Multiple Description Coding for Adaptive QoS Mechanism for Mobile Cloud Computing

Ilan Sadeh\textsuperscript{1}, Stojan Kitanov\textsuperscript{1}, and Danco Davcev\textsuperscript{2}

\textsuperscript{1} University for Information Science and Technology "St. Paul the Apostle", Ohrid, Macedonia
ilansd@yahoo.com, stojankitanov@hotmail.com
\textsuperscript{2} Faculty of Computer Science and Engineering, "Ss. Cyril and Methodious" University, Skopje, Macedonia
danco.davcev@finki.ukim.mk

Abstract. Multimedia transmission over cloud infrastructure is a hot research topic worldwide. It is very strongly related to video streaming, VoIP, mobile networks, and computer networks. The goal is a reliable integration of telephony, video and audio transmission, computing and broadband transmission based on cloud computing. Multiple Description Coding (MDC) is the right approach to pave the way for mobile multimedia and cloud computing. It provides full separation of multimedia and text communication in the Cloud Computing and Mobile Cloud Computing techniques. Like that Multiple Description Coding improves the Quality of Service and provides new service of rate adaptive streaming. This paper presents a new approach for improving the quality of multimedia and other services in the cloud. Furthermore, it introduces a subset of Quality of Service that considers the blocking in multi-terminal multimedia network and fidelity losses.

Keywords: Cloud Computing (CC), Mobile Cloud Computing (MCC), Quality of Service (QoS), Grade of Service (GoS), Quality of Experience (QoE), Multiple Description Coding (MDC), multimedia, rate adaptive, streaming.
Multiple Description Coding for Adaptive QoS Mechanism for Mobile Cloud Computing

Ilan Sadeh¹, Stojan Kitanov¹, and Danco Davcev²

¹ University for Information Science and Technology, “St. Paul the Apostle,”
Ohrid, Republic of Macedonia
² Ss Cyril and Methodius University, Faculty for Computer Science and Engineering,
Skopje, Republic Macedonia
{ilansd@yahoo.com, stojan.kitanov@uist.edu.mk, danco.davcev@finki.ukim.mk}

"This work was partially financed by the Faculty of Computer Science and
Engineering at the “Ss.Cyril and Methodius University.”

Abstract. Multimedia transmission over cloud infrastructure is a hot research
topic worldwide. It is very strongly related to video streaming, VoIP, mobile
networks, and computer networks. The goal is a reliable integration of
telephony, video and audio transmission, computing and broadband
transmission based on cloud computing. Multiple Description Coding (MDC)
is the right approach to pave the way for mobile multimedia and cloud
computing. It provides full separation of multimedia and text communication in
the Cloud Computing and Mobile Cloud Computing techniques. Like that
Multiple Description Coding improves the Quality of Service and provides new
service of rate adaptive streaming. This paper presents a new approach for
improving the quality of multimedia and other services in the cloud.
Furthermore, it introduces a subset of Quality of Service that considers the
blocking in multi-terminal multimedia network and fidelity losses.

Keywords: Cloud Computing (CC); Mobile Cloud Computing (MCC); Quality
of Service (QoS); Grade of Service (GoS); Quality of Experience (QoE);
Multiple Description Coding (MDC); multimedia; rate adaptive; streaming

1 Introduction

Together with the explosive and rapid growth of Internet, mobile networks, mobile
applications, and cloud computing, mobile cloud computing is introduced as a
potential technology for mobile devices. As mobile network infrastructures
continuously improve, their data transmission becomes increasingly available and
affordable, and thus they are becoming popular clients to consume any internet web-
based applications.

The “cloud” in cloud computing is defined as the set of hardware, networks,
storage, services, and interfaces that combine to deliver aspects of computing as a
service. Cloud computing provides delivery of services, storage, software and
processing capacity over internet, reducing cost, increasing, automating systems,
decoupling of service delivery from underlying technology, and providing flexibility
and mobility of information. It provides computing utilities and resources that can be leased and released by the customers through the Internet in an on-demand fashion [1, 2].

On the other hand Mobile Cloud Computing (MCC) refers to an infrastructure where both the data storage and the data processing happen outside of the mobile device. Mobile cloud applications move the computing power and data storage away from mobile phones and into the cloud, bringing applications and mobile computing to not just smartphone users but a much broader range of mobile subscribers. Mobile Cloud Computing (MCC) integrates the cloud computing into the mobile environment and overcomes the obstacles related to the performance (battery life, storage, and bandwidth), environment (heterogeneity, scalability and availability), and security (reliability and privacy) [3, 4].

A General Mobile Cloud Computing Model that is made up of complex network and relationships of and in between Infrastructure Providers, Application/Services Providers, End-Users and Developers all producing and/or consuming applications and/or services on internet is given in [5]. In such a network model different Cloud Computing consumers have different Quality of Service (QoS) requirements of applications against a service provider’s capabilities and usage terms [6]. For future Internet, QoS-aware broadband multimedia applications provided by cloud computing and contents delivery services are expected to be the major traffic generator and consumer.

As a result automated negotiation is needed to accommodate different consumer’s QoS requirements. The result of such a negotiation is a Service Level Agreement (SLA), an electronic contract that establishes all relevant aspects of the service. The SLA contains the QoS requirements of Applications hosted by a cloud based computing platform, such as timeliness, scalability, response-time, throughput, failure probability, and dependability (availability, security, safety, reliability, etc.). The SLA guarantees the corresponding QoS during the applications execution, especially important for sensitive and critical applications. The QoS aspects in a network are influenced by a large number of factors such as: Channel conditions, resource allocation policies, available resources, delay sensitive/insensitive traffic. Therefore an adaptive QoS resource reservation management algorithm is needed to exist in the network.

On the other hand, Multiple Description Coding [7, 8] is also an adequate approach for providing rate-adaptive streaming: content providers can send all the descriptions of a stream without having to think about the receiver limitations. Receivers that don’t have the possibility to sustain a certain data rate, can only subscribe to a subset of the multiple descriptions. This relieves the sender from sending streams with different rates to different receivers.

Multiple Description Coding (MDC) system is an efficient and reliable communication system aiming to improve QoS (Quality of Service) by breaking a message into more distorted replicas, which together contain enough information to reconstruct the original message. A low fidelity reproduction can be obtained if only one channel is intact. Some of the possible applications can be in the field of mobile multimedia, cellular networks and cloud computing in general. This paper will demonstrate why Multiple Description Coding is the right approach to pave the way for Mobile multimedia and Cloud Computing. The goal is a reliable integration
between telephony, video and audio transmission computing and broadband transmission based on cloud computing.

This paper proposes how the Quality of Service for Mobile Cloud Computing can be improved by using Multiple Description Coding. This paper determines the bound for the achievable rate region for Multiple Description Coding systems and presents a new approach for improving the quality of multimedia and other services in the cloud. Furthermore, it introduces a new subset of Quality of Service that takes into account the issues of blocking in multi terminal multimedia network and fidelity losses.

This paper is organized as follows. Section 2 presents the related work. Section 3 explains the advantages of Multiple Description Coding “Send and Forget” Method versus existing classical techniques TCP/IP, UDP and RTP. Section 4 defines a potentially new approach of Quality of Service that can be used in Cloud Computing. Finally Section 5 concludes the paper, and presents the future work.

2. Related Work

Recently, with the popularization and promotion of the Internet, mobile smartphone devices cloud computing and mobile cloud computing, there is a inevitably high demand to transmit image and video at real time in packet-switching networks and narrowband networks. The low computing and processing power of mobile and wireless devices, and the increasingly serious congestion problem in wireless communication networks and the Internet, along with the growing complexity of heterogeneity in networks, and the introduction of Cloud Computing and Mobile Cloud Computing have brought many challenges to the traditional video image coding.

At the moment, the vast majority of state-of-the-art codecs uses Single Description (SD) video coding [9 – 14]. However these SD video coding relies on the traditional TCP/IP protocol, where the system quality seriously declines when the network packet loss is serious. Additionally, excessive Automatic Repeat-reQuests (ARQs) will cause excessive delay, and strong Forward Error Correction (FEC) will also bring additional delay because of its complexity, seriously affecting the real-time play of the video [8].

Because of this the traditional TCP/IP protocol has failed. No future is in that method and no future is in the existing coding methods of Multimedia. In order to provide high-quality video services to the users in wireless mobile terminals, it is necessary to overcome the low operational ability in the terminal and the problem caused by unreliable transmission in the existing network. Therefore it is necessary to design video coding that has low coding complexity and strong error resilient ability. One possible solution of this issue could be a full separation: TCP/IP, or similar protocol for textual communication, and Multiple Description Coding (MDC) for Multimedia communication.

Multiple description coding (MDC) is a Source Coding technique that fragments a single media stream into a few sub-streams referred to as descriptions [7, 8]. The packets of each description are routed over multiple, disjoint paths. In order to decode the media stream, any description can be used. However, the quality improves with
the number of descriptions received in parallel. The idea of MDC is to provide “Send and Forget” method, improve error resilience to multimedia streams and improve Quality of Service.

Besides increased fault tolerance, MDC allows for rate-adaptive streaming: Content providers send all descriptions of a stream without paying attention to the download limitations of clients. Receivers that can’t sustain the data rate only subscribe to a subset of these streams, thus freeing the content provider from sending additional streams at lower data rates.

The pioneering work of Sadeh [15-18] established the Achievable Rate Region of Multiple Description Coding. The exact relations between rates and distortions in all cases were found in that work. Before Sadeh work there were known only inner and outer bounds for the Achievable Rate Region.

Despite the aforementioned advantages of MDC, SD codecs for multimedia communications still dominate in industry. However the cloud computing is a new technology that should adopt this breakthrough, since it is clear that the old protocols such as TCP/IP do not fit. Mobile multimedia and cloud computing provide a very good area for application of Real-Time Multiple description coding.

3 Multiple Description Coding “Send and Forget Method” versus Classical Techniques

One of the important and innovative techniques in data transmission is Multiple Description Coding or MDC [7, 8]. It is a source coding technique that partitions the stream into n substreams (n ≥ 2). These substreams are often referred to as descriptions. The packets of each description are routed over multiple, (partially) disjoint paths. In order to decode the media stream, any description can be used, however, the quality improves with the number of descriptions received in parallel. Using this approach, a certain flexibility can be obtained in recovering from possible errors in transmission.

Since an arbitrary subset of descriptions can be used to decode the original stream, network congestion or packet loss which are common in best-effort networks such as the Internet will not interrupt the stream but only cause a (temporary) loss of quality.

Multiple Description Coding [7, 8] is also an adequate approach for providing rate-adaptive streaming: content providers can send all the descriptions of a stream without having to think about the receiver limitations. Receivers that don’t have the possibility to sustain a certain data rate, can only subscribe to a subset of the multiple descriptions. This relieves the sender from sending streams with different rates to different receivers.

The Achievable Rate Region of Multiple Description Coding, as well as the exact relations between rates and distortions in all cases are given in [15-18].

Multiple Description Coding is much superior than the existing network transmission technologies, such as TCP/IP, UDP and RTP, with regard to with regard to: packet loss, delay, jitter, blocking probability or consistence, bit/packet error rate probability, costs, network congestion.

In the TCP/IP protocol, widely used in computer networks, packet loss is a normal event [19]. When a packet loss occurs, the TCP/IP protocol sends extra packets that
repeat the information lost, multiplying the data rate sent. In the best case, this effort causes an important delay in the transmission, while very often it pushes the entire network into a 'congestion collapse', where an important part of the packets is lost. This causes a degradation of the network throughput.

With the solution proposed in this work, the delay caused by packet retransmission is reduced to zero. "Send and forget" over two parallel channels performs better than TCP/IP, with regard to packet loss, network congestion and delay. Though, the TCP/IP protocol is considered improper for real-time applications such as VoIP, streaming, video-conferencing, etc. For such applications, protocols like the Real-time Transport Protocol (RTP) running over the User Datagram Protocol (UDP) are usually recommended instead [19].

UDP provides a procedure for application programs to send messages to other programs with a limited number of protocol mechanisms. The protocol is transaction oriented, and delivery and duplicate protection are not guaranteed. It has no handshaking dialogues, and thus exposes any unreliability of the underlying network protocol to the user’s program. As it assumes IP as the underlying protocol, there is no guarantee of delivery, ordering or duplicate protection [19].

In situations in which error checking and correction is either not necessary or performed in the application, UDP can be a good choice, because it avoids the overhead of processing at network interface level. Time-sensitive applications often use UDP because dropping packets is preferable to waiting for delayed packets, which may not be an option in a real-time system.

Real-time multimedia streaming applications require timely delivery of information and can tolerate some packet loss to achieve this goal. For example, loss of a packet in audio application may result in loss of a fraction of a second of audio data, which can be made unnoticeable with suitable error concealment algorithms.

An important protocol that defines a standardized format for sending video and audio over IP networks is the Real time Transport Protocol (RTP). The majority of the RTP implementations are built on the User Datagram Protocol [19]. RTP is designed for end-to-end, real-time, transfer of stream data. The protocol provides facility for jitter compensation and detection of out of sequence arrival in data, that are common during transmissions on an IP network. RTP is regarded as the primary standard for audio/video transport in IP networks.

However, the User Datagram Protocol, and as a consequence the RTP protocol, suffer from possible blocking of the transmission, packet loss, bit/packet errors. In addition, there might be a significant variation in delay or jitter, which is the most frustrating parameter for certain devices. With the proposed approach, the average delay will be reduced and, as a consequence, relevant reduction of jitter is expected. These are closely related to QoS.

Sending the message through two (possibly) independent channels is an application of the general idea of Multiple Description Coding (MDC) [7, 8]. Pushing a signal through two parallel channels can only reduce the bit/packet-error rate probability. Even if one of the channels emits a faulty bit/packet, the receiver can still use the bit/packet emitted by the other channel to reconstruct the correct signal. In this case an error detection code can be used to detect the errors, if any, and eventually switch to the correctly transmitted bit/packet.
The approach here proposed in this paper reduces the blocking probability in communication network, and under certain conditions it can be very low. Transmitting the signal through a channel pair, significantly reduces the costs of the physical mean of communication, because now the signal transmission can be with channels of lower capacity.

For all these reasons, the approach proposed in this paper can be considered a very promising technique for use in information networks that support real-time applications such as video streaming, VoIP, mobile telephony, and particularly Cloud Computing, and Mobile Cloud Computing.

4 New Approach to Quality of Service

Quality of Service was defined by the ITU in 1994 [6] to include basic criteria on the characteristic parameters of a connection, such as service response time, loss, signal-to-noise ratio, cross-talk, echo, interrupts, frequency response, loudness levels, and so on. As a subset of telephony QoS is Grade of Service (GoS) criteria to include aspects of a connection in relation to the capacity and coverage of a network. An example is guaranteed maximum blocking probability and outage probability [20].

When considering the computer networks or other kinds of networks for data communication, some kind of resource reservation can be applied. With this approach to Quality of Service, the most important to consider is the possibility to differentiate the available resources for the end user. The level of performance depends on the type of subscription or on the type of user. In some cases, a certain bit rate, delay, jitter, blocking probability, packet dropping probability and/or bit error rate may be guaranteed. When the network capacity is not sufficient, guaranteeing good parameters for the Quality of service can be very important.

This is especially true in real-time streaming audio/video applications, such as VoIP, online gaming and IP-TV, as these often require fixed bit rate and are delay sensitive. The same reasoning applies for other networks with a limited capacity such as the mobile communication networks.

A network or protocol supporting QoS may agree on a traffic contract with the application software and reserve capacity in the network nodes, for example during a particular session in the establishment phase. During the session it may monitor the achieved performance level, for example the data rate and delay, and dynamically control the scheduling priorities in the network nodes. It could potentially release the reserved capacity during a tear down phase. In a best effort network or service usually Quality of Service is not supported. An alternative approach to control mechanisms is to provide a high quality communication over a best-effort network. This is done by over-provisioning the capacity so that it is sufficient for the expected peak traffic load. In such case network congestion is avoided and there is no need for QoS mechanisms.

Another definition for Quality of Service is related to services based on the application level. As an example of such services are telephony networks or video and audio streaming. When considering these kind of services the Quality of Service is used as a mean for the evaluation of the user satisfaction. From this point of view,
QoS is the acceptable cumulative effect on subscriber satisfaction of all imperfections affecting the service. Other terms with similar meaning are the Quality of Experience (QoE) subjective business concept, the required user perceived performance [20] the required degree of satisfaction of the user or the targeted number of happy customers. Examples of measures and measurement methods are Mean Opinion Score (MOS), Perceptual Speech Quality Measure (PSQM) and Perceptual Evaluation of Video Quality (PEVQ).

Here a subset of QoS is presented, namely the pair of: Quality of Fidelity (QoF) and Probability of Blocking, which comprises aspects of a connection related to the capacity and coverage of a network, blocking probability and fidelity of the received digitized voice or video signal. Here is assumed a standard information source with a priori known probability distribution $P$. The information could be divided in half (or other fraction) and sent over two routes. If either route is blocked or disconnected, a reduced fidelity reproduction is still available to the receiver. This concept can be applied to the Quality of Service requirements. The values of the distortion vector $D_0; D_1; D_2$ that guarantee a level of reliable communication system should be weighted by the appropriate probabilities of channel blocking events. The result of the pair

$$P_{\text{total block}} = (p^2, QoF).$$

(1)

can be considered as a new approach to QoS (Quality of Service) taking into account all of blocking and loss of fidelity factors even for the expected peak traffic load.

Let $P_0$ is the probability that both channels function well, $P_1$ is the probability that only the first channel functions well and $P_2$ is the probability that only the second channel functions well. Obviously the probability of total failure (Blocking Event), that is both channels are blocked is the probability $P_{\text{total block}} = 1 - P_0 - P_1 - P_2$. The value of Quality of Fidelity ($QoF$) is defined as the average fidelity in the system

$$QoF = \frac{P_0*D_0 + P_1*D_1 + P_2*D_2}{P_0 + P_1 + P_2}.$$  

(2)

The pair $(P_{\text{total block}}; QoF)$ can be considered as a new subset of QoS (Quality of service) taking into account all possibilities of blocking probabilities and loss of fidelity even for the peak traffic load in multi terminal network.

### 4.1 A Simple Example Using the new Approach Quality of Service

Here will be considered a network with specific distortion vectors $D_0; D_1; D_2$. The distortion measure functions $d_0; d_1; d_2$ are known. The designer should specify the parameters of the channels: probability of channel failure (due to blocking or other reason) and capacity of the channels, according to requirements defined by the new definition of Quality of Fidelity $QoF$. Here is assumed a standard source with a priori known probability distribution $P$ of all letters in alphabet $U$. Given is that all channels are identical, have the same unknown capacity $C$ and the probability of total failure
(like Blocking Event) of a channel denoted by \( P_{\text{block}} = p \). It is assumed that the Failure Events (Blocking) of each channel are independent.

Thus, if there are have two available channels, the probability that both channels are blocked simultaneously is, assuming independence of the events of blocking a channel \( P_{\text{block}} = p^2 \), and it is much lower than the single channel case. Obviously, \( P_0 = 1 - 2p - p^2 \) is the probability that both channels function well, and \( P_1 = p - p^2 \) is the probability that only the first channel is functioning well and the same value is for \( P_2 \) the probability that only the second channel is functioning well.

The rates \( R_1; R_2 \) in the optimal system will be close to the channel capacity \( C \). Here it is necessary to calculate the optimal rates which are actually equal to the channel capacity \( R_1 = R_2 = C \) for a known standard source with probability distribution \( P \) of all letters in alphabet \( U \) (for example: typical human voice speaking English), known values of the distortion vector \( D_0; D_1; D_2 \) and known distortion measure functions \( d_1; d_2; d_0 \), that satisfy the following set of equations.

The optimal solution for \( C \) is located on the boundary of the Achievable Rate Region (ARR). To find the boundary region means to determine the Infimum over all the deterministic mapping from the input process \( u \) to the output processes \( y; v; w \) represented by all possible conditional per letter probability distributions

\[
\begin{align*}
Q &= Q_{\text{max}} \quad (3) \\
R_1 + R_2 &= 2C. \quad (4) \\
R_1 &= C = I_Q(u, v) \quad (5) \\
R_2 &= C = I_Q(u, w). \quad (6)
\end{align*}
\]

such that the average per letter distortions are less or equal to \( D_0; D_1; D_2 \), respectively. That is,

\[
\begin{align*}
E_Q d_0(u, y) &\leq D_0. \quad (7) \\
E_Q d_1(u, v) &\leq D_1, \quad (8) \\
E_Q d_2(u, w) &\leq D_2. \quad (9)
\end{align*}
\]

The value of \( QoF \) the average fidelity in the system will be described by the formula

\[
QoF = \frac{P_0 + P_1 + P_2}{P_0 + P_1 + P_2} \quad (10)
\]
Finally, the value of probability channel failure (due to Blocking Event or something else) denoted by $P_{block} = p$ is easily calculated for a given value of Quality of Fidelity QoF. The conclusion is that the pair

$$P_{total\ block} = (p^2, QoF).$$

is a good criterion for the Quality of Service of communication system characterized by typical standard source $u$ with probability distribution $P$ of letters in alphabet $U$.

Some of the possible applications can be in the field of mobile multimedia, cellular networks and cloud computing in general. A digitized voice or video signal could be divided in half and sent over two routes. If either route is blocked or disconnected, a reduced fidelity reproduction is still available to the receiver. Multiple Description Coding improves Quality of Service and provides new service of rate adaptive streaming.

5 Conclusion and Future Work

This paper demonstrated that Multiple Description Coding (MDC) is the right approach to pave the way for Mobile multimedia and Cloud Computing. Firstly it was provided the advantages of MDC over technologies, such as TCP/IP, UDP and RTP. Then a new approach for improving the quality of multimedia and other services in the cloud was presented. Here, a new subset of Quality of Service, that takes into account issues of blocking in multi-terminal multimedia network and fidelity losses was introduced. Finally it was concluded that Multiple Description Coding system is an efficient and reliable communication system aiming to improve QoS (Quality of Service) by breaking a message into more distorted replicas, which together contain enough information to reconstruct the original message. A low fidelity reproduction can be obtained if only one channel is intact. Some of the possible applications can be in the field of mobile multimedia, cellular networks and cloud computing in general. A digitized voice or video signal could be divided in half and sent over two routes. If either route is blocked or disconnected, a reduced fidelity reproduction is still available to the receiver.

In order to provide high-quality video services to the users in wireless mobile terminals it is necessary to design video coding that has low coding complexity and strong error resilient ability. One possible solution is to make a full separation between multimedia and textual communication: TCP/IP, or similar protocol for textual communication, and Multiple Description Coding (MDC) for Multimedia communication.

Despite the aforementioned advantages of MDC, SD codecs for multimedia communications still dominate in industry. However the cloud computing is a new technology that should adopt this breakthrough, since it is clear that the old protocols such as TCP/IP do not fit. Mobile multimedia and cloud computing provide a very good area for application of Real-Time Multiple description coding.

In future we plan to test the subjective perspective to determine if the new improved QoS metrics are correct.
References