

Understand VXML / CVP HTTP Cache for Media Files

Contents

[Introduction](#)

[Prerequisites](#)

[Requirements](#)

[Components Used](#)

[Background Information](#)

[Gateway Prompt Caching Considerations](#)

[How to determine if the gateway is caching properly](#)

[Calculate the FreshTime](#)

[Removing Stale Cached Entries](#)

[Audio-Prompt Load Command](#)

[Date Time](#)

Introduction

This document describes the Voice Extensible Markup Language (VXML) / Customer Voice Portal (CVP) Hypertext Transfer Protocol (HTTP) cache for media files.

Prerequisites

Requirements

Cisco recommends that you have knowledge of these topics:

- VXML Gateway
- CVP

Components Used

This document is not restricted to specific software and hardware versions.

The information in this document was created from the devices in a specific lab environment. All of the devices used in this document started with a cleared (default) configuration. If your network is live, make sure that you understand the potential impact of any command.

Background Information

In HTTP Client Cache, there are two types of cache involved in storing media files: the IVR Media Player cache and the HTTP Client cache.

The server cache setting overwrites the HTTP client settings, these parameters are sent from the server either via http message headers, or through vxml application scripts.

Gateway Prompt Caching Considerations

Step 1. When audio prompts are stored on an HTTP media server, proper gateway prompt caching methods are necessary to optimize both the performance of the gateway and network bandwidth consumption. Gateway performance decreases by approximately 35-40% if caching is disabled entirely.

To configure caching on the gateway set the following on the gateway:

```
..ivr prompt memory 15000
```

```
..http client cache memory file 500
```

```
..http client cache memory pool 15000
```

Note: The **http client cache memory file** represents the largest size prompt file (in Kbytes) that can be cached. In general, customer prompts larger than 500K (about a minute in length) should be broken up into smaller, more manageable pieces to facilitate loading and caching. For example, queue music could be a repetitive loop of a 30 second prompt. Note also that because the prompts are streamed, the prompt will not be cached unless the whole prompt is played. Therefore, it is recommended that you make prompts a manageable size.

Step 2. Synchronize the datetime between the gateway and the HTTP media server.

Note: Synchronization does not have to be exact, but at least within a minute or two. Times that are not synchronized can cause prompts to never refresh or they will refresh with every call, both of which are undesirable behaviors.

Step 3. On the media server, set the content expiration (for example 15 minutes).

Note: In IIS, this is done under the HTTP Header tab. The gateway prompt will be refreshed after this time period. The period chosen should reflect how oftener r re-record prompts and how long you are willing to wait to have the new prompt load after modification.

Programs > Administrative Tools > IIS Manager

Navigate to the .wav file you want to modify.

Right click > **Properties > HTTP Headers**

Enable Content expiration.

How to determine if the gateway is caching properly

To determine if you have properly configured gateway caching, do the following:

The IIS log on the media server records every time a client requests a prompt. If caching is set up correctly, these requests appear approximately every X minutes (X is whatever was defined as the refresh interval in Step 3 above) for any particular prompt. The log is located at:

C:\WINNT\system32\LogFiles\W3SVC1\ex*

or

Do **show http client cache** on the gateway. The **Fresh Time** column must equal the refresh time period set on the HTTP media server. For example, if the refresh period was set to 15 minutes, this must say 900 seconds. The **Age** column shows how many seconds have passed since the prompt was last refreshed. In general, this number is less than the **Fresh Time**. However, if no call has ever accessed the prompt recently, this number can be greater than the fresh time. Prompts are only refreshed when triggered by a call and the prompt **Fresh Time** has expired. If the **Fresh Time** is a very high value, the only way to remove the prompt from cache (other than the hidden commands) is to reload the gateway.

It's much easier just to add the header as real HTTP Header via IIS.

This can be done via IIS 6 or 7.

<http://weblogs.asp.net/joelvarty/archive/2009/03/23/force-ie7-compatibility-mode-in-ie8-with-iis-settings.aspx>

Calculate the FreshTime

There are several variables that can affect the FreshTime of a file, such as: http message headers from the server, and cache refresh value configured via the CLI, etc. So how do we know which value a file uses for its FreshTime? The FreshTime of a file is determined in the following precedence:

1) When a file is downloaded from the http server, if one of the http message headers contains the following:

Cache-Control: max-age = <value in seconds>

Then the above <value in seconds> will be used as the FreshTime for this file.

2) If (1) is not present, but the following two headers are included in the http message:

Expires: <expiration date time>

Date: <Current date time>

Then the difference <expiration date time> - <Current date time> will be used as the FreshTime for this file.

3) The HTTP/1.1 spec, RFC 2616 (HyperText Transport Protocol), recommends that either of the http message headers as described in (1) or (2) above be present. If the server fails to send both (1) or (2) in its http response, then we will take 10% of the difference between Date and Last-Modified from the message headers:

Last-Modified: <last-modified date time>

As previously mentioned, a stale cached entry is removed by its owner on an as-needed basis, when:

- The cached entry becomes stale; and
- Its ref count is zero (0), i.e., no one is using this cached entry; and
- Its memory space is needed to make room for other entries

This means the http client and IVR Media Player must manage and control their cached entries in this way in non-streaming and streaming modes, respectively. What if the http client needs to clean up some stale entries to regain space in the cache memory pool but it is not the owner of those files? This becomes the responsibility of the http client cache background ager.

The http client cache background ager wakes up every 5 minutes. If the total memory used for the cached entries exceeds 70% threshold of the configured cache memory pool size, the ager will walk through every cached entry. If the entry is still fresh, it will leave it alone. If the entry is stale and has no reference to it, i.e., ref count = 0, the http client deletes the entry on its own because it is the legitimate owner of that entry. If the stale entry has a reference count 1 on it and it has no parent or child linked to it, meaning the file is not in the middle of refresh download, the http client calls back to notify the Media Player to release this stale entry.

Audio-Prompt Load Command

Sometimes, it can be desirable or necessary to manually download an audio file into the router. By now we are already told that the router does not automatically go to the http server to refresh stale cached entries. Those entries are refreshed only when they are needed. A manual download can overcome this issue.

Another scenario a manual download may be useful is to preload a large audio prompt in non-streaming mode. This can be done before the first call is received so that the caller does not experience any delay of prompt loading.

To manually download a particular audio file, type the following CLI command:

audio-prompt load <url>

The above <url> is where the audio file resides on the server. Of course, the http client cache is expected to be properly configured to save this file in the cache.

Note: If the <url> is an active prompt, i.e., currently in play, this CLI does not take effect.

Date Time

Also, please make sure the datetime between the gateway and the HTTP media server are synchronized. This is a must.

Warning: Do not use **clear http client cache** in the VXML GW. If this command is invoked on very loaded/active VXML GW, it is known to cause problems, memory corruption and crashes. Basically, use of **clear ip http client cache all** is not recommended. What it does is

it refreshes all the entries from the cache, and what happens is it creates and deletes nodes from the cache linked-list which causes some problems. The command is in process of being removed from IOS. The recommended command is **set http client cache stale**, what this command does is it just refreshes the newly changed part of the cache.