Voxpilot presents:

Interactive Services for 3G and IMS Networks: The Next Wave

Voxpilot Whitepaper, May 2006
By Dave Burke, CTO

Abstract

New capabilities in 3G mobile handsets and networks are fuelling a resurgence in video telephony. This enables the delivery of a new world of exciting mobile video applications and, for the operator and service provider, generates increased ARPU. This whitepaper explores the background to this resurgence, the driving technologies, and discusses Voxpilot's pioneering solution for making Interactive Voice and Video Response (IVVR) applications readily available on both 3GPP IP Multimedia Subsystem (IMS) and pre-IMS network architectures.

1. Introduction

By the end of 2006, third generation (3G) handsets will achieve double-digit penetration in Europe and, by the end of 2010, three in five mobile users will have adopted 3G[1]. In Japan, NTT DoCoMo have signed up more than 23 million subscribers to its 3G FOMA service and expect to transfer all of its mobile phones to 3G by 2010.

With 3G come new technologies for real-time streaming and conversational video that has naturally led to a renewed interest in video telephony. Concurrent with this, new video telephony features have enabled Interactive Voice and Video Response (IVVR) applications to be conveniently delivered to a wide audience. The real-time nature of a conversational, multimedia interface is ideal for delivering interactive, network-based applications and services such as:

- Video content streaming
- Video contact centers
- Infotainment portals
- Video blogging
- Video mail
- DVB-H Back Channel

For example, a video contact center solution may use an IVVR service for interacting with a caller before routing them to an agent. While on hold, video advertising may be delivered. Once the agent answers, he or she may be able to assist the caller further by streaming video clips down to the caller's phone while coaching them through the content in real-time.

For the user, classical IVR is evolving to a richer, more intuitive, and more effective medium. For the operator and service provider, IVVR means "monetizable multimodality" - a robust, multimodal interface by which to reach subscribers. Furthermore, since IVVR involves a phone call, it leverages existing revenue collection paradigms such as premium rate numbers and short codes and hence eases the realization of increased ARPU.

In tandem to the developments in the mobile

[1] Source: Frorrester Research, 2006

video application space, the Third Generation Partnership Project (3GPP) is forging ahead with developing the IP Multimedia Subsystem (IMS) network architecture. IMS is the 3GPP's vision for a converged IP network architecture that merges cellular and Internet technologies to uniformly deliver voice, video, and data. IMS addresses issues of quality of service, billing, mobility and rapid service delivery. It is equally applicable across fixed and mobile networks and hence is the forerunner candidate for providing true Fixed Mobile Convergence (FMC). Given the impending evolution of wireless and wireline networks, it is of critical importance that enduring IVVR solutions support both the pre-IMS networks of today as well as next-generation IMS networks.

There are several key technologies enabling the delivery of IVVR.. In the next section, we review these technologies and, in Section 3, we take a closer look at the unique solutions available from Voxpilot for delivering IVVR.

2. Technologies

2.1 Pre-IMS: 3G-324M Video

Today's 3G video telephony employs 3G-324M technology originally developed by the ITU under the umbrella standard H.324M and adopted by the 3GPP for Release 99 networks. 3G-324M requires a 64 kbit/s circuit-switch, bi-directional data channel right up to the mobile handset. Since the modern PSTN supports such data channels, it is possible to place 3G-324M calls from mobile to mobile that also traverse the PSTN. 3G-324M relies on existing signaling mechanisms to establish the data channel such as ISUP and ISDN.

There are three principal components of 3G-324M (see Figure 2.1):

- A H.223 multiplexor
- A H.245 control channel
- Audio and video coders

The job of the H.223 multiplexor is to multiplex audio, video, and control data down the 64 kbit/s constant bit-rate channel. The role ofH.245 is to provide control functions such asterminal capability exchange and media channel management. The audio and video codecs compress the media suitable for transmission over low bit-rates, while at the same time, maintaining sufficient fidelity. Today's 3G networks employ the H.263 and MPEG-4 codecs for video with compression ratios in the region of 40-60:1 and the AMR and G.723.1 codecs for audio with compression ratios in the region of 5-10:1. In Europe, all terminals are required to support H.263 and AMR at a minimum, thereby providing baseline interoperability.

H.245	Video	Audio
H.223 MUX		
64 kbit/s circuit		

Figure 2.1: Components of a 3G-324M client

In summary, 3G-324M as a technology presents several benefits:

- Guaranteed quality of service
- Easy to use (no specialized client required, short codes can be used)
- Standard revenue collection approach (premium rate numbers, etc)
- Simple to deploy

As a result of these benefits, 3G-324M is the preferred approach for delivering IVVR over 3G terminals today.

2.2 IMS Architecture

IMS is a standardized architecture that employs voice- and video-over-ip technology based on a 3GPP profile of SIP, and runs over the standard packet-based IP network. Figure 2.2 presents a simplified view of the IMS architecture where we have divided the network into three parts: the service layer, the control layer, and the access network.

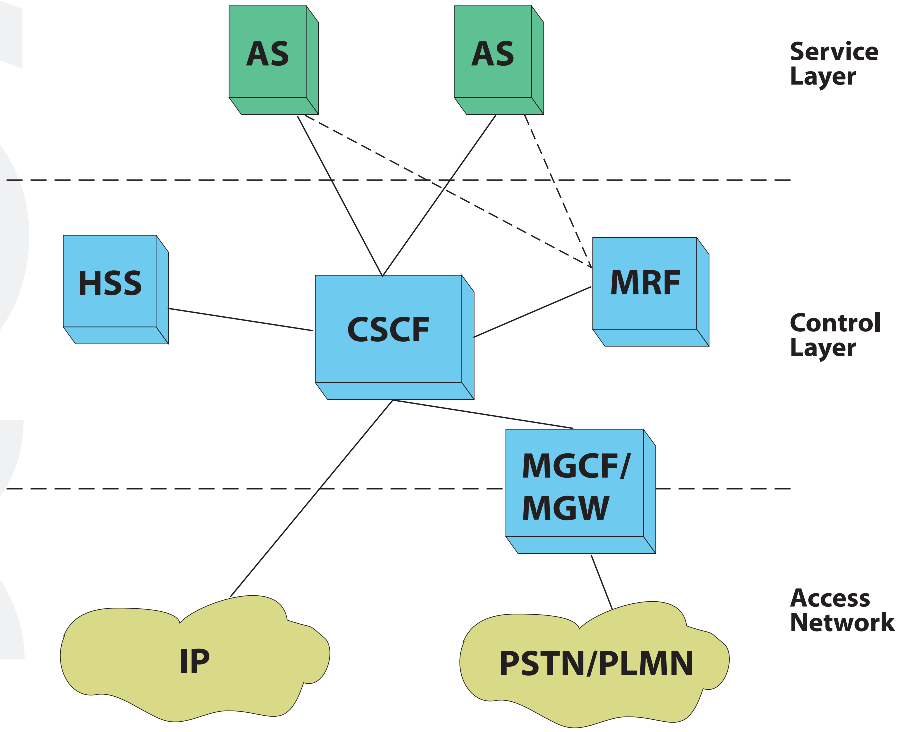

Figure 2.2: Simplified view of the IMS Architecture

IMS applications are hosted in the service layer. This layer consists of SIP Application Servers (AS) which execute IMS applications and services by manipulating SIP signaling and interfacing with other systems. The AS may also include HTTP capabilities allowing it to also perform the role of a content server for resources such as media files and VoiceXML application scripts. Typically, the AS will offer a programming language and framework for creating new services, for example Java SIP and HTTP Servlets.

The control layer of the IMS network consists of nodes for managing call establishment, management, and release. At the heart of the control layer is a specialized SIP server called the Call Session Control Function (CSCF) - all SIP signaling traverses this essential node. The CSCF inspects each SIP message and determines if the signaling should visit one or more application servers en route to its final destination. The CSCF interacts with the Home Subscriber Service (HSS), which provides a central repository of user-related information (an evolution of the Home Location Register). When the CSCF or AS requires media capabilities, it routes the signaling to the Media Resource Function (MRF). which provides centralized media processing capabilities (described in the next section). Finally, the Media Gateway (MGW) together with its controlling node, the Media Gateway Control Function (MGCF), interfaces with circuit switch networks.

The access network is comprised of IP routers and legacy PSTN switches that provide access to the IMS network both from contemporary IP telephony devices and older circuit switch devices respectively. IP devices compatible with IMS incorporate a SIP user agent which is used to place voice or video calls toward the network - this user agent naturally supports IVVR.

2.2.1 Media Processing

As telecom network architectures have evolved, there has been a strong tendency toward centralizing the media processing capabilities into a specialized logical node. Generically known as the "Media Server", the IMS network architecture equivalently labels this entity the "Media Resource Function" (MRF) and further decomposes it into two sub-nodes called the Media Resource Function Controller (MRFC) and Media Resource Function Processor (MRFP). The MRFC exposes a SIP interface to the other components of the network and provides signaling and control functions, while the MRFP provides media processing capabilities.

The MRF provides media processing services such as media stream origination (e.g. announcements, speech synthesis), media stream processing (e.g. DTMF/speech recognition, audio/video recording, transcoding), and media mixing (audio/video conferencing). The MRF is typically invoked by applications running on the AS via the CSCF. The interface exposed by the MRFC component of the MRF to the network is the so-called Mr interface and is based on SIP. Media terminates on and originates from the MRFP via the Mb interface which principally employs the Real-Time Protocol (RTP) for transporting media over IP networks.

In modern deployments, large-scale speech recognition and synthesis services are provided by specialized speech servers controlled by an IETF protocol called the Media Resource Control Protocol (MRCP). Typically, the MRFC will incorporate an MRCP client to control the distributed speech servers and the MRFP will supply media to and receive media from the speech servers.

2.3 VoiceXML

Offering advanced multimedia interfaces requires complex orchestration of media processing features. VoiceXML provides a powerful and convenient language for programming both voice and video dialog interfaces for the caller that abstracts the developer from low-level complexities and enables applications to be brought to market quickly.

VoiceXML is a markup-based, declarative programming language standardized by the W3C for creating speech-based telephony applications. VoiceXML supports dialogs that feature synthesized speech, digitized audio, recognition of spoken and DTMF key input, recording of audio, basic telephony call control, and mixed initiative conversations. In particular, VoiceXML enables Web-based development and content delivery paradigms to be used with interactive voice response applications.

More recently, there has been a move toward using VoiceXML for IVVR dialogs. Since VoiceXML is independent of the media it operates on, adding video media requires no changes to the language itself. Rather, by simply substituting a multimedia file container in place of an audio file container, and assuming the underlying media processing functions support it, VoiceXML can be used to create rich IVVR dialogs. A popular multimedia file format is the 3GP container which supports the video and audio codecs used on 3G networks.

With a VoiceXML IVVR application, one has the choice of locating the media content on a HTTP content server or on an RTSP streaming server. The HTTP approach brings with it advanced caching features and is ideal for serving pre-recorded media. The RTSP approach, which results in a real-time media stream from the RTSP streaming server, is more expensive to deploy and is best suited for streaming live video feeds such as from a professional video capture solution or webcam.

Today, VoiceXML is sufficiently powerful to create compelling IVVR applications. Future versions of the VoiceXML language will evolve to include new features for enabling even richer applications - designs for which are currently underway at the W3C.

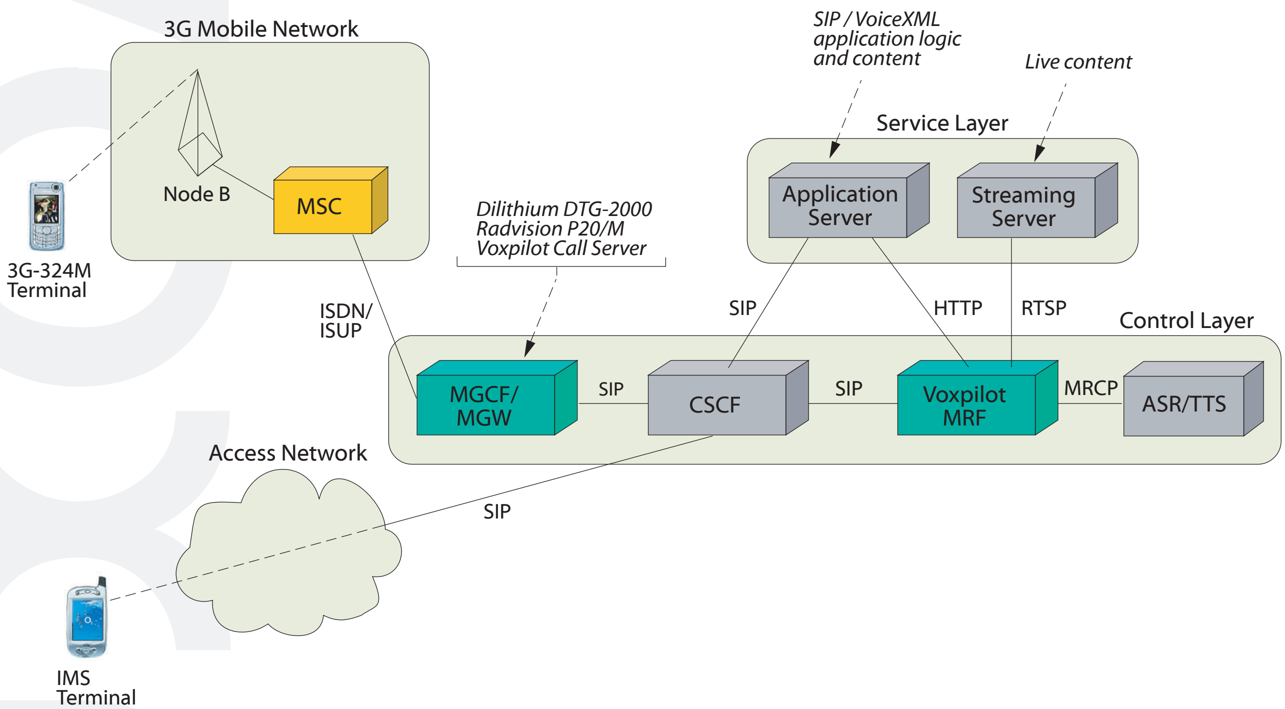

Figure 3: Voxpilot IVVR solution for IMS and pre-IMS networks

3. The Voxpilot Solution

The Voxpilot MRF, a component of the Voxpilot Open Media Platform (OMP), is the key enabler for delivering IVVR services on both IMS and pre-IMS networks - see Figure 3.

Within the Voxpilot MRF, the MRFC provides:

- a SIP interface to the network,
- a VoiceXML interpreter for processing VoiceXML application instructions
- an MRCP client for controlling distributed speech servers
- an RTSP client for interoperating with leading RTSP streaming servers such as Helix Mobile Server and Darwin Streaming Server.

The Voxpilot MRFP provides host-based media processing capabilities including:

- audio and video playback
- audio and video recording
- DTMF recognition services
- media forking
- transcoding.

The Voxpilot MRF exposes a standard SIP interface[2] which allows VoiceXML applications to be triggered from the AS or CSCF. The AS specifies the initial VoiceXML script URI within SIP and the MRF downloads the script via HTTP and commences processing it. Data may be returned to the AS either over HTTP or SIP. VoiceXML is an ideal language for IMS deployments because it employs the Web model allowing applications to be hosted on content servers within the IMS service layer. Moreover, locating the VoiceXML interpretation functions inside the MRFC focuses the media control burden away from the AS and CSCF nodes and close to the media processing functions located in the MRFP and MRCP-controlled speech servers.

The Voxpilot MRF runs equally well under either the Red Hat Enterprise Linux or Windows Server 2003 operating systems on standard Intel hardware. The software-only Voxpilot MRF is ideal for deployment on Advanced Telecom Computing Architecture (ATCA) blades for carrier-grade availability.

[2] Voxpilot implements the IETF Internet-Draft "SIP Interface to VoiceXML Media Services" (draft-burke-vxml), which is based on RFC 4240.

Voxpilot relies on 3G-324M media gateways to interoperate with pre-IMS, circuit-switched networks. The Voxpilot MRF is fully compatible with general-purpose, leading multimedia gateways such as those provided by Dilithium and Radvision. As part of the OMP solution, Voxpilot also offers its own 3G-324M gateway functions packaged within the Voxpilot Call Server - a combined multimedia gateway and CCXML soft-switch - ideal for IVVR service node deployments in smaller enterprise networks.

About Voxpilot

Voxpilot is a provider of IMS-Ready, VoiceXML and Video-In-VoiceXML technology that brings the next generation of interactive telecommunication services to enterprises, integrators, carriers and service providers. Our flagship product, Voxpilot Open Media Platform (OMP), has been installed in a variety of networks across Europe, America, Africa and Asia. Combining advanced telephony features with Internet capabilities and Web technologies, the OMP replaces classic Interactive Voice Response and enables speech-based self-service, conferencing, multimedia services, and Interactive Voice and Video Response (IVVR) solutions such as VideoMail and content delivery applications. Voxpilot's customers include France Telecom, Wind, Cegetel, Monaco Telecom, BT, NMS Communications, Swisscom, Ericsson, Cofiroute, Skype, PostFinance (Swiss Post Telephone Bank), and Universal Music Mobile.

For more information on our products and services, please contact us at the numbers below or sales@voxpilot.com and visit our web site at http://www.voxpilot.com

voxpilot

France:
4, Avenue des Martyrs de Soweto,
91300 Massy
France.
Tel: +33 1 69 75 78 00
Fax: +33 1 69 75 78 01

Ireland:
6-9 Trinity Street,
Dublin 2,
Ireland
Tel: +353 1 209 1919
Fax: +353 1 209 1920

USA:
8 Cog Hill Drive
Simpsonville, SC 29681
USA
Tel: +1 864 627 1442
Fax: +1 864 627 1222

Germany, Austria, Switzerland, Italy:
c/o Plurinnova,
Wetlistrasse 2,
8032 Zürich,
Switzerland.
Tel/Fax: +41 43 534 91 63