An Advanced Video Interactivity Framework for IPTV Services

Lesang V. Dikgole, Neco Ventura
Department of Electrical Engineering
University of Cape Town, Rondebosch, South Africa
Tel: +27 21 6502813, Fax: +27 21 6505296
email: {lesdikgole, neco,}@crg.ee.uct.ac.za

Abstract-Research has highlighted a need for a comprehensive interactivity framework for Internet Protocol Television (IPTV) services. Standardized IPTV architectures do not adequately address this need. Current video access techniques deployed over the Internet such as download and content-caching techniques offer limited video interactivity functions. This paper presents an advanced interactivity framework for IPTV services. The proposed framework is designed over the IP Multimedia Subsystem (IMS) architecture. The IMS based interactivity platform demonstrates that the advanced interactivity framework can be implemented using standardized protocols such as SIP and RTSP over the IMS based IPTV system.

Index Terms—Interactivity, IMS, SIP, RTSP, IPTV

I. INTRODUCTION

IPTV services are expected to be deployed in Next Generation Network (NGN) infrastructure. Most of these NGN based services will offer higher level of service and video interactivity compared to traditional television services. However, video websites offering High Definition (HD) video broadcasts and media will challenge the adoption of NGN based IPTV services.

Research is currently being done on how next generation television systems should be deployed to rival current satellite based systems (e.g. Multi-choice) and Internet based video services. Most of the research has focused on how IPTV systems should be developed to support services and video interactions currently being offered by satellite based systems such as DVB [1]. Qualitative research done by Shin et al [2] has shown that most prospective IPTV users will be looking for a new paradigm of interactivity in IPTV services. The *new* paradigm of interactivity will be supported by the packet based IP infrastructure. IPTV users will be expecting to 'interact' with IP services (such as the web, video and audio conferencing). Users will also expect a higher level of interactivity with the video content. The platform used to deploy these *interactive* services will be a comprehensive NGN architecture that has various service enabling technologies to support full integration and interworking of services. These NGN architectures will also guarantee a high Quality of Service (QoS) to avoid service disruptions that are not expected in robust television systems.

Service selection and live service configuration functions such as switching from a live multicast session to a unicast session are expected in NGN based IPTV service. These functions are called *service interactivity* functions throughout this paper. Service interactivity may also include interacting with other services such as presence. Other interactivity functions expected in NGN IPTV services are trick modes (i.e. play, pause, etc); and video adaptation functions which allow users to change the video quality or the streaming rate during a media session. These are called *video interactivity* functions throughout this paper.

The Multimedia Home Platform (MHP) DVB-J API [Ibid. 1] has been standardized to allow for various applications to be developed to enable more interactivity over TV set-top boxes (STB). Most of the solutions offered in the DTV were however developed more than 7 years ago (ca. 2002) and are based on the MPEG2 codec. Many advances have been made in video codec and transmission technologies since then. The H.264/AVC and SVC codecs have emerged to be superior codecs.

A comprehensive video interactivity framework that will rival interactivity functions possible in DVD players and Blue-ray Discs is needed in NGN based IPTV platforms for IPTV to be successful. Due to the ever changing nature of mobile network connections, future interactivity should also include on-the-fly video adaptation. IPTV service and content providers will also need to have control on the video interactivity signals so that users will be disallowed from activating a forward request, for example, during an advertisement.

The Video on Demand (VoD) service and various interactive functions such as trick functions require large streaming and processing resources on the service provider network. Unicast streams need to be set-up to enable each client to have control on the media. Multiple media sessions (and interactivity functions) initiated by many clients will utilize large amounts of network bandwidth and incur additional processing load on the server. The processing workload will increase exponentially as the number of sessions increase. This requires a framework for *management* of interactive functions.

The focus of this paper is to develop a video interactivity framework for IPTV services. The following questions will be addressed: what is the best video access technique for video interactivity; what are the video interactivity requirements; what is the video interactivity framework and how can it be implemented and managed? Section II assesses the impact of video access techniques on interactivity functions. Section III presents the proposed advanced video interactivity framework. Section IV discusses video interactivity over the TISPAN IMS IPTV architecture. Section V presents the design of the advanced interactivity framework over the IMS. Conclusions will be provided in Section VI.

II. CURRENT VIDEO ACCESS TECHNIQUES

In this section, various video access techniques will be discussed. There are three major video access techniques used for providing unicast video streams: download, content-caching and live streaming.

Download only technique allows interactivity only after the full video has been downloaded. Trick mode functions will only be played on the downloaded video. Fig. 1 shows the flow diagram for the download technique. The client selects the content and sends a video request to the server. The video is then downloaded to the user. Interactivity is allowed *after* the full length video has been downloaded. The download only technique is used in online television sites such as AOL, ABC, and HBO. Many online movie download websites such as CINEMA NOW also only offer the download only option. There are also many other online content retail sites (e.g. Apple online) that are already offering or considering to offer video download options.

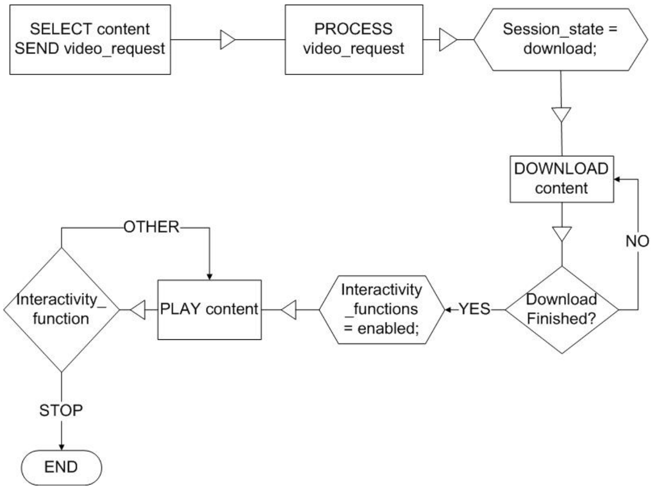

Fig. 1: Download Technique

The content-caching technique exploits various video buffering/caching mechanisms. The video content is essentially being streamed to the user, but most of the content is buffered locally before play. This means that the content is essentially being 'downloaded' from the server and not necessarily streamed as it may appear to be. Limited interactivity is allowed during the video session. Interactive functions allowed are limited to pause, play, rewind.

Other more bandwidth demanding functions such as *fast* forward, and *slow* forward are not allowed. Fig. 2 shows the flow diagram for the content-caching technique. During streaming, only a few video interactivity functions that act on the cached content are possible. Functions such as forward that require action on non-downloaded video sequences are not possible. Live video adaptation is also not fully integrated as content access has to be re-started to select a video of different quality.

Video websites such as Youtube and Google video rely on video caching techniques to stream content to video consumers. Most of the content is streamed to users indiscriminately. This provides little leverage for the content provider to monitor and manage interactivity on the streaming content.

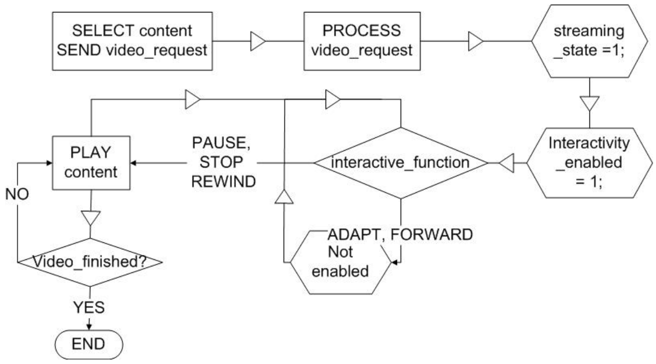

Fig. 2: Content-caching Technique

Live streaming technique commences streaming of content when the user makes a content request. Live streaming allows full interactivity during the video session and only requested streams are streamed to users. The live streaming state diagram in Fig. 3 shows the state permutations for the live streaming technique. Interactivity is controlled during streaming. The video server keeps track of the streaming user and can enable or disable an interactivity function during streaming. Since interactive modes such as fast-forward are executed on the server, the service and content providers have better control on streaming content.

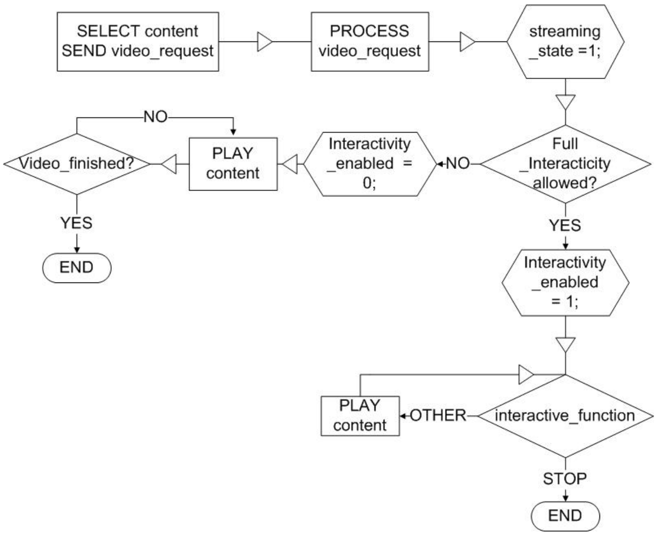

Fig. 3: Live Streaming Technique

III. THE ADVANCED VIDEO INTERACTIVITY FRAMEWORK

This section discusses requirements for video interactivity and proposes the advanced video interactivity framework for IPTV services.

A. Requirements for Video Interactivity Framework

Consider the following NGN scenarios (with requirements) for video interactivity:

Scenario 1: Consider a user requesting only to watch a middle scene in video of length 90min. The bandwidth used to download the scene should be minimal.

Requirement: a video interactivity framework should take into account the possibility of a user watching certain scenes of a video without watching the full length video. The content access technique used should ensure minimal bandwidth usage. Live streaming technique significantly reduces bandwidth usage as only requested video streams can be sent to the video client for play-out.

Scenario 2: Consider users requesting a video of length 90min. Users want to play interactive commands such as session set-up, play, and pause. Interactivity and set-up delays should be minimal.

Requirement: Content on demand systems should be designed to minimize set-up, pause and play delays. For example, the set-up delay will be unacceptable when the download technique is used. Content-caching and live streaming can significantly reduce delays.

Scenario 3a: Consider a number of users who requested to watch a movie. Most users did not watch the full movie as they did not enjoy it.

Scenario 3b: Users are using the interactive commands: forward, fast-forward and rewind. Users who forward some part of the movie do not desire to pay for 'fast' tracked content.

Requirement: Video on demand systems should be designed to integrate interactivity with video access. This will significantly reduce the amount of resources used (bandwidth, processing load, etc) on the service provider site. The consumer will only pay for the content, bandwidth, services and interactivity functions used.

B. Video Interactivity Functions

Video interactivity functions can be classified into three types: trick play, advanced interactivity and video adaptation functions. Fig. 4 shows a taxonomic use case of service interaction and video interactivity functions. *Service switching*–from multicast to unicast or vice versa-is a service interaction function.

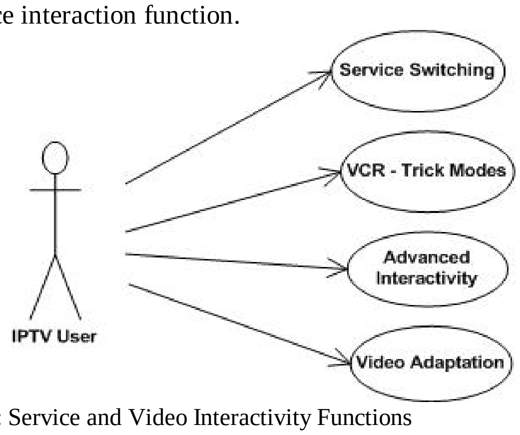

Fig. 4: Service and Video Interactivity Functions

1) Trick Modes

Trick modes include functions such as pause, play, forward and rewind.

Pause plainly involves momentarily discontinuing a video stream. Play will then resume the video.

Forward can be performed in three ways: 1) the video can be played faster so that it plays in 'fast' motion-this is equivalent to increasing the frame-rate, 2) certain video frames can be discarded to enable faster playback, 3) selective playback could be used by playing selected video sequences (i.e. scenes). Rewind reverses the playback rate, discards frames in reverse and can similarly perform selective playback in backward motion.

2) Advanced Interactivity

Advanced interactivity functions refer to functions that are not normally included in every interactive video service. These include scene selection, delayed trick modes and server initiated trick modes.

Scene selection is used when the user wants to move to a particular scene (or video sequence) without having to fast-forward to a particular video position. This interactive feature is especially desirable in robust video systems that charge users for the amount of time the user spent watching a video. Users may only be interested in watching certain sections (scenes) of the video and it is thus imperative that any modern video system provides this functionality.

Other trick functions may be disabled for a certain period and then allowed to execute later. *Delayed trick modes* may be executed by the client to provide convenience of scheduling television advertisements. For example, a user could perform a pause function and play a 'delayed play' function. This may be necessary when the user wants to leave the television screen (e.g. to briefly answer a door bell) for a certain period. This will be equivalent to setting the duration (period) of a pause function. This period or 'interlude' can be used by the service provider to stream customized advertisements to the client.

Other related 'delayed' trick modes can include *slow* forward and slow backward *motion* video. This will be enabled by allowing the client to play media at a slower play rate. The frame rate could also be increased to enhance the quality of the slow video sequence. The client can thus be able to play the video at any particular play rate. Slow motion video sequences can be executed on the video server by reducing the frame-rate.

Server initiated interactivity functions can also be desired. These can be useful for providing integration between the VoD service and other services. For example, if the user receives a phone call while watching a video, the video can be paused (automatically by the server) so that the user will be able to answer the call and resume watching later.

3) Video Adaptation Functions

Video adaptation functions are used to adapt video streams during streaming. The user can choose to upgrade or downgrade the video quality during streaming. These functions provide the user with more options and flexibility over the adapting video streams. Most video adaptation schemes do not give the client control over video stream adaptation. They either employ the use of complex IP level mechanisms that adapt video streams on the fly or rely on the streaming server to detect the changes on the network to adapt the streams accordingly. Most of these mechanisms are not widely adopted due to their high level of complexity and limited flexibility in adapting to specific needs of users. Client initiated video adaptation is therefore desired.

C. Characteristics of Video Interactivity Functions

The following characteristics of video interactivity functions (VIFs) can be drawn:

- VIFs require *quick responses* i.e. a pause functions will be expected to have low latency e.g. less than 1s
- VIFs *directly interact* with the video signal i.e. the video scene changes by being paused or altered in quality
- VIFs involve more than simple trick modes (e.g. PAUSE) they also include more *complex functions* such as scene selection and video adaptation. These functions require video interactivity to be integrated into the service to allow for controlled management of interactivity resources.

IV. VIDEO INTERACTIVITY OVER THE IP MULTIMEDIA SUBSYSTEM

The IP Multimedia Subsystem (IMS) is increasingly being adopted as an NGN platform for IPTV services by standardization bodies such as TISPAN, ITU-T and Open IPTV [3][4][5]. The IMS offers various service enabling technologies such as presence and management of user profiles. The IMS also promises to provide consumers with high Quality of Service (QoS) guarantees. There are many IPTV solutions proposed in literature for future interactive IPTV services. The IMS however presents the most credible platform for IPTV video interactivity:

- It offers high level control and management of services offered to the user
- It provides efficient user friendly charging systems
- It is an increasingly adopted standard for IPTV deployment

We adopted the TISPAN/IMS based platform for IPTV as a platform for the video interactivity framework (shown in Fig. 5).

Additional entities in the adopted platform are discussed below.

A. The IMS Architecture

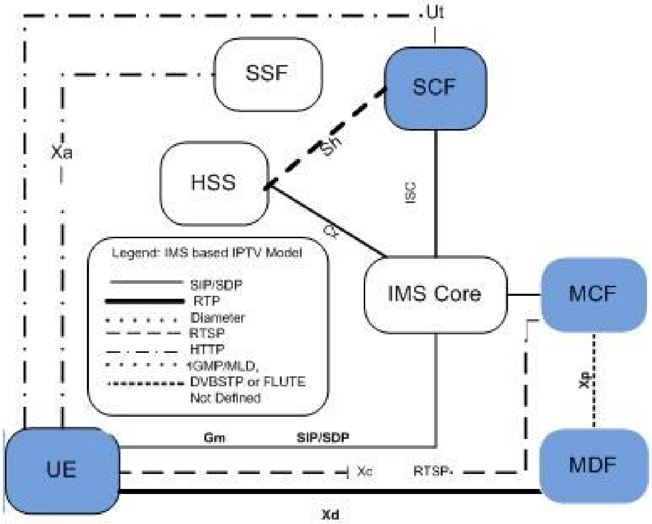

Fig. 5: TISPAN IMS Architecture [3]

User Equipment (UE): The UE is a functional entity that provides the user with access to IPTV services.

Home Subscriber Server (HSS): complete information about particular subscribers is stored in the HSS. This includes profiles, policies, subscriptions and preferences for video subscribers.

IMS Core: consists of the Call Session Control Functions responsible for controlling signaling messages and provisioning services.

Service Control Function (SCF): The SCF is a functional entity that provides IPTV service logic and the functions required to support execution of such logic.

Service Selection Function (SSF): An SSF is a functional entity that provides service selection information to the UE.

Media Control Function (MCF): The MCF is a functional entity that provides the UE with functions required to control media flows and manages the MDFs under its control.

Media Delivery Function (MDF): The MDF is a functional entity that delivers content data to the UE.

B. IMS Video Interactivity

The IMS-based TISPAN/Content-on-demand IPTV standard provides a promising interactive framework for IPTV services. The Session Initiation Protocol (SIP) [6] protocol is standardized for session set-up, while Real Rime Streaming Protocol (RTSP) [7] is used for controlling media streams.

Interactive RTSP functions such as PAUSE, PLAY, REWIND, FORWARD and STOP are possible to effect on the current standard.

According to TISPAN, the MCF is responsible monitoring and managing 'trick modes' [Ibid. 3]. Interactivity management is therefore performed at the MCF

C. Limitations of the current TISPAN Architecture

Although the current standardized IMS architecture allows for the use of the live streaming technique without incurring delay, jitter and service-disruption associated with streaming, the architecture does not provide a complete framework for video interactivity and for interactivity management.

The RTSP and SIP inter-working in the TISPAN presents challenges. RTSP only supports trick modes. Advanced interactivity requests such as live video adaptation are not fully supported. Interactivity management is also limited as RTSP requests are not rationed to limit the overuse of interactivity related resources. There is insufficient integration of the media session with the IPTV service profiles to enable advanced interactivity modes such as service switching and scene selection.

V. IMS BASED VIDEO INTERACTIVITY FRAMEWORK

The proposed architecture comprises of the following new entities that form part of the proposed Advanced Interactivity Platform (AIP):

- AIP-SCF – logically located at the SCF (shown in Fig. 6) , responsible for servicing SIP requests from the client, requests are then forwarded to the MCF
- AIP-MCF – logically located at the MCF, responsible for executing the video interactivity framework. Management of VoD resources is performed here.

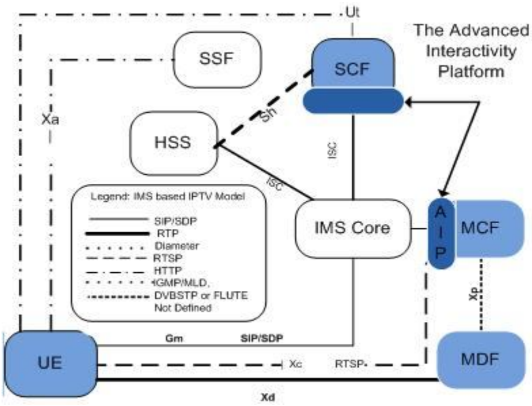

Fig. 6: The Advanced Interactivity Platform (AIP)

New interactive modes such as delayed paused are handled by the SCF. The user preferences are sent to the SCF during session set-up. A video client will send a request to the SCF when requiring video adaptation. The SCF will then send a request to the MCF to update the video stream.

A. System Design

1) Interactivity Functions

Table I below shows the processes and protocols used to implement interactivity functions. The standardized RTSP and SIP protocols are used for interactive functions. In conformance to the TISPAN standard, the RTSP protocol is used for trick modes. The SIP protocol is used for the video interactivity functions. A re-INVITE request will be sent with either an XML scheme (to update the codec profile at the MCF) or as a SIP protocol attribute value. The RTSP protocol allows for method extensions (new parameters) as well as creation of new methods [Ibid. 7]. This characteristic can be exploited to implement faster video interactivity requests that will be sent directly to the AIP-MCF.

TABLE I
INTERACTIVITY PROCESSES AND PROTOCOLS

Interactivity Function	Process	Protocol Used
SETUP	Send session setup request to the SCF	SIP
PAUSE	Send PAUSE request to the MCF	RTSP
PLAY	Send PLAY request to the MCF	RTSP
FORWARD	Send FORWARD request to the MCF	RTSP
REWIND	Send REWIND request to the MCF	RTSP
STOP	Send STOP request to the MCF	RTSP
	Advanced Trick Modes	
SCENE selection	Send scene_number # request to the SCF	RTSP/SIP
SLOW/FAST MOTION	PLAYRATE request to the SCF	RTSP/SIP
PRE_SET interactivity	Send session preferences to the SCF	RTSP/SIP
	Video Adaptation	
CHANGE video RATE	Send ADAPT, re-INVITE request to the SCF	SIP

Fig. 5 below shows the profile architecture used to manage interactivity. The standard TISPAN based architecture has variables for keeping track of the programme or movie ID and media delivery status. The proposed extension of the profile is the interactivity profile.

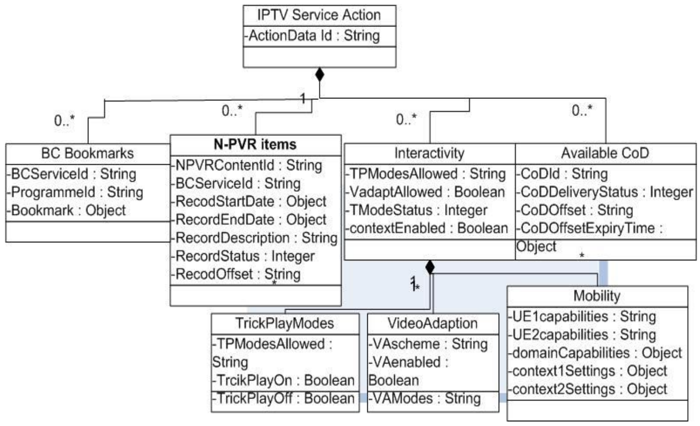

Fig. 5: User profile architecture

The interactivity profile variables include information about comprehensive call information such as which interactivity functions are allowed or disallowed at the server. For every client interactivity request, the MCF will evaluate whether the request is allowed. This will not only increase security in the media function but will also restrain users from making large number of interactivity requests that will overload the MCF. The mobility profile is also used for keeping track of user mobile contexts. This can be used in conjunction with the video adaptation profile to adapt to the environment of a video client. Unified video interactivity management is achieved by linking the SIP call information with the RTSP session information. All the interactivity requests generated by either protocol will therefore be managed and controlled by a single entity – the AIP-MCF.

2) The Interactivity Platform

To demonstrate how the interactivity platform is implemented, Fig. 6 shows an interactivity request for video adaptation.

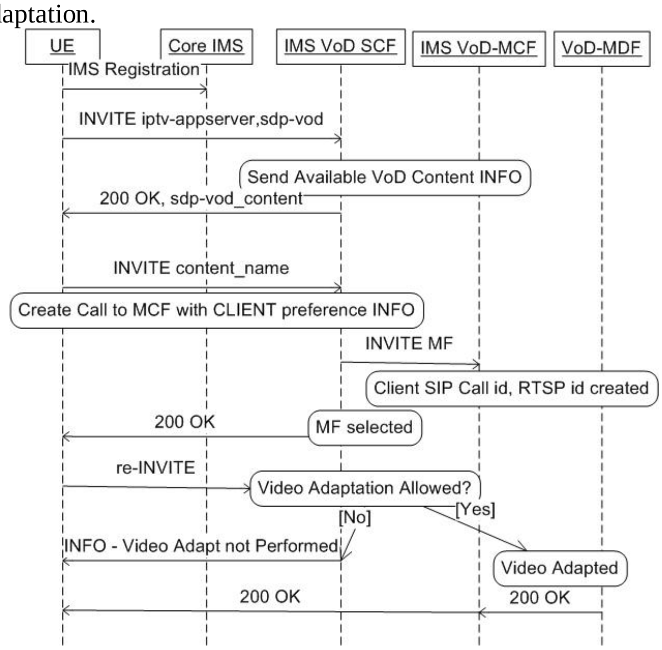

Fig. 6: Video Adaptation Example

After the User Equipment (UE) registers with the core, a video on demand session is initiated by sending an invite to the IPTV application server (in this case the SCF). The

SCF then sends the latest VoD content to update UE's Electronic Program Guide (EPG).

The client navigates through the available VoD content and then chooses a particular content of interest by sending an INVITE requests to the SCF to set-up a VoD session.

The SCF searches for the appropriate MCF and initiates the media session. The selected MCF collects the necessary client details such as the user interactivity profile information from the SIP call. A VoD session is then initiated and an OK message is send to the UE.

To perform video adaptation, the client sends an SIP re-INVITE request to the SCF. The request is examined at the SCF. The SCF uses the user profile architecture to check if video adaptation is allowed. If video adaptation is allowed, the client request is forwarded to the MCF for execution. If video adaptation is not allowed, a SIP INFO message indicating that video adaptation is not allowed is sent to the UE.

B. Advantages of the Proposed System

1) Unified Video Interactivity Framework

By linking session parameters to interactivity requests (by means of SIP call info and RTSP session states), the interactivity framework is logically coherent within the SIP-based IMS IPTV service requests. This presents a unified framework to both the user and the service provider when provisioning interactivity functions in the IMS context.

2) Interactivity Processing

The AIP adopts the use of the current TISPAN standard and the RTSP protocol. Due to its simplicity, the RTSP protocol is very efficient for processing interactivity modes.

3) Better Controlled VoD platform

Stream processing is managed by the AIP-MCF by use of the advanced profile architecture. This ensures better video interactivity management.

4) Better Interactivity (new modes)

The proposed system comprises of interactivity functions such as fast-forward and scene selection that are not currently incorporated in most VoD systems and in the current TISPAN standard.

VI. CONCLUSIONS

Current Internet based video streaming systems do not adequately address video interactivity requirements. Live streaming offers the most promising platform for advanced video interactivity. Due to its high Quality of Service (QoS) guarantees, The IMS-based framework provides an adequate platform for employing the live streaming technique. The current IMS standard lacks a comprehensive framework for video interactivity. The proposed advanced interactivity framework offers advanced interactivity functions which are coupled with better managed content selection and video adaptation procedures. Contributions in this paper are 1) a discussion on current video access techniques, 2) advanced video interactivity framework, and 3) an IMS based Advanced Interactivity Platform (AIP) for the implementation and management of the video interactivity framework. Future work will include the full implementation of the AIP, SCF and MCF entities. The platform will also be tested over different network interfaces and contexts.

REFERENCES

[1] Digital Video Broadcasting (DVB); Multimedia Home Platform (MHP) Specification 1.0.3, ETSI TS 101 812 V1.3.1 (2003-06)

[2] Dong Hee Shin, "Potential user factors driving adoption of IPTV, What are customers expecting from IPTV?", Elsevier-Technological Forecasting & Social Change 74 (2007), pp 1446–1464

[3] TISPAN Technical Specification, IMS-based IPTV stage 3 specification, ETSI TS 183 063 V2.1.0, June 2008

[4] International Telecommunication Union - Telecommunication standardization Sector IPTV Focus Group IPTV (WG1) C-0608. FG IPTV-C-0610, 4th FG IPTV meeting: Bled, Slovenia, 7 – 11 May 2007.

[5] Open IPTV Forum – Functional Architecture – V 1.1, APPROVED Jan 15, 2008.

[6] Rosenberg *et al* "SIP: Session Initiation Protocol" *RFC 3261*, June 2002

[7] H. Schulzrinne *et al* "RTSP: Real Time Streaming Protocol" *RFC 2326*, April 1998

Lesang V. Dikgole received his Electrical and Co mputer Engineering Bachelor of Science (*honours*) degree in 2007 from the University of Cape Town (UCT). He is presently studying towards his Master of Science degree in Electrical Engineering at UCT. His research interests include NGN networks, IPTV, Video on Demand and Digital Rights Management System