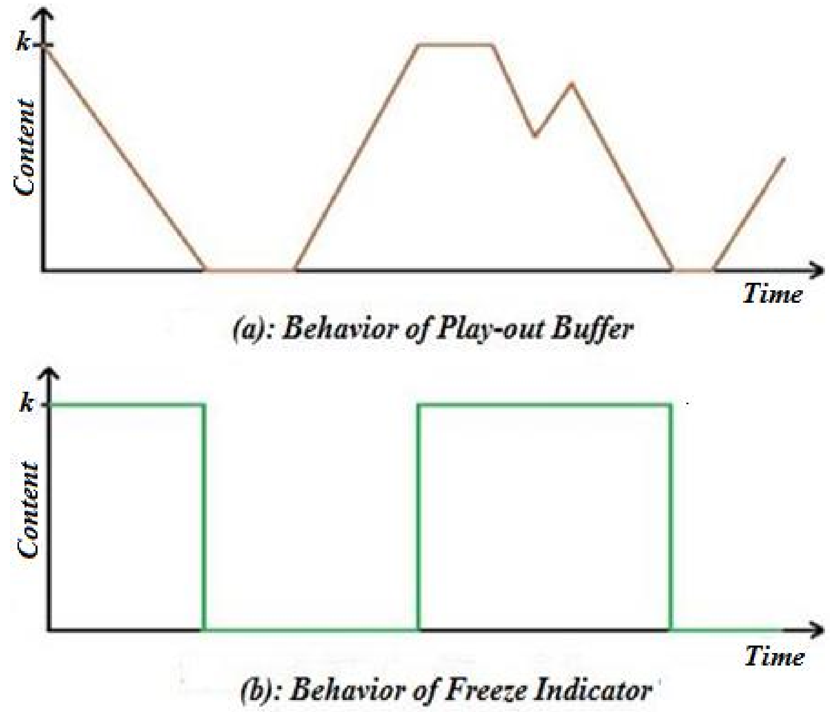


Fig 3.3: Dependence of freeze indicator on play-out buffer.

4 SIMULATIONS

In this section, the simulation of the proposed video transmission model is presented. MATLAB is used as a simulation framework to model the behavior of the play-out buffer and freeze indicator.

As explained in the previous chapter, data flows arrive into the play-out buffer at a constant input rate C_{on} from an on-off source and flows out at a constant output rate C_{off} when the freeze indicator is full.

Let each ON state time T_{on} and OFF state time T_{off} be a random number from $\mathrm{exprnd}\,(\mathrm{ET_{on}})$ and $\mathrm{exprnd}\,(\mathrm{ET_{off}})$, where these two functions define exponential distribution with mean parameter ET_{on} and ET_{off} , respectively. In another words, T_{on} and T_{off} can be written as

$$\mathrm{T_{on}} = \mathrm{exprnd}\,(\mathrm{ET_{on}}) \tag{4.1}$$

$$\mathrm{T_{off}} = \mathrm{exprnd}\,(\mathrm{ET_{off}}) \tag{4.2}$$

The time required to completely fill an empty play-out buffer and vice versa is defined as

$$T'_{on} = \frac{k}{C_{on}} \tag{4.3}$$

$$T'_{off} = \frac{k}{C_{off}} \tag{4.4}$$

Two conditions are possible for T_{on} and T_{off}: either $T_{on/off}$ values are $'large'$ enough to fill/empty the play-out buffer completely ($T_{on/off} \geq T'_{on/off}$) or they are too $'short'$ to empty/fill the play-out buffer in a given time ($T_{on/off} < T'_{on/off}$). The notations $'large'$ and $'short'$ are used to indicate these conditions. Figure 4.1 illustrates a large T_{off} condition.

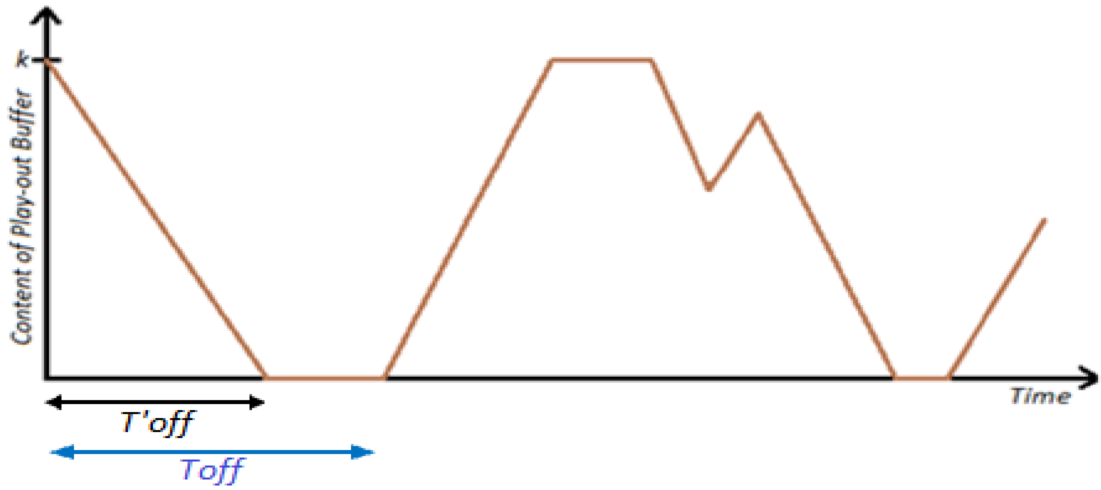


Figure 4.1: Illustration of T'_{off} and large T_{off} values for play-out buffer.

The developed model is simulated in MATLAB with parameters shown in Table 4.1 to explain the behavior of different states;

Table 4.1: Simulation parameters for results shown in Figure 4.2 and Figure 4.3.

Parameter	***Value***
Buffer Size k $[Du]$	3
Mean On Time ET_{on} $[Tu]$	2
Mean off Time ET_{off} $[Tu]$	2
Input Rate C_{on} $[Du/Tu]$	1
Input Rate C_{off} $[Du/Tu]$	1
Number of On & Off state changes	7

As Figure 4.2 and 4.3 show, the simulation starts with both the play-out buffer and freeze indicator completely full. Next, the system enters an OFF state with short T_{off} condition during which the content of play-out buffer decreases with rate C_{off}. Since the buffer did not completely become empty during this state, the freeze indicator remains full and the video keeps playing. Then, the system enters ON state during which the content of play-out buffer remains at the same level since $C_{on} - C_{off} = 0$. This is followed again by an OFF state during which the content of play-out buffer decrease with rate C_{off} until it finally becomes completely empty.

At this moment, the freeze-indicator suddenly becomes empty and the video playback is stopped. During the next ON state with short T_{on} condition, the play-out buffer's content increases with rate C_{on} only as no video is currently being played. Since the play-out buffer did not become completely full during this state, the freeze indicator remains empty. This is again followed by 2 OFF/ON states with short conditions, during which the play-out buffer is not able to reach threshold k, keeping the freeze-indicator empty, i.e. no video playback during these states.

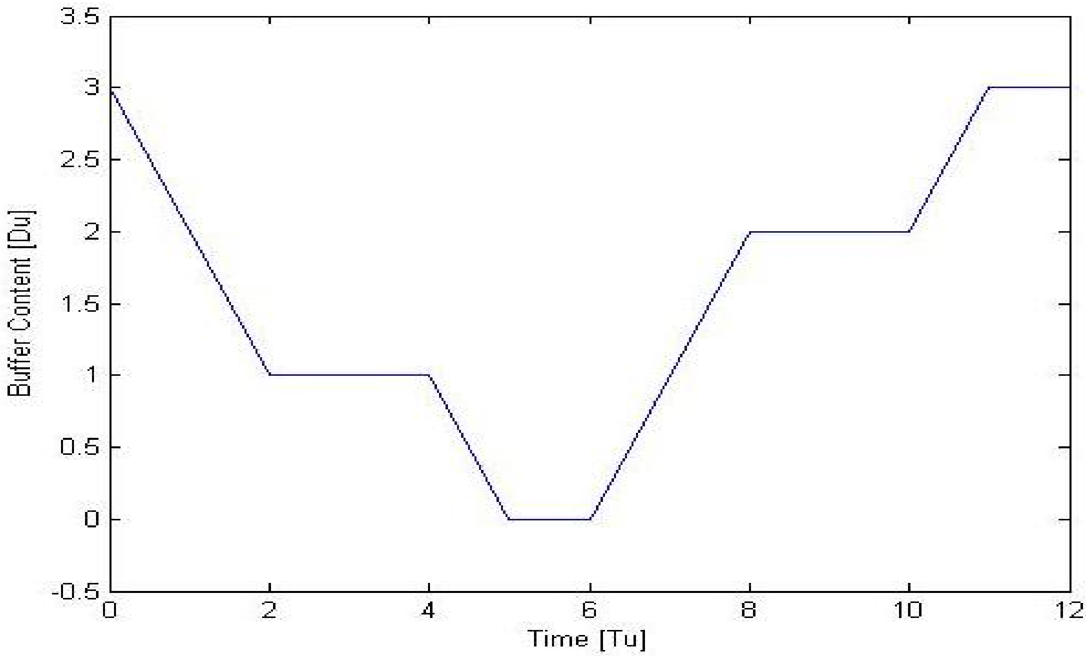


Figure 4.2: Content of play-out buffer C(t).

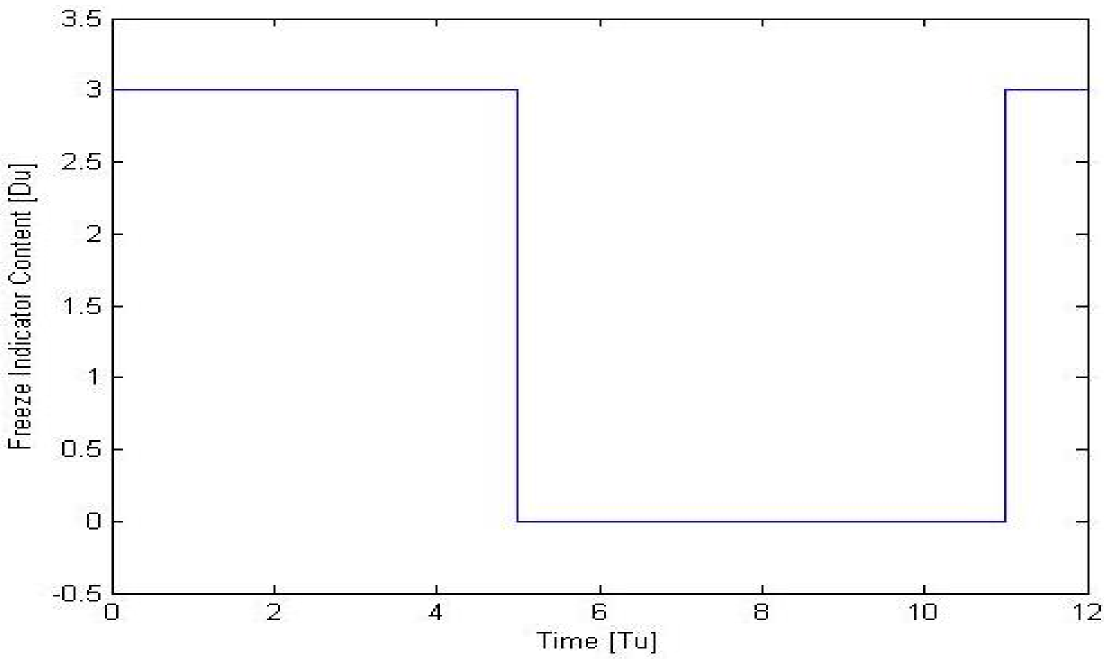


Figure 4.3: Content of freeze indicator F(t).

In order to explain the real simulation, four combinations of T_{on} and T_{off} are possible over a large number of ON/OFF state changes:

1) Large ET_{off}, Large ET_{on}
2) Large ET_{off}, Short ET_{on}
3) Short ET_{off}, Large ET_{on}
4) Short ET_{off}, Short ET_{on}

These possible combinations are simulated for the proposed model with two different buffer sizes and identical C_{on}/C_{off} rates. For each simulation, the probabilities $Pr[F = k]$ and $Pr[F = 0]$ are calculated and the simulation runs for 100,000 numbers of ON/OFF state changes. In each simulation scenario, the pre-condition is the play-out buffer and freeze indicator are full in the beginning.

Scenario 1: Large ET_{off}, Large ET_{on}

In this scenario, the system enters OFF stats with large T_{off} and enters ON states with large T_{on} condition, meaning the time that data flows into the buffer (T_{on}) or goes out of the buffer (T_{off}) is large enough to make the play-out and freeze indicator completely full or empty. Table 4.2 represents parameters in this scenario.

Table 4.2: Simulation parameters for results shown in Figure 4.4 and Figure 4.5.

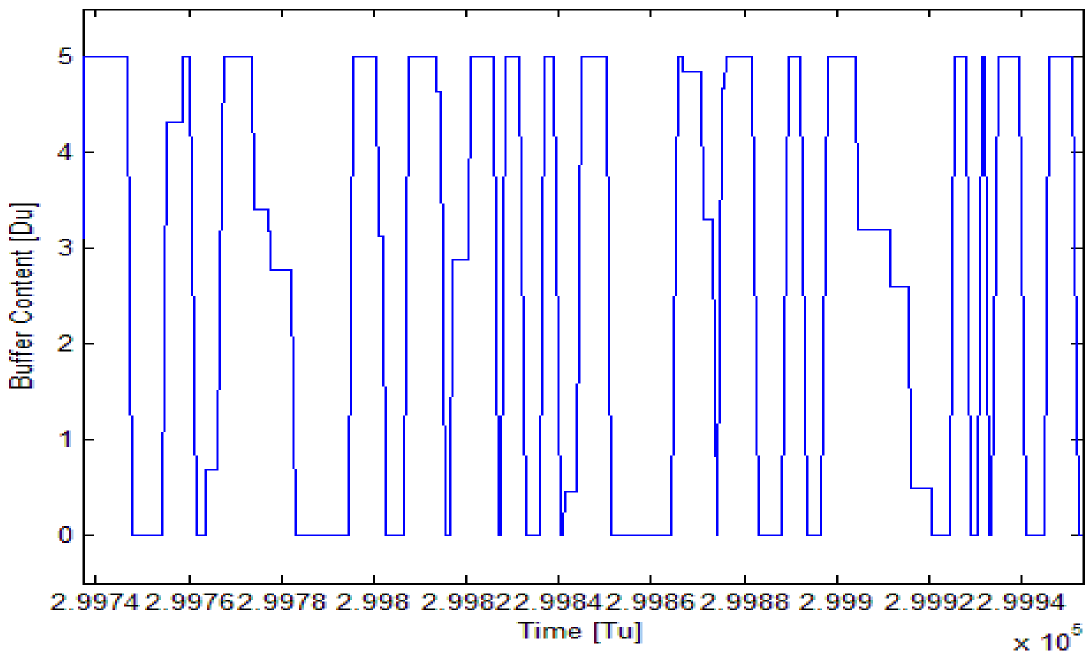
Parameter	***Value Group1***	***Value Group2***
Buffer Size k $[Du]$	5	10
Mean On Time ET_{on} $[Tu]$	3	3
Mean Off time ET_{off} $[Tu]$	3	3
Input Rate C_{on} $[Du/Tu]$	4	4
Input Rate C_{off} $[Du/Tu]$	4	4
Number of On & Off State Changes	100,000	100,000

The results in Table 4.3 show, when ET_{off} and ET_{on} are large enough, almost increasing the buffer size has no effect on freeze indicator probabilities. For instance, the probability of video playback in both cases when k=5 and k=10 is the same number. The reason is, both T_{off} and T_{on} are large enough to make both play-out buffers (k=5, k=10) completely full or empty during a single state.

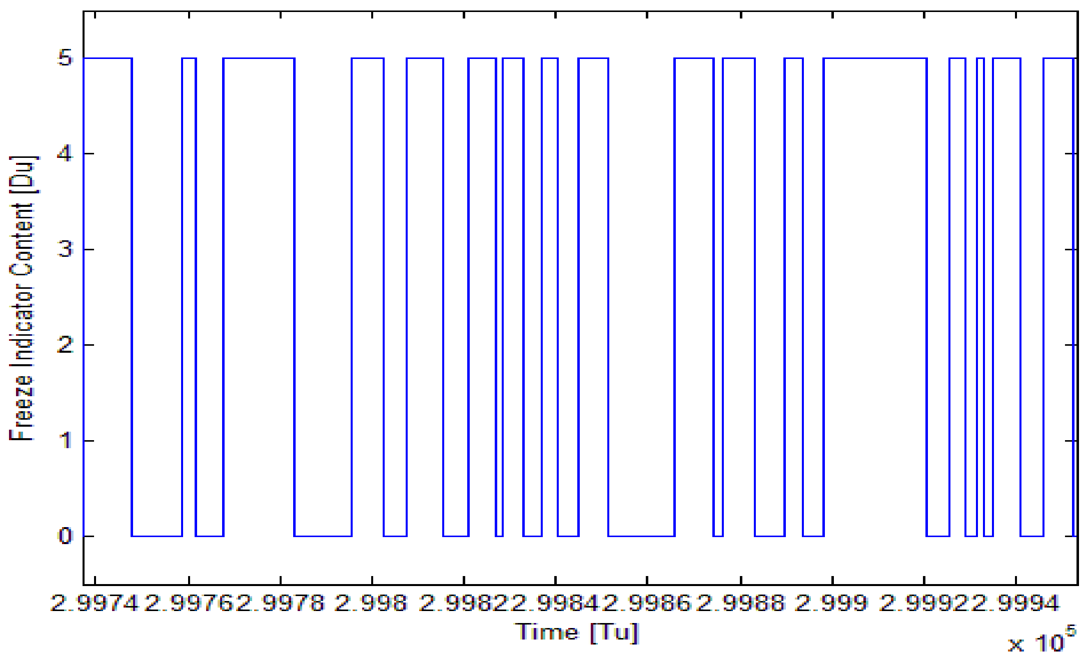
Table 4.3: Probability of full or empty freeze indicator for k=5 and k=10.

Parameters	***Results***	***Results***
Buffer Size	$Pr[F = k]$	$Pr[F = 0]$
k = 5	0.50	0.50
k = 10	0.50	0.50

Figure 4.4 and Figure 4.5 display the behaviors of play-out buffer and freeze indicator for the last 20 state changes when k=5 and k=10.

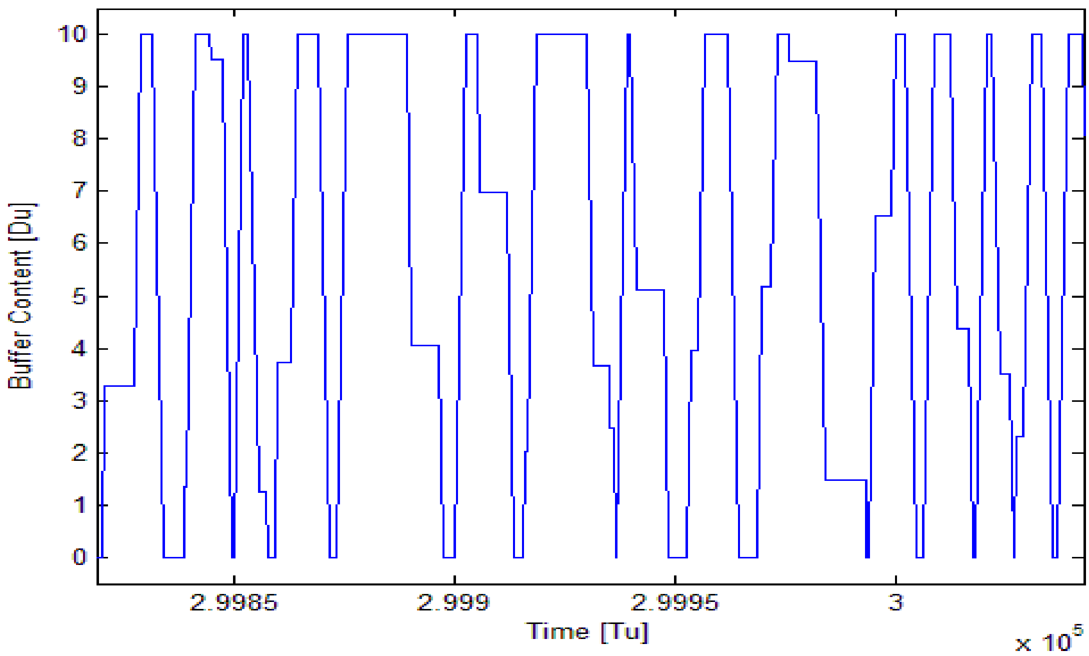


(a) Behavior of Play-out Buffer

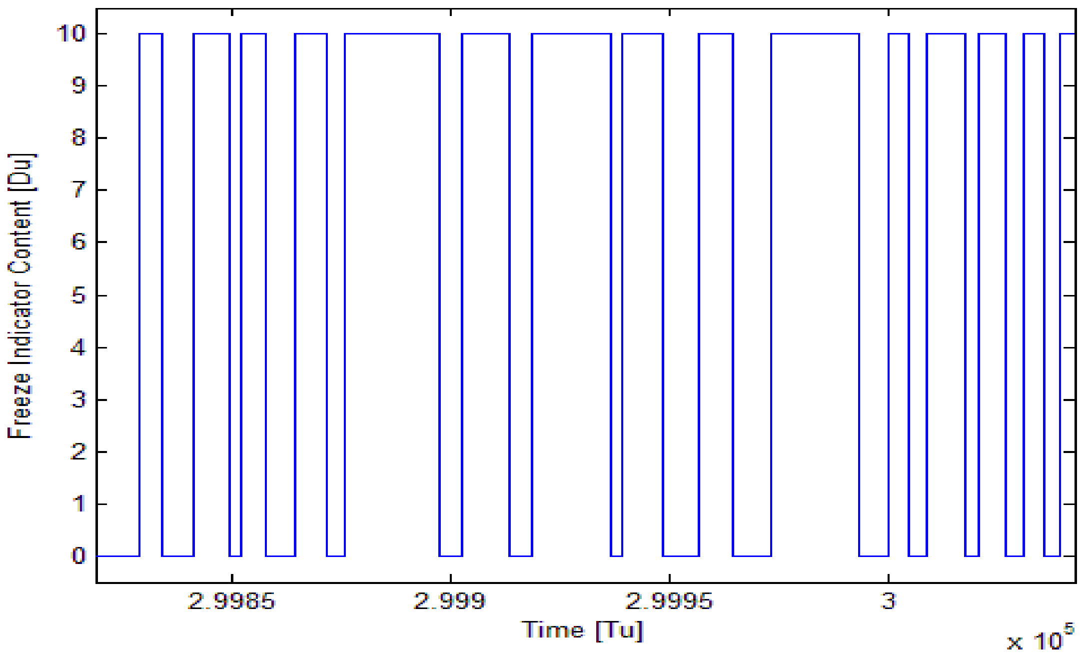


(b) Behavior of Freeze Indicator

Figure 4.4: Content of play-out buffer and freeze indicator for last 20 state changes (k=5).



(a) Behavior of Play-out Buffer



(b) Behavior of Freeze Indicator

Figure 4.5: Content of play-out buffer and freeze indicator in the last 20 state changes (k=10).

Scenario 2: Large ET_{off}***, Short*** ET_{on}

In this scenario, the system enters OFF states with large T_{off} and enters ON states with short T_{on} condition. This means the time that data goes out of the buffer (T_{off})

is large enough to make the play-out buffer and freeze indicator completely empty. In addition, the time that data flows into the buffer (T_{on}) is too short to make the play-out buffer and freeze indicator completely full. Table 4.4 represents the parameters in this scenario.

Table 4.4: Simulation parameters for results shown in Figure 4.6 and Figure 4.7.

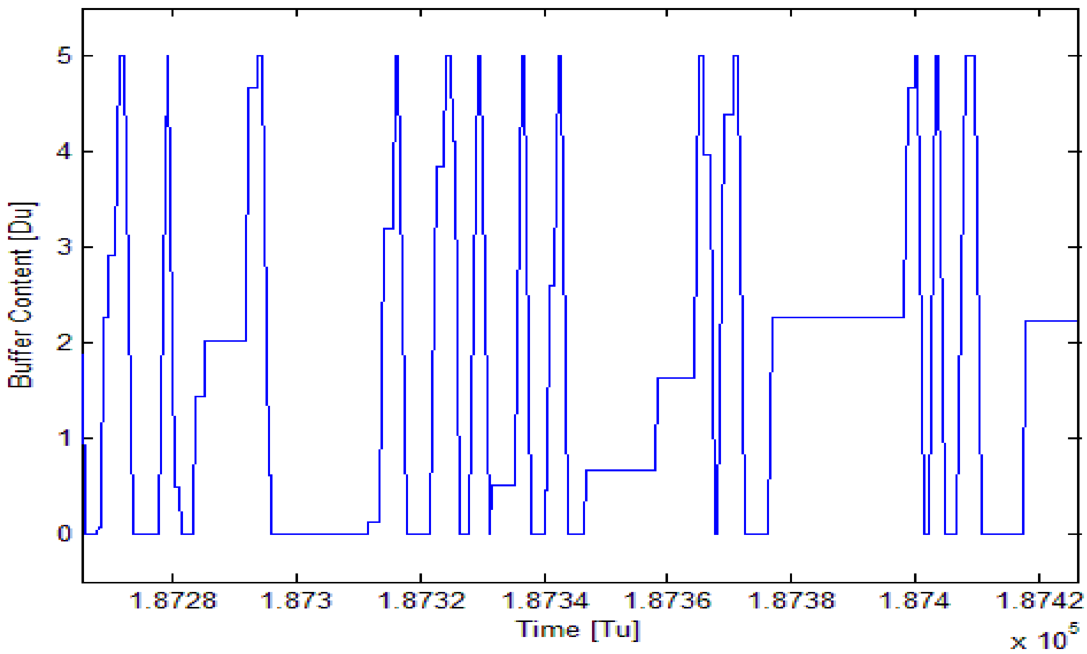
Parameter	***Value Group1***	***Value Group2***
Buffer Size k (Du)	5	10
Mean On Time ET_{on} (Tu)	0.75	0.75
Mean Off Time ET_{off} (Tu)	3	3
Input Rate C_{on} (Du/Tu)	4	4
Input Rate C_{off} (Du/Tu)	4	4
Number of On & Off State Changes	100,000	100,000

The results in Table 4.5 explain when ET_{off} is large and ET_{on} is short, increasing the buffer size has little impact on freeze indicator probabilities. Since T_{off} is large enough to make both buffer sizes empty, for the larger buffer size (k=10) the probability of video playback is a bit more than the smaller buffer size (k=5). Because for the larger buffer size, it needs a bit longer time to make the buffer completely empty during a single state.

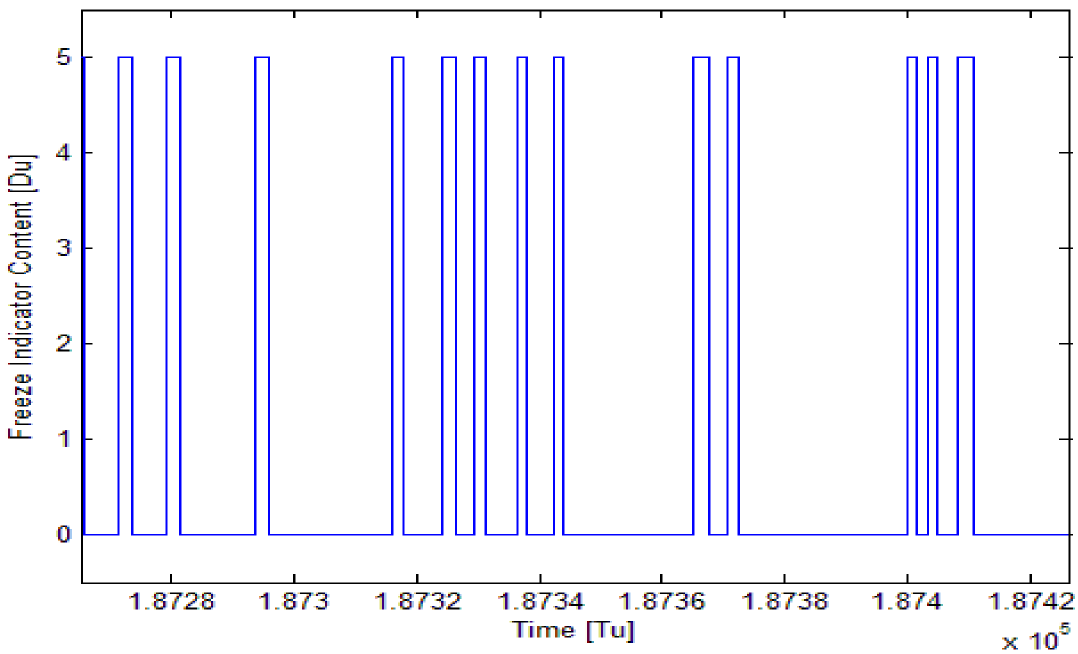
Table 4.5: Probability of full or empty freeze indicator for k=5 and k=10.

Parameters	***Results***	***Results***
Buffer Size	$Pr[F = k]$	$Pr[F = 0]$
k = 5	0.18	0.82
k = 10	0.21	0.79

Figure 4.6 and Figure 4.7 display the behaviors of play-out buffer and freeze indicator for the last 20 state changes when k=5 and k=10.

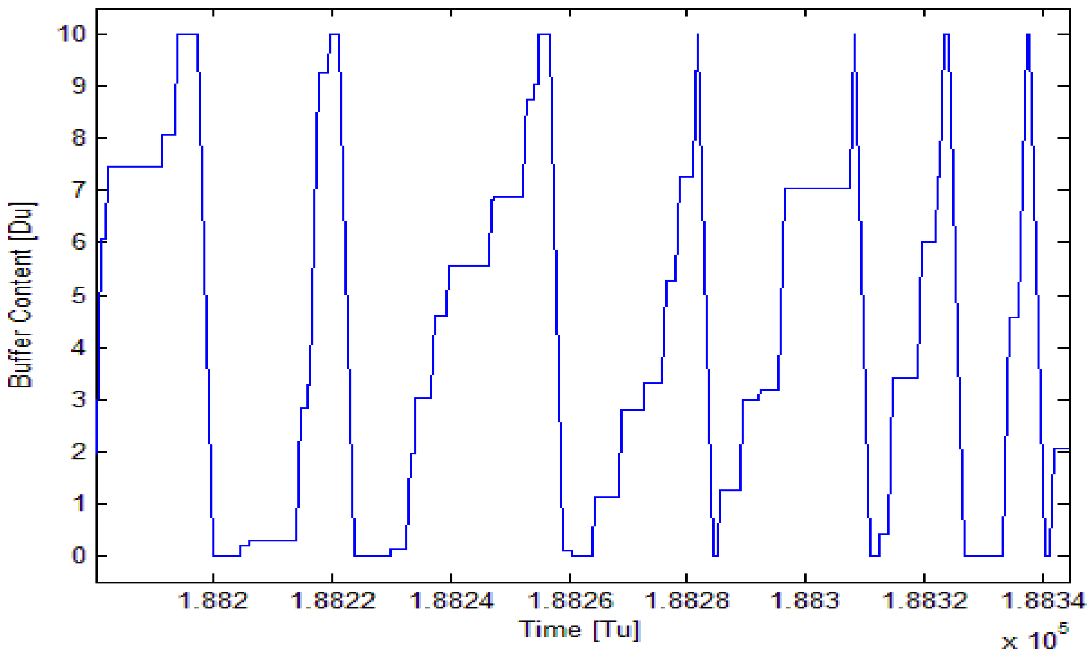


(a) Behavior of Play-out Buffer

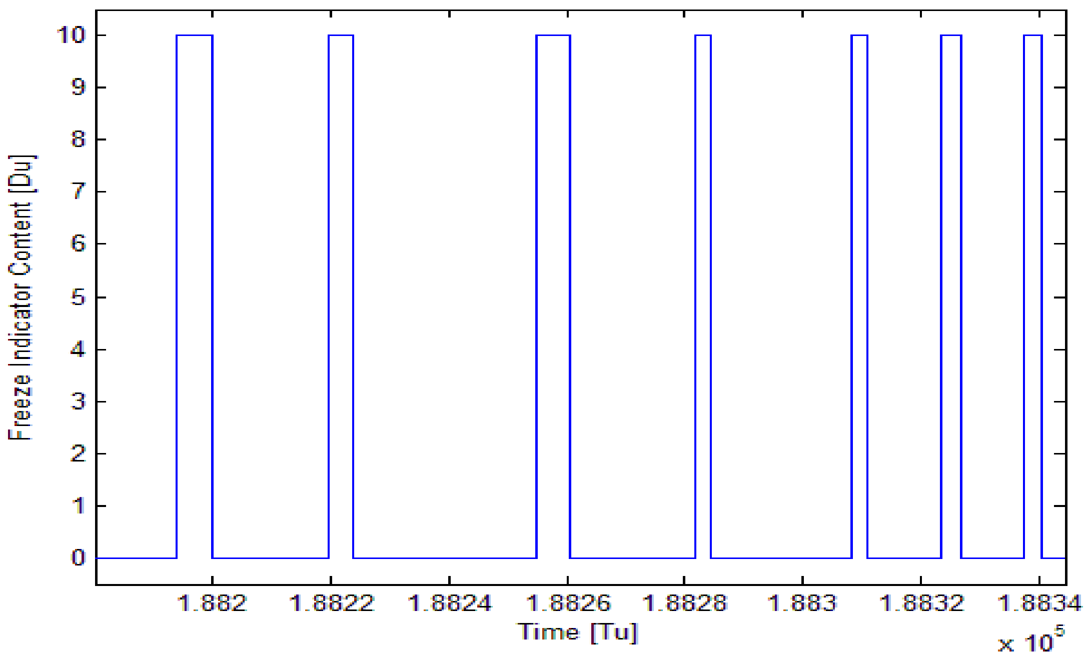


(b) Behavior of Freeze Indicator

Figure 4.6: Content of play-out buffer and freeze indicator in the last 20 state changes (k=5).



(a) Behavior of Play-out Buffer



(b) Behavior of Freeze Indicator

Figure 4.7: Content of play-out buffer and freeze indicator in the last 20 state changes (k=10).

Scenario 3: Short ET_{off}, Large ET_{on}

In this scenario, the system enters OFF states with short T_{off} and enters ON states with large T_{on} condition. It means, the time that data flows into the play-out buffer is large enough to fill the play-out buffer and freeze indicator completely and the time data goes out of the play-out buffer is too short to make the play-out buffer and freeze indicator completely empty. Table 4.6 represents the parameters in this scenario.

Table 4.6: Simulation parameters for results shown in Figure 4.8 and Figure 4.9.

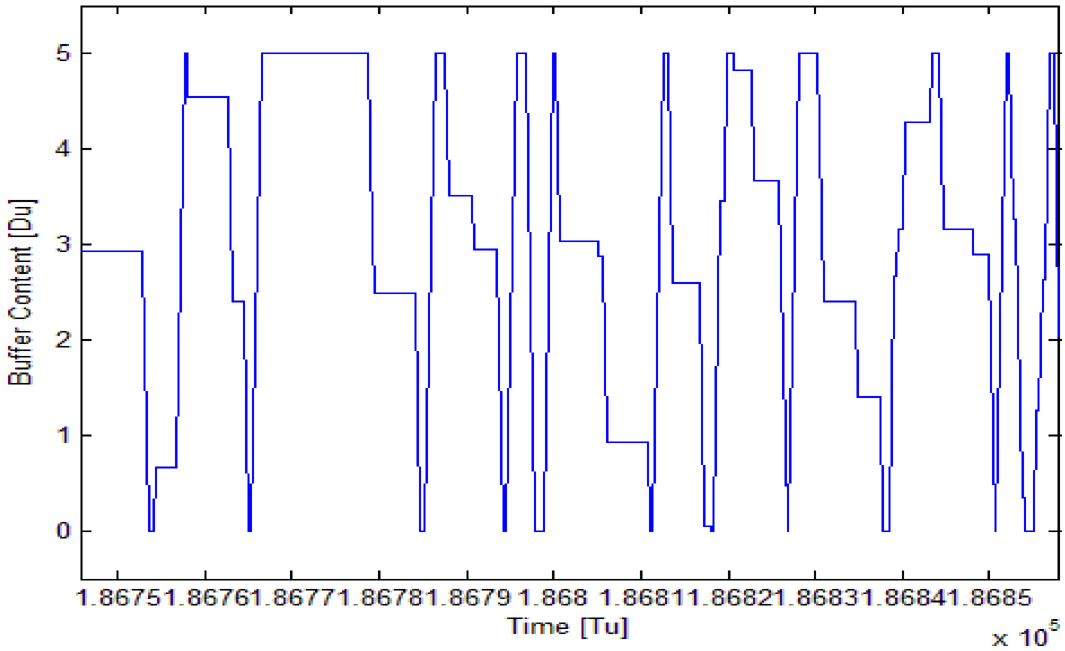
Parameter	***Value Group1***	***Value Group2***
Buffer Size k (Du)	5	10
Mean On Time ET_{on} (Tu)	3	3
Mean Off Time ET_{off} (Tu)	0.75	0.75
Input Rate C_{on} (Du/Tu)	4	4
Input Rate C_{off} (Du/Tu)	4	4
Number of On & Off State Changes	100,000	100,000

The results in Table 4.7 show when ET_{off} is short and ET_{on} is large, increasing the buffer size has slightly negative effect on freeze indicator probabilities. For example when k=5, the probability of video playback is slightly higher compared to k=10. Although, T_{on} is large enough to make both buffer sizes completely full, but for the smaller buffer size it takes shorter time to become completely full and also the play-out buffer remains full for a longer time.

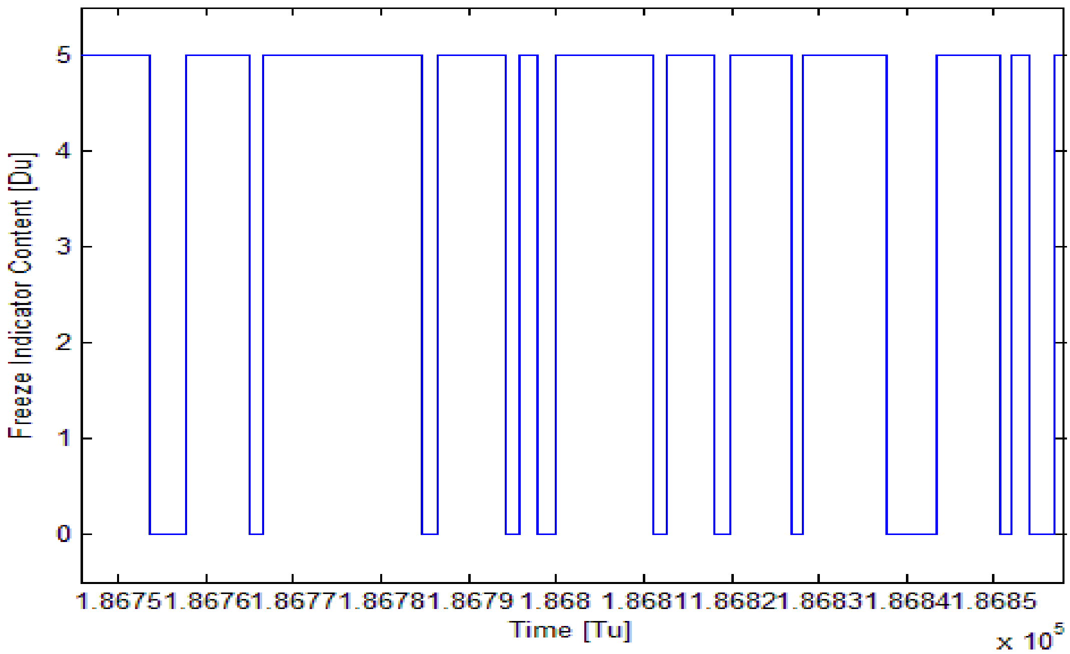
Table 4.7: Probability of full or empty freeze indicator for k=5 and k=10.

Parameters	***Results***	***Results***
Buffer Size	$Pr[F = k]$	$Pr[F = 0]$
k = 5	0.8125	0.1875
k = 10	0.79	0.21

Figure 4.8 and Figure 4.9 display the behaviors of play-out buffer and freeze indicator for the last 20 state changes when k=5 and k=10.

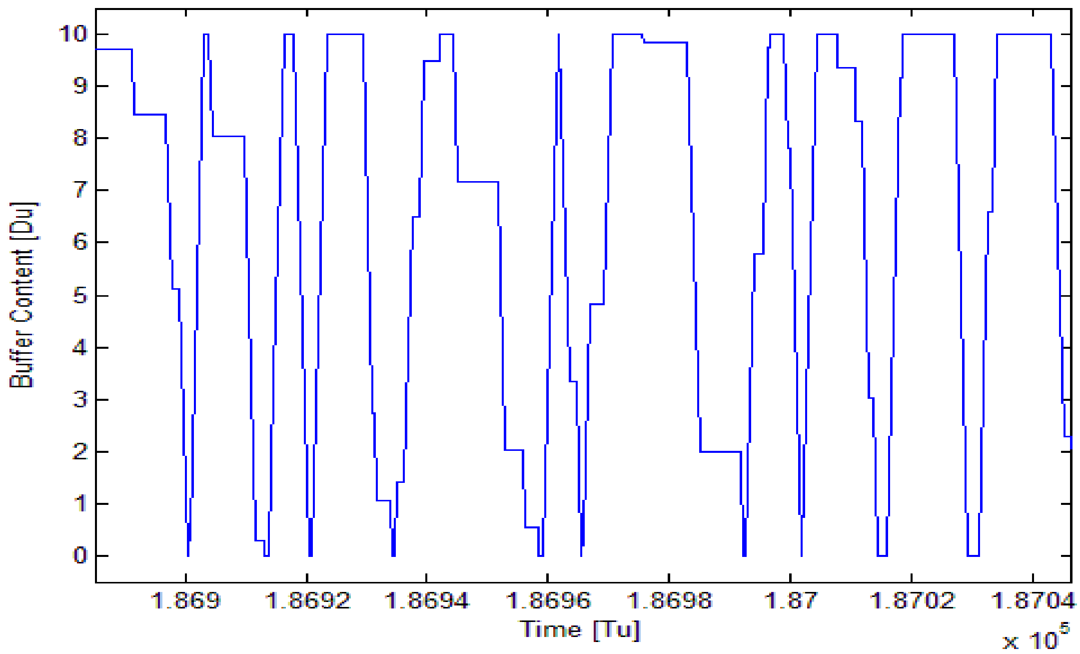


(a) Behavior of Play-out Buffer

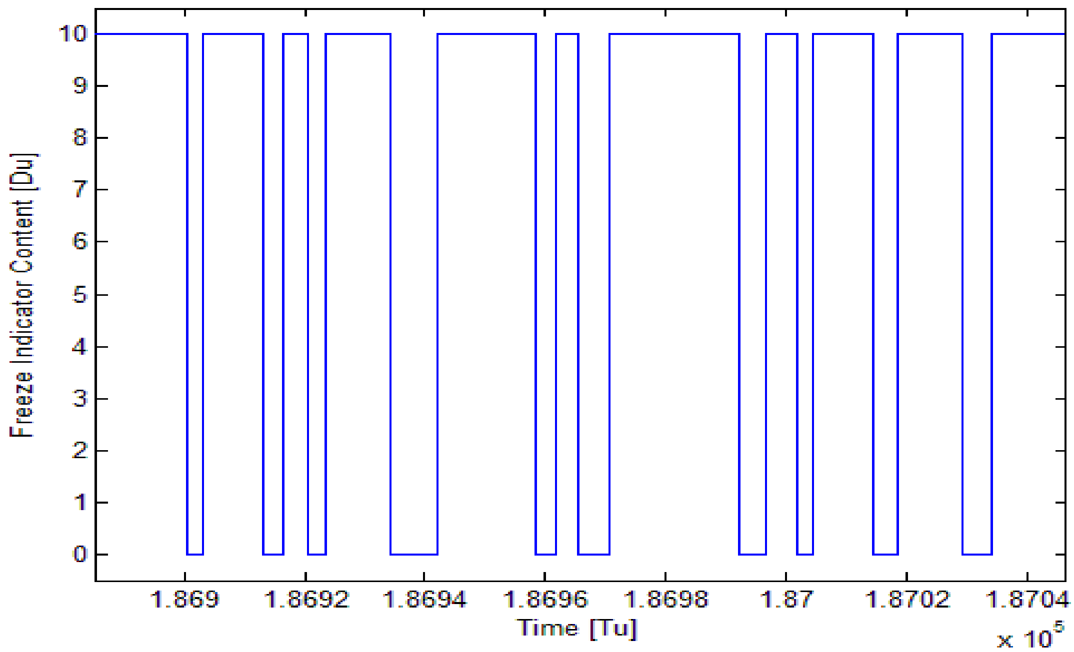


(b) Behavior of Freeze Indicator

Figure 4.8: Content of play-out buffer and freeze indicator in the last 20 state changes (k=5).



(a) Behavior of Play-out Buffer



(b) Behavior of Freeze Indicator

Figure 4.9: Content of play-out buffer and freeze indicator in the last 20 state changes (k=10).

Scenario 4: Short ET_{off}, Short ET_{on}

In this scenario, the system enters OFF states with short T_{off} and enters ON states with short T_{on} condition. This means, the time that data flows into the play-out buffer or goes out of it, is too short to make the buffer and freeze indicator completely full or empty. Table 4.8 represents the parameters in this scenario

Table 4.8: Simulation parameters for results shown in Figure 4.10 and Figure 4.11.

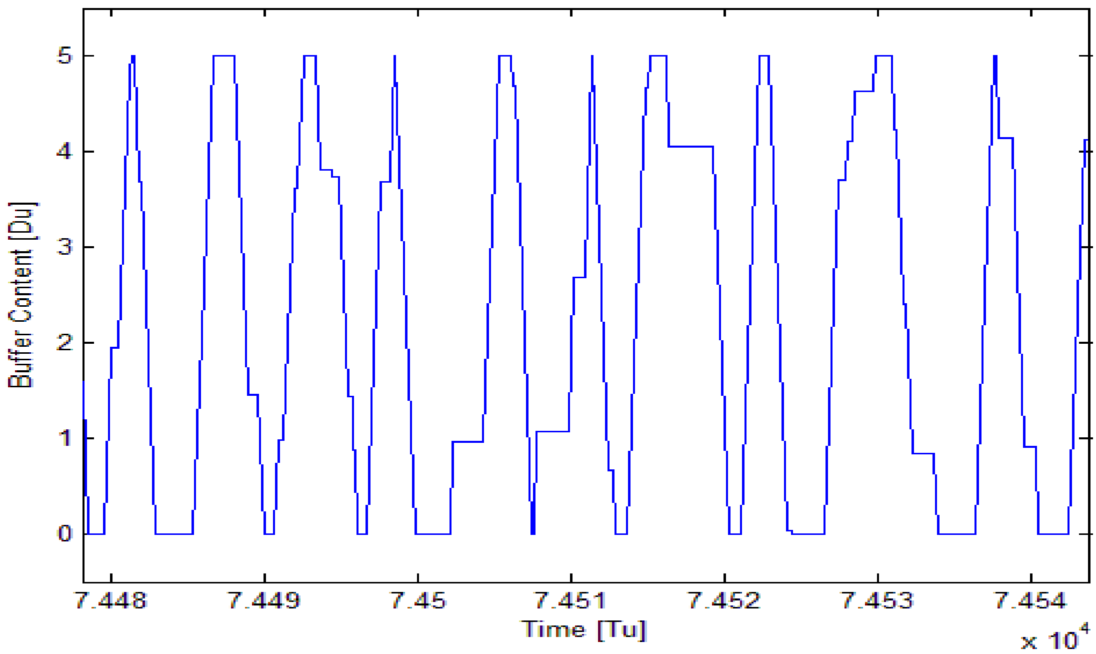
Parameter	***Value Group1***	***Value Group2***
Buffer Size k (Du)	5	10
Mean On Time ET_{on} (Tu)	0.75	0.75
Mean Off Time ET_{off} (Tu)	0.75	0.75
Input Rate C_{on} (Du/Tu)	4	4
Input Rate C_{off} (Du/Tu)	4	4
Number of On & Off State Changes	100,000	100,000

The result in Table 4.9 show when ET_{off} and ET_{on} are too short, increasing the buffer size has negative effect on freeze indicator probabilities. For instance, when k=5, the probability of video playback is higher compare to k=10. The reason is, although T_{off} and T_{on} are too short to make both buffer sizes completely full or empty, for the larger buffer size (k=10) it takes longer time to become full or empty in a single state and also the play-out buffer remains full or empty for a shorter time.

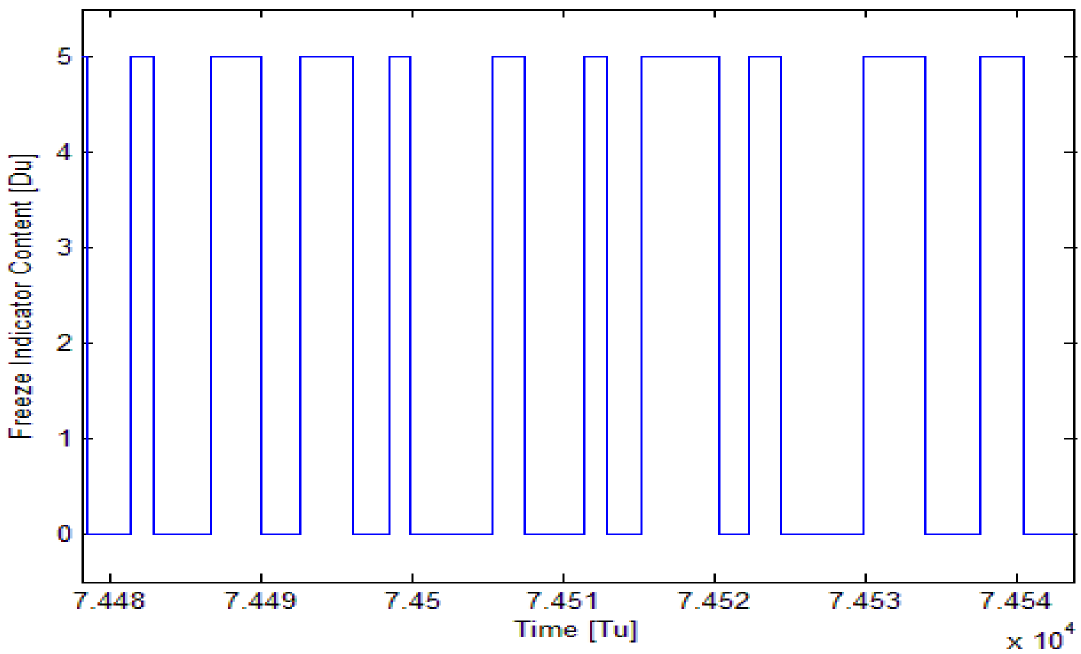
Table 4.9: Probability of full or empty freeze indicator for k=5 and k=10.

Parameters	***Results***	***Results***
Buffer Size	$Pr[F = k]$	$Pr[F = 0]$
k = 5	0.50	0.50
k = 10	0.46	0.54

Figure 4.10 and Figure 4.11 display the behaviors of the play-out buffer and freeze indicator for the last 20 state changes when k=5 and k=10.

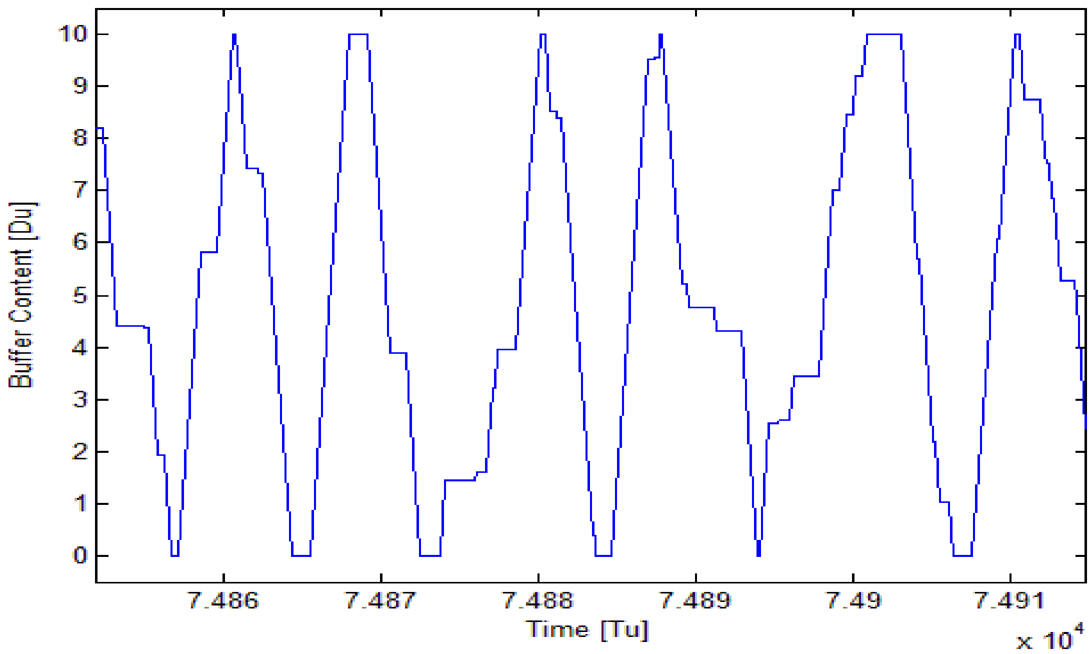


(a) Behavior of Play-out Buffer

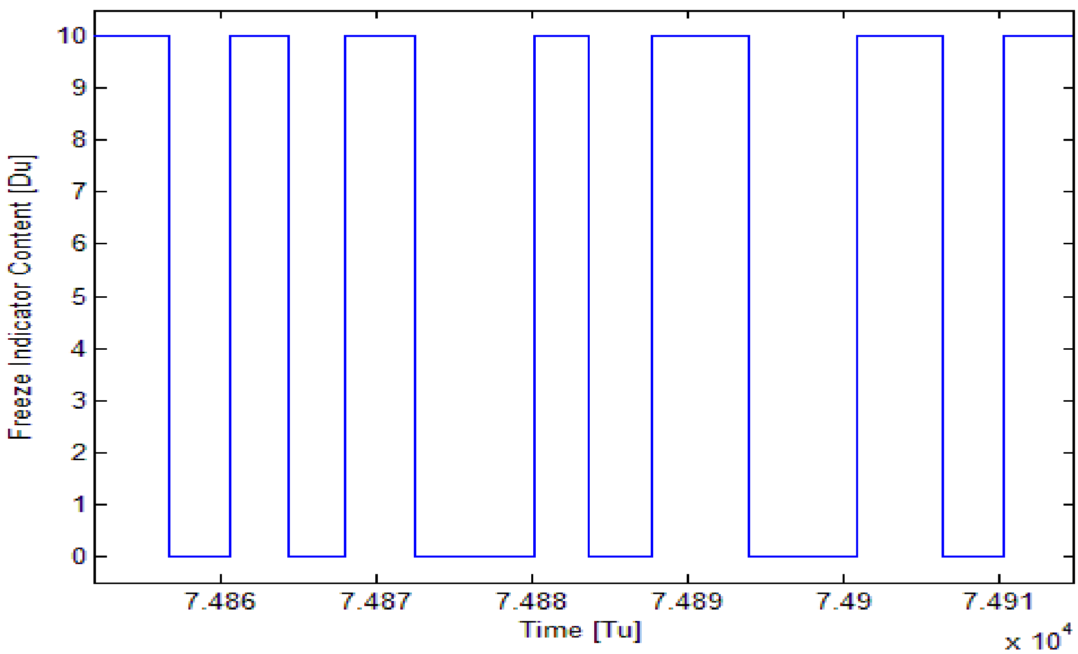


(b) Behavior of Freeze Indicator

Figure 4.10: Content of play-out buffer and freeze indicator in the last 20 state changes (k=5).



(a) Behavior of Play-out Buffer



(b) Behavior of Freeze Indicator

Figure 4.11: Content of play-out buffer and freeze indicator in the last 20 state changes (k=10).

Figure 4.4 below presents a summary of results.

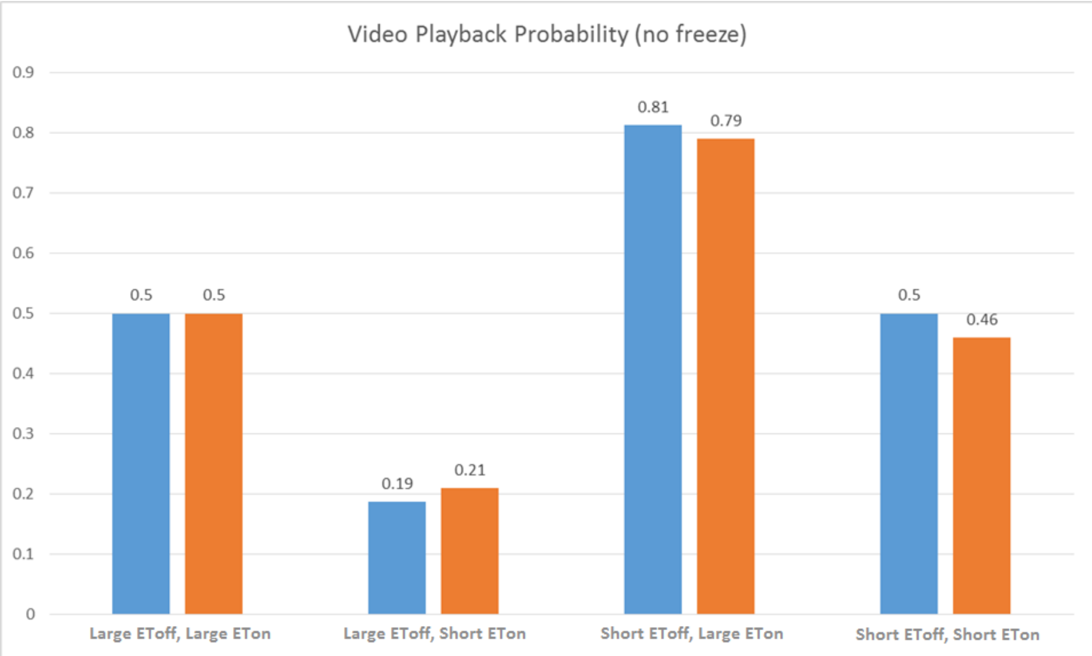


Fig 4.4: Freeze indicator full probability under different scenarios.

The probability of video playback is highest in the case when ET_{on} is large and ET_{off} is short because the average T_{on} is large enough to completely fill in the play-out buffer most of the times. In this case, the probability is slightly impacted by an increase in buffer size 'k' because the smaller play-out buffer is already not becoming empty. The video playback probability is lowest in the case when ET_{on} is short and ET_{off} is large because the higher value of T_{off} is making the play-out buffer empty most of the time. An increase in buffer size in this case will have a little impact on the playback probability since the larger buffer size needs a bit longer time to completely empty the buffer. The playback probability is also comparatively small in both cases where ET_{on} and ET_{off} are large or ET_{on} and ET_{off} are short because the average T_{on} or T_{off} is either large to completely fill/empty or short to completely fill/empty the play-out buffer in each cycle. Like the previous case, an increase in play-out buffer size, when both ET_{off} and ET_{on} are short has negative impact on playback probability of freeze indicator.

The dominating role of average off-time ET_{off} in freezing video playback is clear from the results. Freezing is affected by the duration of outages (off-times) in the data streams that are carrying the video traffic. An increase in average outage/off time cannot be compensated by an increased buffer size. In most cases, an increased buffer size has a negative impact of the freeze probability.

5 CONCLUSIONS AND FUTURE WORK

We created a fluid flow model for video transmission over a mobile link and derived an approximate formulate for freeze probability. The model is simulated and the results show that video playback freeze issue cannot be directly controlled by increasing the play-out buffer size. We also see that under different possible combinations of on/off times, a larger play-out buffer generally has a negative impact of freeze probability.

However, the main lesson learned from this work is that we should focus on addressing issues that play a dominant role is lowering the freeze probability, i.e. reducing the off-times as compared to on-times than increasing the play-out buffer size.

To elevate QoE of streamed video over a mobile link, the buffer lengths should be kept small. It is better to sacrifice resolution just in order to put less bandwidth pressure on weak mobile link and to keep the video coming. This leads to less problems with outages, in particular less frequent and shorter off times. Shifting the focus from off to on times is much better than increasing the buffer size.

This work introduces a fluid flow modeling approach for mobile video transmission which abstracts video streams to fluid streams that allows for an approximation of formula for freeze probability in the system. This could serve as a starting point for more advanced research, where this model is extended with other factors that could potentially contribute to freeze.

Additionally, average off/on-times can hardly be predicted in real-life situation due to their exponential nature and memory-less character. A future work could be to validate the findings from this model via experimentation through an actual video transmission over a mobile link. QoE parameters should also be quantitatively validated through specific user tests.

6 REFERENCES

[1] Deloitte, "Digital Media: Rise of On-demand Content," 2015. [Online]. Available: https://www2.deloitte.com/content/dam/Deloitte/in/Documents/technology-media-telecommunications/in-tmt-rise-of-on-demand-content.pdf.

[2] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015–2020 White Paper," 2016. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual networking-index-vni/mobile-white-paper-c11-520862.html.

[3] K. Kerpez, D. Waring, G. Lapiotis, J. B. Lyles and R. Vaidyanathan, "IPTV Service Assurance," IEEE Communications Magazine, vol. 44, no. 9, Sept. 2006 pp. 166-172.

[4] H. Zepernick, L. Lundberg, P. Arlos, M. Pettersson and M. Fiedler, "QoE-Based Cross-Layer Design of Mobile Video Systems: Challenges and Concepts," Proc. of International Conference on Computing and Communication Technologies, July, 2009.

[5] T. Hossfeld et al., "Initial Delays vs. Interruptions: Between the Devil and the Deep Blue Sea," Proc. of International Workshop on Quality of Multimedia Experience, Yarra Valley, Australia, July 2012.

[6] G. Hablinger and M. Fiedler, "Why Buffers in Switching Systems do not Essentially Improve QoS: An Analytical Case Study for Aggregated on-off Traffic," Proc. Internet Performance Control of Network Systems, SPIE, vol. 4865, Boston, MA, July 2002, pp.47-58.

[7] J. Gettys, "Bufferbloat: Dark Buffers in the Internet," IEEE Internet Computing, vol. 15, no. 3, May/June 2011, pp. 95-96.

[8] C. Chirichella, D. Rossi, "To the Moon and back: are Internet bufferbloat delays really that large?", Computer Communications Workshops (INFOCOM WKSHPS), 2013, IEEE Conference, pp. 417-422.

[9] M. Fiedler, "On the limited potential of buffers to improve quality of experience", Pervasive Computing and Communications Workshops (PERCOM Workshops), March, 2014, IEEE International Conference, pp. 419-424.

[10] S. Mastoureshgh, "Measurement and Method for Receiver Buffer Sizing in Video Streaming", Doctoral dissertation, Worcester Polytechnic Institute, USA, 2012.

[11] M. Youssefi, "Video Quality in IP Based Mobile Systems," Department of Signal and Systems, Master Thesis, Chalmers University of Technology, Göteborg, Sweden, 2008.

[12] R. Casellas, “Performance Evaluation of MPLS Load Sharing”. Proc of Workshop on the Design and Performance Evaluation of 3G Internet Technologies – IEEE MASCOTS 2002 Workshops, Fort Worth, USA, 2002.

[13] W. Scheinhardt, "Markov-Modulated and Feedback Fluid Queues," Phd Thesis, University of Twente, Twente, The Netherlands, 1998.

[14] Z. He, and H. Xiong, “Transmission Distortion Analysis for Real-Time Video Encoding and Streaming over Wireless Networks,” IEEE Transactions, 2006, pp. 1051-1062.

[15] A. Ziviani, B. E. Wolfinger, J. F. Rezende, O. C. Duarte, and S. Fdida, “Joint Adoption of QoS Schemes for MPEG Streams”. Multimedia Tools Appl. Vol. 26, May 2005, pp. 59-80.

[16] 3GPPTSG-SSA. Transparent End-to-end Packet-switched Streaming Service (PSS), Protocols and Codecs (Release 9). TS 26.243 V9.4.0, 2010.

[17] Adobe. HTTP Dynamic Streaming on the Adobe Flash Platform. White Paper, 2010.

[18] Youtube LLC. http://www.youtube.com/, 2011.

[19] A. Ben-Huroch, J. Weston, “A User’s Guide to Support Vector Machines,” in Methods in Molecular Biology, Vol 609, 2010, pp. 223-239.

[20] A. J. Cadzow, “Foundations of Digital Signal Processing and Data Analysis,” MacMillan, New York, 1987.

[21] J. Beran, R. Sherman, M. Taqqu and W. Wilinger, “Long-Range Dependence in Variable-Bit-Rate Video Traffic,” IEEE Transactions on Communications, vol. 43, no. 234, April 1995, pp. 1566-1579.

[22] T. Berger, “Rate Distortion Theory, Mathematical Basis for Data Compression,” Engleweed Cliffs, NJ: Prentice Hall, 1971.

[23] J. H. Sarker, “Models for Video Traffic,” Helsinki University of Technology, Finland, 1999.

[24] K. De Vogeleer, “On the Quality of Delivery,” Locenciated Thesis, Blekinge Institute of Technology, Karlskrona, Sweden, 2012.

[25] J. W. Bosman, R. D. Van Der Mei and R. Nunez-Queija, “A Fluid Model Analysis of Streaming Media in the Presence of time-varying Bandwidth,” Proc of the International Tele traffic Congress, 2012.

[26] G. Liang and B. Liang, “Balancing Interruption Frequency and Buffering Penalties in VBR Video Streaming,” Proc. of IEEE International Conference on Computer Communications, May 2007, pp. 1406-1414.

[27] A. Parandehgheibi, M. Medard, S. Shakkottai and A.E. Ozdaglar, ”Avoiding Interruptions- QoE Trade-Offs in Block-Coded Streaming Media Applications,” The Computing Research Repository abs/1001.1937, 2010.