Voice Portal Services in Packet Network and VoIP Environment

Wu Chou and Feng Liu
Avaya Labs Research
233 Mt. Airy Road, Basking Ridge, New Jersey, NJ 07920, USA
{wuchou, fliu1}@avaya.com

Abstract
In this paper, we study the voice portal services in packet network and VoIP environment. An extensible VoIPTeleServer for VoIP in SIP (Session Initiation Protocol) environment is described. It is based on the concept of dialogue system and web convergence that separates the channel dependent media resources from the application creation environment. It supports XML based service applications for multiple channels including voice, DTMF, IM and chat over IP. Special attention is given to the adverse effect of delay, jitter and packet loss for voice portal services over IP. In particular, case studies of DTMF service in voice portal under adverse channel conditions are performed. The compounding effects of multiple channel impairments to DTMF in voice portal services over IP are revealed. The potential high error rate indicates that the data redundancy method as proposed in RFC 2198 is needed for DTMF in order to achieve reliable voice portal services over IP.

1. Introduction
The explosive growth of the IP network and the need to support multimedia/multimodal converged communication services have greatly accelerated the acceptance of VoIP as an alternative to PSTN (Public Switched Telephony Network). Session Initiation Protocol (SIP) is an emerging standard in VoIP. It is an application-layer control protocol that can establish, modify, and terminate multimedia sessions (conferences), such as Internet telephony calls [1]. It is regarded as the successor to H.323 for IP-based telephony. The approach of using Session Description Protocol (SDP) [2] in SIP for session information exchange offers great flexibility for setting up multimodal/multimedia applications. The session in SIP can consist of many forms of communications, including voice, DTMF, IM (instant messaging), and even video, if the user agent of two communication parties support. This makes SIP a favorable communication environment for converged communication services.

One scenario in enterprise service interaction can start from voice dialogue, where speech recognition (ASR), text-to-speech (TTS), and natural language understanding are applied to initiate the engagement. Visual browsers, such as HTML, WAP or others, that provide rich text and graphics, can be brought in as the interaction develops. To support the enterprise communication services in the native VoIP SIP environment, it requires a voice portal service platform be part of a “multimedia technology integration platform” that uses voice communication to initiate and integrate multimedia or multimodal communication. Such a platform will manage the state of visual and voice interaction, allowing devices to connect and transit smoothly between different modalities, and maintain the application context and service integrity [6].

Unlike other protocols, SIP is a text-based protocol, which is much like HTTP and with extensibility for multimedia communication. The relation of SIP with HTTP and SMTP allows it to reuse many Internet elements and bring voice application under the Internet umbrella. Besides its simplicity, SIP provides a mechanism to control and (together with media transport protocols, such as RTP) transmit voice/video over the same IP network that carries data. SIP’s integrated capabilities of presence, availability, location, mobility, and IM can greatly extend the reach and power of enterprise multimodal, multimedia communication services.

The adoption of VoIP in converged communication presents many new technical challenges for voice portal services. The voice portal service over IP departs from the traditional PSTN in that all media data (voice, DTMF, video, etc.) are transmitted over IP as data packets embedded as payload in the RTP data stream. A voice portal service platform over IP needs to manage and control the media transmission/receiving, under the adverse channel condition of packet loss, jitter, delay, etc. The data from multiple data sources need to be multiplexed and demultiplexed in order to transmit and reconstruct the original media streams and transmitted control sequences for further processing. It is an active research area to understand the impact of the channel condition on the communication service over IP, and develop methods that alleviate the service impairments for reliable services.

In this paper, we study the voice portal service platform over IP as part of a Multimedia Technology Integration Platform (MTIP), and address the issues of supporting voice communication in the native SIP environment over IP. DTMF service is an important part of voice portal service, which allows the user to use DTMF to communicate with the voice portal for mission critical applications (e.g., password, service entry selection, etc.). It is critical for voice portal service deployment in noisy environment or for applications that require privacy is a major concern. The DTMF service over IP is typically based on the IETF RFC 2833 specification. It is achieved by injecting DTMF data service packets into the same data stream that carries the packets of all other medias. As a result, the same packet loss, jitter, and delay during the transmission, and it can behave very differently in the situation of voice portal services over IP.

The organization of this paper is as follows. We survey the related work in Section 2. The architecture of VoIPTeleServer over SIP for voice portal services that support voice/DTMF/IM over IP is described Section 3. In Section 4,
we study voice and DTMF services over IP in more detail, and we summarize our findings in Section 5.

2. Related Work

Integrated voice portal services in SIP environment is an active research area. In [9], an approach is described to integrate the VoiceXML based voice dialogue services with SIP. A SIP user agent is directly connected to a VoiceXML browser, making interactive voice response applications accessible by a SIP phone. The VoiceXML browser controls the interaction and fetches a VoiceXML page from a web server upon receiving a SIP INVITE. This page is passed on to a VoiceXML interpreter that is connected to text-to-speech (TTS) to synthesize a system prompt. The TTS output is sent out to the caller via RTP. The DTMF input is transmitted back through RTP to VoiceXML browser that invokes a DTMF detector to handle the DTMF input. However, from voice service portal perspective, it is rather primitive. The realized communication modes were TTS, DTMF and voice recording, which are quite limited even for the voice modality. The platform was directly wired to the input channel, and lacks an extensible service integration infrastructure.

Voice service quality over IP is an area that received a lot of attention. In [4], it is studied the audio packet loss over IP for speech recognition. The packet loss was modeled by the Gilbert model, which consists of two states. One of the states (state 1) represents a packet loss, and the other state (state 0) represents the case where packet is correctly transmitted. This model is characterized by two parameters: $p$, the probability from state 0 to state 1, and $q$ from state 1 to state 0. Speech recognition performance degradation for different speech coders (e.g. G.723.1 and G.711) under packet loss was studied. However, this study does not consider the transmission condition, where packet loss, delay and jitter occur at the same time.

Moreover, DTMF is typically regarded as reliable in PSTN, and much is to be learned in VoIP environment for voice portal services. The issue becomes acute when considering the adverse channel condition of packet loss, delay and jitter during the transmission over IP. In addition, the DTMF transmission over IP is based on IETF RFC 2833 [3]. The reliability of this method depends on the way how the DTMF is entered (e.g. the minimum duration of the key pad touch, etc.), and how the media stream is unpacked at the voice portal site to recover the transmitted control sequence from the payloads of multiple media sources in the RTP stream.

Our study in this paper provides an inside look into these critical issues for creating voice portal services over IP under the native SIP environment. We describe a generic portal service platform architecture, which is extensible and can support integrated portal services (e.g. voice, DTMF, IM, chat, etc.) over IP. We study in detail the DTMF services, implemented according to IETF RFC 2833, under the adverse channel condition of having packet loss, jitter and delay in transmitting payloads for voice portal services.

3. An Extensible Voice Portal Architecture for Converged Services over IP

In our approach to converged communication, the VoIP portal service platform is a component of the MTIP. Fig. 1 depicts the MTIP in a converged network environment. It shows that the MTIP in our approach can be accessed from PSTN phones, mobile phones and SIP phones through different interfaces. For MTIP, VoIPTeleServer is the interface of MTIP to the VoIP SIP environment. It is responsible for information exchange between MTIP and SIP networks. MTIP in our approach has a Resource Manager that coordinates all the resources, including PSTNTeleServer, ASR, TTS, etc. From the MTIP Resource Manager perspective, VoIPTeleServer is one of the portal service resources it manages. The Resource Manager is the only component on MTIP that VoIPTeleServer directly interacts with. The communication between VoIPTeleServer and MTIP is through TCP/IP. From SIP network point of view, VoIPTeleServer is exposed as a SIP endpoint that can be reached by other SIP clients. The difference is that VoIPTeleServer is a SIP server that can communicate with multiple SIP clients at the same time and manages the session and media transmission for each application simultaneously. Since MTIP connects to other SIP resources, including Registrar, IM server, and Presence server, etc., it can easily leverage these new SIP service capabilities in voice portal service over IP by integrating these resources with VoIPTeleServer.

3.1 VoIPTeleServer Structure

Based on the functionalities, our VoIPTeleServer consists of SIP User Agent, RTP handler and TeleManager as shown in Fig. 2. The SIP interface is a SIP User Agent built on a SIP stack, handling all SIP related signaling, such as session setup, session termination, session modification, call transferring, call cancellation, etc. The RTP handler consists of media transmitter, media receiver and DTMF detector. The media
transmitter is to package media data received from MTIP and send to SIP clients over RTP. The media receiver is responsible for receiving RTP packets, extracting media data and sending it to MTIP according to commands issued by MTIP. The DTMF detector handles DTMF input from the RTP stream. The TeleManager is the interface of VoIPTeleServer to MTIP which manages the mapping between calls and channels, interprets the commands from MTIP, and sends events to MTIP according to call status and media handlers.

The asynchronous communication is realized by adopting an event notification mechanism. Basically VoIPTeleServer sends different events to MTIP to notify current call status and media status, and MTIP issues commands to direct the VoIPTeleServer what to do next. It is VoIPTeleServer’s responsibility to notify MTIP of execution result. MTIP determines the next action according to application logic and notification from VoIPTeleServer.

The media format is controlled by MTIP. MTIP directs the VoIPTeleServer as to what kind of media formats it accepts and in what kinds of format it sends media data. VoIPTeleServer has to conduct format conversion, if different formats are used in RTP transmission. The communication between MTIP and TeleManager is through TCP/IP connection. This makes our platform easily deployable in a distributed environment. The separation of the channel interfaces, resources, and interaction into a three-tier infrastructure in our approach has the advantages of sharing the processing resources over different channels, easy integration cross multiple modalities, and having a markup language layer for using XML in multimodal services (e.g. VoiceXML, Hybrid XML, EMMA, etc.) [6-8].

### 4. Voice Portal Services over IP

Dual tone multifrequency (DTMF) service is an important service for a voice service portal. In VoIP systems, DTMF tones can be transported from one terminal to the other in basically three ways [9]. One approach is, at the sender side, to generate and encode the tone as voice, and send the tone as audio RTP packets. At the receiving end, a touch-tone detection algorithm is deployed to detect the DTMF tone, as does in the PSTN telephony network. However, it cannot be guaranteed that the low-rate voice codecs can reproduce these tone signals accurately enough. In addition, the network conditions such as packet loss, delay and jitter, may significantly affect the quality of the in-hand DTMF tones. Both factors make it hard for the receiver to detect the tone.

The second method, according to IETF RFC 2833, is to translate the tone as a telephony event and transmit it in RTP packets [3]. The third method is to transport the DTMF tone in the SIP INFO message. As the SIP messages can be transported by TCP, the network condition has little effect on this method. However, it requires some architecture change of the SIP terminal and it is against the principle that the user data should not be transmitted through the signal channel in SIP [10]. Therefore, method two based on IETF RFC 2833 is the most supported method for SIP based voice portal services over IP.

In RFC 2833, a DTMF tone is represented as a telephone event, and the payload format of the RTP packet carrying this event is specified. According to RFC 2833, an audio source starts transmitting DTMF event packets as soon as it recognizes such an event, and, during the life span of the event, it transmits it every 50 ms thereafter or every packet interval for the audio codec used for this session, if it is known. The end of the event (one DTMF tone) is transmitted three times. This ensures that the duration of the event can be recognized correctly even if the last packet for an event is lost.

#### 4.1 Case Study - DTMF Service over IP under Adverse Channel Conditions

We studied the service impairment of DTMF services over IP under the adverse channel conditions of packet loss, jitter and delay. To understand the effects of the network conditions on the DTMF services, we did not use the redundancy transmission scheme. The DTMF signal was encoded according to IETF RFC 2833, and transmitted over IP during the session of life voice call using SIP/RTP.

#### 4.2 Experimental setup

Two SIP terminals, AVSP3 and AVSP4, were connected through a router. We installed the NIST Net [11] on the router to emulate the network QoS (delay, loss, and jitter) impairments for transmission over IP. NIST Net is a general-purpose tool for emulating performance dynamics in IP networks. The tool is designed to allow controlled, reproducible experiments with network performance sensitive/adaptive applications and control protocols in a simple laboratory setting.

#### 4.3 Packet Loss vs. DTMF Detection Errors

During the experiments, AVSP3 sent DTMF tones periodically to AVSP4. The RTP packets are sent every 20ms. In our DTMF detection method, if several consecutive DTMF packets were lost, RTP receiver would replace them with noise packets, and determine the DTMF event was finished, even though it was not. Thus when the RTP packet of the same event arrived later, it was detected as a new DTMF event. In this case, an insertion error was generated. In most cases, the DTMF tone duration was in the range from 80–500ms, or 6–27 RTP packets (including 2 more RTP packets indicating the end of the event). The insertion error as a function of the

![Figure 2: VoIPTeleServer architecture](image-url)
packet loss is tabulated in Table 1, which shows how insertion errors in DTMF services increased with packet loss.

**Table 1 Insertion Error vs. Packet Loss**

<table>
<thead>
<tr>
<th>Packet Loss(%)</th>
<th>0</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insertion Error(%)</td>
<td>0</td>
<td>0.25</td>
<td>0.65</td>
<td>1.45</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

During the experiments, the duration of the tone is 150 ms, and the interval between two tones is 60 ms.

### 4.4 Tone duration vs. DTMF service impairment

The relation between the tone duration and DTMF service impairments over IP was studied. Results were listed in Table 2. It can be seen from Table 2 that the longer the tone lasts, the more likely an insertion error can occur. The packet loss in this experiment was set to 20%.

**Table 2. Insertion Error vs. Tone Duration**

<table>
<thead>
<tr>
<th>Tone Duration(ms)</th>
<th>100</th>
<th>150</th>
<th>200</th>
<th>300</th>
<th>400</th>
<th>500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insertion Error(%)</td>
<td>0.50</td>
<td>1.45</td>
<td>2.35</td>
<td>3.75</td>
<td>5.85</td>
<td>9.8</td>
</tr>
</tbody>
</table>

### 4.5 Packet delay vs. DTMF service impairment

From the method that DTMF is transmitted in the RTP stream, a short but constant delay would not have a significant impact on DTMF services. However, if the delay is not a constant, i.e., in case of a jitter, the order of the packets arrived at the receiver may be out of sequence, and errors may occur. To study the impact of jitter to DTMF services, the distribution of the packet delay in our experiments were set to be a normal distribution characterized by its mean and variance. Table 3 tabulates the DTMF service impairments in relation to the increase of the random delay variance at mean of 100ms.

**Table 3. Insertion Error vs. Packet Delay and Jitter**

<table>
<thead>
<tr>
<th>Mean(ms)</th>
<th>100</th>
<th>100</th>
<th>100</th>
<th>100</th>
<th>100</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sigma</td>
<td>0</td>
<td>5</td>
<td>10</td>
<td>20</td>
<td>30</td>
<td>40</td>
</tr>
<tr>
<td>Insertion Error(%)</td>
<td>0</td>
<td>0.4</td>
<td>2.2</td>
<td>8.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 4.6 Joint effects of delay, jitter and packet loss to DTMF service impairment

We studied a more realistic situation of adverse channel condition, where packet loss, delay, and jitter all happened during the transmission. In this study, the DTMF tone duration is 150ms, with 60ms interval. The mean value of the packet delay is 100ms. The relation between the DTMF detection error rate, packet loss, and variance of the random delay are presented in a 3-D plot in Fig. 5. The color bands in Fig. 5 correspond to different values of the variances in the jitter distribution which was modeled by a normal distribution during the experiments. Results indicate that the compounding effects of the adverse channel condition of having packet loss, delay, and jitter at the same time can have a much significant impact on the DTMF detection error rate for voice portal over IP.

### 5. Summary

In this paper, we studied the voice portal services in packet network and VoIP environment. An extensible VoIP Teleserver was described. In particular, case studies of DTMF service in voice portal under adverse channel conditions are performed. The compounding effects of multiple channel impairments to DTMF in voice portal services over IP were revealed. The potential high error rate of DTMF service indicates that the data redundancy method as proposed in RFC 2198 is needed for DTMF in order to achieve reliable voice portal services over IP.

### 6. References