Dialogic
Making Innovation Thrive™

White Paper

Considerations for Creating Streamed Video Content over 3G-324M Mobile Networks

Considerations for Creating Streamed Video Content over 3G-324M Mobile Networks

White Paper

Executive Summary

To provide the best possible video quality to mobile users who are likely to be watching video on 3G-324M or other mobile multimedia streaming environments with restricted video rates, special care must be taken in creating video content. This white paper provides basic information about video characteristics and video streaming techniques, and discusses guidelines for authoring content. A short section describing Dialogic® products that support video streaming is included.

White Paper

Considerations for Creating Streamed Video Content over 3G-324M Mobile Networks

Table of Contents

Considerations for Creating Streamed Video Content over 3G-324M Mobile Networks

White Paper

Introduction

Creating multimedia content for successful transmission over a mobile network with restricted data rates, such as the 3G-324M network, requires knowledge and skill. It involves an understanding of the fundamentals that affect video quality, such as codec choice and compression, and the use of specialized tools, such as the Dialogic® hmp3gp utility, which can be used to verify that the contents of a video file are properly specified for a target audience.

Video Basics

Because the bandwidths of mobile networks are limited, the video and audio data must be compressed significantly. This section covers the basics of compressed video and its characteristics within video networks.

Video Bandwidth

In multimedia streaming applications, video compression is used because uncompressed video requires enormous amounts of data. In fact, standard definition (SD) movies on DVD or Blu-ray are already in a compressed format that provides a bandwidth of 4 - 6 Mbps. For mobile streaming networks, which can require data rates as low as 30 kbps, this means that the video must be compressed hundreds of times or more to achieve the required bandwidth.

With the growth of mobile multimedia streaming, it is important to work within the bandwidth limitations of the network and the capability of the endpoint. Although 3G and next generation networks provide much higher bandwidths to mobile devices, as more and more endpoints use these networks for multimedia tasks, conformance to target audience bitrates will become even more important than today.

Networks Used for Video Delivery

Table 1 describes the network environments used for delivering video services:

Network	Bandwidth available	Terminals	Codecs	Image size
3G-324M	64 kbps	Video handsets	H.263, MPEG-4, H.264	QCIF, CIF
3G wireless data	256-768 kbps	Video handsets, smart phones	H.263, H.264, MPEG-4	QCIF, CIF
Broadband IP	768 kbps	Smart phones, soft client on PC	H.264	QCIF, CIF
Enterprise	2-5 Mbps	Soft client	H.264	CIF, 4CIF, HD
WiMax, LTE	2-100 Mbps	PC, TV, portable devices	H.264	CIF, 4CIF, HD

Table 1. Networks Used for Delivering Video Services

Video Codecs

A video codec, short for enCOder/DECoder, is used to compress video data to achieve a very low bitrate. Examples of video codecs include MPEG-2 (used in DVD compression); and H.263, MPEG-4, and H.264 (found in low bandwidth mobile and IP video telecommunications networks). To achieve such low bitrate video, video codecs use both lossless and lossy compression techniques.

Lossless compression techniques allow the original source video to be reproduced from the compressed video. One type of lossless compression technique, called Huffman coding, is used in many data compression algorithms.

Lossy compression techniques do not provide this ability. With lossy compression, the less significant information digitally represented in the video is removed (lost). An advantage of lossy compression is that it can compress the video to the very low data rates required for wireless networks. Lossy compression first removes the data that cannot be perceived by human vision; however, as more information is removed from the source signal, the result can be a reduction of video quality.

Compression efficiency is the ability of a codec to encode more video characteristics into a data stream described by fewer bits. The more efficient a codec is at compression, the better the quality and sharpness of the video, because more can be described with less data. For example, H.264 has more efficient video compression than H.263 and MPEG-4, its predecessors. In general, H.264 provides better video quality at a lower bandwidth, but at the cost of encoder complexity and greater processing requirements for decoding at an endpoint.

For more information about video codecs, see *Selecting the Video Codec*, later in this white paper.

Video Profiles and Levels

Video codec standards have defined profiles and levels to describe the common characteristics of video within specific ranges. A video endpoint usually supports well-defined profiles and levels for the video, as this is a common parameter to use when designing the video capabilities of the endpoint. Supporting a higher profile or level usually implies support of the lower levels. The video format resolutions common for video telephony, 3G-324M, and mobile networks are Sub-QCIF (88 x 72 pixels), QCIF (176 x 144 pixels), and CIF (352 x 288 pixels). The profiles and levels also specify the maximum number of frames per second (frame rate) and bits per second (bitrate) that an endpoint needs to support to be standards compliant.

Tables 2 and 3 describe the video codecs commonly used for IP video telephony and 3G-324M functionality.

Codec	Profile	Level	Video Format	Frame Rate	Bitrate
H.263	Profile 0	Level 10	Sub-QCIF QCIF	Up to 15 fps	Up to 64 kbps
	Profile 0	Level 20	Sub-QCIF QCIF CIF	Up to 30 fps Up to 30 fps Up to 15 fps	Up to 128 kbps
	Profile 0	Level 30	Sub-QCIF QCIF CIF	Up to 30 fps	Up to 384 kbps
MPEG-4	Simple	0,1	Sub-QCIF QCIF	Up to 15 fps	Up to 64 kbps
	Simple	2	Sub-QCIF QCIF CIF	Up to 30 fps Up to 30 fps Up to 15 fps	Up to 128 kbps
	Simple	3	Sub-QCIF QCIF CIF	Up to 30 fps	Up to 384 kbps
H.264	Baseline	1,1b,1.2,1.3	Sub-QCIF QCIF CIF	Up to 30 fps	Up to 128 kbps Up to 192 kbps Up to 768 kbps

Table 2: Commonly Used IP Video Codecs

Codec	Profile	Level	Video Format	Frame Rate	Bitrate
H.263	Profile 0 Profile 3	Level 10	Sub-QCIF QCIF	Up to 15 fps	Up to 64 kbps
MPEG-4	Simple	0	Sub-QCIF QCIF	Up to 15 fps	Up to 64 kbps
H.264	Baseline	1	Sub-QCIF QCIF	Up to 15 fps	Up to 64 kbps

Table 3. Commonly Used 3G-324M Video Codecs

Video Streaming Basics

Compressed video that is sent across a data network is called a video stream. It is important to understand the basics of video compression and the effect that it has on the video stream at low bandwidths. This section covers the characteristics of the video stream and the constraints imposed by the network and endpoints that help define the target audience bitrate.

Video Compression Techniques

Video codecs use various compression techniques to fit a video signal into the allotted channel bandwidth. These compression techniques can affect the resulting quality of the video in different ways. An understanding of encoding principles can help a content provider determine what content will look best on a mobile device, and highlight some of the expected tradeoffs when producing multimedia files.

Quick bandwidth reduction can be achieved by using video compression techniques such as:

- Removing statistical redundancies
- Reducing resolution size (for example, CIF ➔ QCIF)
- Using fewer frames per second (for example, 15 fps ➔ 10 fps)

Further bandwidth reduction can be achieved by leveraging the patterns within the video data and removing redundancies. Image compression relies on discarding information that is indiscernible to the viewer. Motion compensation provides interpolation between frames, using less data to represent the change. The goal of a video encoder is to remove redundancies in the video stream and to encode as little data as possible. To achieve this goal, the encoder samples the video stream in two ways:

- In time intervals from consecutive frames (temporal domain)
- Between adjacent pixels in the same frame (spatial domain)

A video decoder pieces the video stream together by reversing the encoding process. The decoder reconstructs the video stream by adding together the pixel differences and frame differences to form a complete video.

This is an overly simplified look at compression, but it is useful to remember that a compressed video stream provides the deltas between previously encoded data, instead of a complete representation of each frame.

Compressed Video Stream

A compressed mobile video stream consists of a mix of two types of encoded frames:

- I-frames (also called key frames or intraframes)
- P-frames (also called predictive frames or interframes)

I-frames

An I-frame represents a complete picture snapshot in time. It is also called a key frame, because it represents the reference key of the video stream. All pixels that describe the picture are defined in the I-frame. A video decoder ***must*** start with an I-frame to decode the video stream because without an I-frame, a video decoder has no reference to determine how video pixels have changed since the previous frame. For this reason, compressed video recordings normally do not start until an I-frame is received by the video recording device.

P-frames

P-frames represent the change in the picture since the previous frame, which can either be an I-frame or a prior P-frame. The amount of data in a P-frame is many times smaller than the amount of data in an I-frame. If a video starts decoding on a P-frame at an endpoint, a user might see either scrambled video or no video, because there is no reference frame. Alternatively, a user could see a block by block creation of the changed macroblocks (units of 16 x 16 pixels) refreshed on the screen.

Considerations for Using I-frames and P-frames

Both I-frames and P-frames are important in compressed video streams. I-frames provide a reference for a decoder, as well as seek points for DVR controls; but a video stream with many I-frames will likely exceed bitstream limits. On the other hand, although using longer stretches of P-frames keeps the video stream compressed significantly, P-frames do not provide a good mechanism for recovery at the decoder if frames are missed, lost, or corrupted.

Video Streaming Container (.3gp)

When streaming multimedia files to mobile handsets, the audio and video data must be placed in the proper format for streaming. The container format for mobile multimedia streaming is the .3gp file, defined by the 3rd Generation Partnership Project (3GPP), http://www.3gpp.org, for delivery to mobile devices. Because the bandwidths of video telephony networks are limited, the video and audio data included in a .3gp file is compressed significantly.

Within the .3gp container, video can be encoded with specific video codecs specified by the 3GPP. Dialogic® products support .3gp files with content encoded with the H.263, MPEG-4, and H.264 video codecs. Dialogic® products also support the AMR-NB and G.723 audio codecs. AMR-NB is required per the specification and is much more popular than G.723, due to limited handset support for G.723. In order for Dialogic® products to stream a .3gp file's contents, the file must be specified with the video and audio codecs supported by Dialogic.

Video Streaming Constraints

Video streaming is constrained by the network channel capacity, 3G-324M channel bandwidth, and endpoint capabilities.

Network Channel Capacity

Network channel capacity refers to the amount of multimedia data that can be transported over a mobile network or channel from the server to the endpoint. The amount of data can fall within a large range when considering all types of mobile networks and radio channel capabilities. Mobile broadband networks are growing significantly, but there are many types of technologies between a server and the endpoint that can dictate the channel capacity. Dealing with network capacity is probably the biggest concern when working with 3G-324M networks.

3G-324M Channel Bandwidth

3G-324M networks support the transport of multimedia over a circuit switched connection. An advantage of a 3G-324M multimedia connection is that it supports a constant Quality of Service (QoS), because it reuses the PSTN circuit timeslot as a data channel to support multimedia streaming. A drawback is that the maximum amount of data that can be transported over a single PSTN timeslot is 64kbps, even when the timeslot is used only for data traffic.

To add to this limitation, the 64 kbps data channel is made up of 3G-324M data, which multiplexes video data, audio data, control logical channel data, and overhead data used for data link level framing.

The audio channel for 3G-324M is usually AMR-NB, which can vary from 4.7 kbps to 12.2 kbps, but is normally held at constant rate of 12.2 kbps to optimize audio quality. As a general rule, the video bandwidth allotted for the video channel is less than 42 kbps.

For the 3G-324M channel, constant bitrate (CBR) encoding is used over variable bitrate (VBR) encoding to avoid bitrate spikes. Any bandwidth utilization above 50 kbps, even for a very short duration, can cause data drops or data delay through the channel and can have dramatic effects on video quality. This means that the ceiling to the video channel bitrate must be significantly lower than 50 kbps to account for any spikes in bitrate. Large spikes in bitrate are more detrimental to video quality than adjustments in the overall average. Some encoding tools are better than others at conforming to a constant bitrate, especially for tough video sequences. It is usually best to keep the average video bitrate between 35 kbps and 42 kbps to account for the variations.

Equally important, the size of the frames must be scaled to the data rate of the file being played. A large frame could take a considerable amount of time to be sent on a low bandwidth network, and the video stream at the device can freeze while waiting for the entire frame to make it through the network. For example, a 5000 byte video frame takes approximately one second to arrive over a 40 kbps 3G-324M network, and the endpoint must wait until the entire frame arrives before it can be rendered. Consequently, for 3G-324M networks, typical I-frame sizes are normally between 1000 bytes and 2000 bytes.

Endpoint Capabilities

Different endpoints can have different capabilities for displaying and rendering streamed video. Most endpoints are standards-compliant and designed with profiles and levels in mind. However, endpoint capability can be limited by many factors, such as processing power, multimedia chipset capability, and the number of memory buffers available for multimedia processing.

Video endpoints have video buffers sized for the application. For streaming applications, real-time memory buffers are sized smaller than download and play buffers to balance the tradeoff between play latency and real-time rendering. The smaller buffers mean that it is important for data to be packaged in smaller chunks, so there is no delay reading the data while trying to render it in real time.

Therefore, for real-time streaming, it is beneficial to keep the size of video frames scaled appropriately. Issues can occur when trying to stream large I-frames to the mobile device, especially at the onset of video clips. If the video frame exceeds the capacity of endpoint streaming buffers, the video can freeze or lose the data necessary to render the video effectively.

Producing Content for Mobile Delivery

There are a few details that need to be considered when producing the final compressed video content for mobile networks. This section covers the tradeoff between using video transcoding and streaming the video natively from files. It also describes some considerations for choosing source material, for those who have control over the mobile content they intend to stream.

Choosing Between Transcoded and Native Video Streaming

One important factor to consider is whether to use transcoded video streaming or native video streaming for playing files. If the application streams video natively, you must take into consideration the network channel capacity when creating files, including the limitations of the 3G error prone radio channel and endpoint capability.

Transcoded Video Streaming

Transcoded video streaming involves adding a video transcoder to the video streaming path. An advantage of using a video transcoder is that source files can be stored in different formats, and the transcoder can perform transcoding, transrating, and/or scaling to match the negotiated endpoint video characteristics.

A video transcoder is required for advanced streaming capabilities, such as image overlay, text overlay, and video conferencing. In addition, video transcoding provides better performance when an endpoint requests Video Fast Updates (VFUs), because the transcoder can generate an I-frame on request to refresh a video endpoint. The VFU feature is especially valuable in providing a video refresh mechanism to mobile endpoints over wireless networks that might experience bit errors due to radio frequency data loss that can corrupt the video stream. By contrast, native files are used as they are stored, and thus do not have a dynamic capability to refresh video beyond what I-frames are already stored in the file.

As noted earlier, Dialogic® video transcoding uses a lossy compression technique. Therefore, it is best to produce video streams to be transcoded at a higher bitrate than the desired destination bitrate. This allows the Dialogic® transcoder to properly conform to a destination bitrate and use the bits in the most efficient way for the end audience. If possible, it is also a good practice to produce files that have the same frame rate and resolution size as the endpoint. Using a higher resolution input usually requires more processing to remove pixels. Often, there is not much benefit to producing higher frame rate files, because the extra frames are the first data dropped in the transcoding algorithm.

Taking these recommendations into account, to provide a video at QCIF resolution, 40 kbps and 10 fps out to a mobile endpoint, it is preferable to start with a QCIF 128 kbps 10 fps file instead of a CIF 128 kbps 15 fps file. This reduces the processing burden on the transcoder while allowing the Dialogic® transcoder to properly conform to the destination bitrate.

Native Video Streaming

Native video streaming uses pre-encoded files that have the characteristics required by the network and remote endpoint. Native streaming situations require that the characteristics of the audio and video data stored within the .3gp file match the capability expected by the endpoint or the network. In native streaming situations, the compressed data stored in the .3gp file contains the actual compressed data that will be streamed over the network and uncompressed at the mobile device.

Some benefits of streaming natively include higher achievable density, reduced latency, and overall reduced system cost due to more channels and less processing per port. However, a drawback is that it is up to the user to pre-encode files in the proper format for endpoints and make sure that the characteristics within the files are common for many different endpoints. There will likely be a need to have multiple pre-encoded files to cover the characteristics of different endpoints, networks, and video codecs. This is true, for example, when supplying files to different networks such as 3G-324M and IP, as well as for different codecs such as H.263, MPEG-4, and H.264. Separate pre-encoded files would need to be developed for each codec and bandwidth requirement. The optimal settings for native files depend on the characteristics of the audience.

Using Codec Profiles and Levels

The video characteristics within the files must match the characteristics required by the endpoint, especially when streaming native files to multimedia endpoints. Endpoints can normally support a wider range of video profiles and levels than those used for real-time streaming to cover the ability to play downloaded multimedia. However, network capacity normally adds more constraints to the video streaming capability through the network. The capacity of the network depends on many different factors. Some networks, like 3G-324M, are limited in throughput, but have a well-defined channel capacity. Other networks, like mobile broadband networks, are more difficult to specify, because they depend on the radio frequency technology to the endpoint, the carrier's network throughput, and the capabilities of the endpoints.

At low bitrates, there can be a significant tradeoff between fast action and video clarity. The more bits it takes to describe the scene changes in the video, the fewer bits there are to describe the detail within the video sequence. This is especially true for fast-paced content and content with frequent scene changes, such as movie trailers, music videos, and action scenes.

Important: Video content produced for download or for high bandwidth environments, such as TV or cinema, does not always translate well to low bandwidth mobile streaming formats.

For multimedia streaming, it is beneficial to customize video content for the mobile environment whenever possible. Producing video content for the mobile wireless network can be tricky, but there are general rules that can help keep the characteristics of the video better suited to handle video compression techniques.

Guidelines for Creating or Choosing Video Content

The following guidelines provide some considerations for creating or choosing video content suited to play successfully on mobile devices:

- Start with the highest quality source video.
- Reduce the amount of data that needs to be encoded.
- Reduce the amount of work the encoder needs to perform.

Starting with the Highest Quality Source Video

Start with the highest quality video possible and avoid re-encoding material that has low bandwidth characteristics. Video anomalies in the source video file are magnified during the compression and encoding of frames. If possible, an uncompressed video source is the best starting place.

Reducing the Amount of Data that Needs to be Encoded

Content created for mobile networks should limit the frequency of scene changes. Although scene changes do not always correspond to new I-frames and a complete refresh of the screen, they still correspond to greater spikes in bitrate in order to describe the changes between frames. In general, use video sequences that have scene changes greater than two seconds apart, and avoid videos that have scene changes fewer than one second apart. Also, limit fast panning and zooming sequences. The faster the panning or zooming in the video sequence, the greater the differential in motion from one frame to the next. This results in a higher average bitrate for the fast moving sequences.

Reducing the Amount of Work that the Encoder Needs to Perform

Encoder inefficiencies can occur when extra bandwidth is used up trying to encode video anomalies or areas of the picture that are not the primary video subject. Video encoders work hard to encode noisy or grainy video that has many light and dark pixels, because encoding algorithms normally treat all areas of the picture with equal priority. Eliminating video noise and extra brightness from the source video can improve an encoder's efficiency.

When possible, keep the foreground subject distinct from the background by maintaining a depth of view. Dark homogeneous background colors keep the video subject in focus while isolating it from the background. People on camera should stay away from white clothing, pin stripes, and complex patterns as each of these can cause unwanted video distortion and lower the overall quality by making it harder for the encoder to do its job effectively.

Choosing Video Characteristics for Mobile Delivery

When a video file is streamed natively through the network, it must be encoded properly to fit within the network and endpoint constraints. Generally, users need to use a video encoder conversion tool to produce a file with the proper video characteristics for the intended audience. This section describes some of the settings available in various encoder tools and how these settings apply to a low bandwidth mobile network like 3G-324M.

Choosing a Video Encoder Tool

After choosing or producing a video for the mobile environment, use a video encoder content creation tool to convert the file into the .3gp format for mobile streaming. Video encoder creation tools that have a greater number of parameters make it more convenient to tweak settings and achieve better quality, especially at very low bandwidths. Although there are many video encoder creation tools that can convert to .3gp format, not all of these tools can convert to .3gp files with target bitrates under 42 kbps, which is required for the channel capacity of 3G-324M networks. Video encoder creation tools that can achieve sufficient video compression include Helix Mobile Producer, Vidiator Xenon, and QuickTime Pro. Each of these tools has a varying degree of success at very low bitrates, but, in general, they can be used to produce 3G-324M files adequate for native streaming.

Specifying Video Encoder Characteristics

When creating a video for a 3G-324M network, you must specify the following video characteristics:

- Video codec
- Video resolution
- Frame rate
- I-frame distance
- Target bitrate

You can also optionally specify whether hint tracks are used, and the number of macroblocks to use as Intrablocks (I-blocks).

Selecting the Video Codec

As mentioned earlier, the target video codec is based on what profiles are supported on the network and at the endpoint. When choosing a codec, consider what endpoints are being used in the network, the desired level of interoperability, and the level of encoder efficiency (often corresponding to video quality) desired at the endpoint.

Choose one of the following codecs:

- H.263
- MPEG-4
- H.264

H.263 baseline profile 0 is the most common video codec on 3G-324M networks, because the 3GPP standard requires all endpoints to support it. Therefore, it can be especially beneficial to use H.263 baseline profile 0 with native streaming, to assure that the video content stored in the file can be rendered at the majority of 3G-324M endpoints.

The MPEG-4 video codec adds advanced error detection and correction services to the services that H.263 baseline profile 0 provides. Because of these techniques, MPEG-4 can offer better video quality than H.263, and it is supported by the majority of endpoints. On the downside, Decoder Configuration Information (DCI) must be passed from the encoder to the decoder to properly decode the video bitstream. On 3G-324M networks, the DCI is passed in the control protocol, and therefore, all natively streamed video files must have the same DCI embedded in the video for compatibility when playing multiple files to remote endpoints. Transcoded streaming solutions do not have this limitation, because the transcoder applies the same DCI to every file within the same session.

The new H.264 video compression standard greatly improves on encoder efficiencies. H.264 yields better picture quality, while significantly lowering the required bitrate. H.264 is considered a major enhancement to video compression standards and the optimal codec for low bitrate applications. However, because H.264 was recently added to the 3GPP specification, most 3G-324M endpoint vendors have been slow to adopt it as of this writing.

Setting the Video Resolution

As discussed earlier, the target video resolution of the file is based on the audience profile. For most mobile networks, the target video resolution is QCIF resolution (176 x 144 pixels), because it is well suited for mobile screens and mobile network bandwidths. To limit video distortion and possible encoder inefficiencies, make the original video dimensionally close in proportion to the target resolution.

Setting the Frame Rate

Keeping both the peak and average bitrates down might include tweaking settings such as the target frame rate and maximum keyframe distance. For constant bitrate files, a lower number of frames per second, like 10 fps, produces better video clarity at the cost of a noticeably choppier video. At 15 fps, the average bitrate is higher, but the peak bitrate of the video may be reduced. Higher frame rates are normally not recommended for 3G-324M networks unless accompanied by a reduction in the target bitrate, because the increased data rate may be too much for the endpoint memory buffers to handle. For 3G-324M networks, a target frame rate between 6 fps and 10 fps is suitable.

Setting the I-Frame Distance

Maximum I-frame distance is the interval is used to create full reference video frames at a periodic rate. The smaller the distance between I-frames, the higher the resulting bandwidth, because I-frames require more bits to describe the entire video frame. Lengthening the distance between I-frames helps reduce the overall bitrate of the file, but there is a longer time to recovery if the video data gets corrupted.

A key to controlling bitrate spikes is to reduce the number of transitions in the video that would cause a greater number of I-frames. For mobile networks, it is generally acceptable to have an I-frame interval between 4 and 10 seconds to balance the bandwidth constraint and the need to provide complete frame updates over error-prone radio networks.

Specifying the Target Bitrate

The goal of video encoding is to keep the video at a constant bitrate, as close to bitrate of the target audience as possible without exceeding the channel capacity. Video encoder tools include a target bitrate, which is used to conform to the audience and which, for most video encoder tools, is an average target bitrate rather than a peak bitrate. Periodic bursts in video data can cause an overflow of the video channel capacity. It is often important to reduce the overall target bitrate to guarantee that a peak bitrate does not exceed the maximum channel capacity. The more effective the encoder tool is at keeping the average constant and the peaks at the average, the higher the specified average can be. For 3G-324M network files, set the target bitrate between 35 kbps and 38 kbps to ensure that the video encoder can keep the bitrate peaks under 50 kbps, unless you know that the maximum bitrate will not be exceeded.

Using Hint Tracks

Files used for media streaming can contain hint tracks (also called streaming server profile), which describe how the media data within the files should be packetized. Using hint tracks reduces the pre-processing required to stream a file. The data within the hint tracks provides pacing and synchronization for real-time media streaming.

I-Block Refresh

A macroblock is a compressed video block of 16 by 16 pixels. Some video encoders provide the ability to encode a number of macroblocks as Intrablocks (I-blocks). This enables decoders to refresh the changed I-blocks within a frame, instead of refreshing the whole frame. Encoding macroblocks as I-blocks provides a method for error resiliency and error recovery time in case a frame gets lost or corrupted over the radio channel. It can also produce more consistent frame sizes and lower the average bitrate. The tradeoff is that it may take a few seconds to update the screen while the I-blocks refresh.

Analyzing Video for Conformance to Targets

This section describes how to use the Dialogic® hmp3gp utility to verify that the .3gp file conforms to the targeted video characteristics.

Dialogic® hmp3gp Utility

The Dialogic® hmp3gp utility is a tool that converts multimedia data between Dialogic® multimedia format files and 3rd Generation Partnership Project (3GPP) format files. (The multimedia format files are used on Dialogic® Host Media Processing (HMP) platforms.) Through specific command line arguments, hmp3gp can also generate a comma separated value (.csv) file with detailed information about each video frame in the .3gp file. You can import the .csv file into a spreadsheet program and use the data to analyze the file's video frame characteristics.

The hmp3gp utility is a valuable tool for analyzing .3gp files, because it can help determine if a .3gp file is suitable for the 3G-324M or mobile streaming environment. It is available on the Dialogic website at http://www.dialogic.com/products/ip_enabled/download/multimedia/omf.htm.

Video Analysis using Dialogic® hmp3gp

The process in Table 3 can be used as a guideline for analyzing video statistics using hmp3gp. Use this process to determine if an input .3gp file has the proper video characteristics for the 3G-324M network.

Note: There are no hard and fast rules for analyzing video statistics. Videos can sometimes violate one or more of these guidelines without the user noticing a degradation in quality.

Step	Action
1	Create a.3gp file from the video source using one of the following video encoder conversion tools. • Helix Mobile Producer • Vidiator Xenon • QuickTime Pro For guidelines on specifying video encoder characteristics during the file creation process, see *Specifying Video Encoder Characteristics*, earlier in this white paper.
2	Use the hmp3gp -d option to convert the.3gp file into separate Dialogic® multimedia format video (.vid) and audio (.aud) files. For example: **`hmp3gp -d1 file.vid file.aud input.3gp`**
3	Use the hmp3gp -rd option to generate a comma separated value (.csv) file from the Dialogic® multimedia format video (.vid) file. For example: **`hmp3gp -rd file.vid`**
4	Optionally, import the .csv file to a spreadsheet program like Microsoft® Excel®. For more information, see *Importing the .csv File to a Spreadsheet Program*, later in this white paper.
5	Analyze the data for conformance, as described in *Analyzing the Data for Conformance*, later in this white paper.
6	Determine whether the video file satisfies the network requirements. A video file is likely to be satisfactory if all of the following conditions are met: • Peak moving average bitrate is below the network channel capacity. • I-frame frequency is within tolerances. • Peak frame size is below the maximum endpoint and network constraint.
7	If the .3gp file conforms to the targets, then it can be streamed from the Dialogic® platform. If the .3gp file is not acceptable, tweak the video encoder settings to re-encode the video using lower target bitrate, fewer frames per second, or longer I-frame distance.

Table 3. Analyzing Video Statistics using the Dialogic® hmp3gp utility

Figure 1 illustrates the process of performing video analysis with hmp3gp:

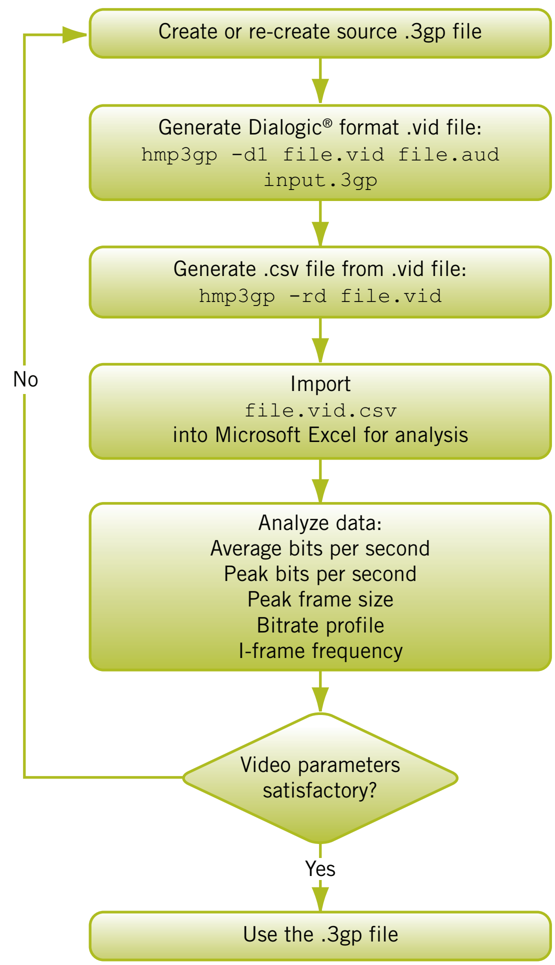

Figure 1. Video Analysis using the Dialogic® hmp3gp utility

Importing the .csv File to a Spreadsheet Program

You can optionally open the .csv file in a spreadsheet program like Microsoft® Excel® for analysis. The .csv file output contains statistics for the video file, with each row representing a video frame. The end of the .csv file contains summary data over all frames in the file.

Table 4 shows the .csv file's video frame statistics:

Index	Time	Frame Size (Bytes)	Frame Size (Bits)	Avg Bytes Per Sec	Avg Bits Per Sec	I-Frame	Time Since Last I-Frame (Sec)	Average Time Since Last I-Frame (Sec)
1	0	1320	10560	1320	10560	1	0	0
2	0.111111	23	184	1343	10744	0	0.111111	0
3	0.222222	23	184	1366	10928	0	0.222222	0
4	0.333333	23	184	1389	11112	0	0.333333	0
5	0.444444	36	288	1425	11400	0	0.444444	0
6	0.555556	366	2928	1791	14328	0	0.555556	0
7	0.666667	255	2040	2046	16368	0	0.666667	0
8	0.777778	375	3000	2421	19368	0	0.777778	0
9	0.888889	885	7080	1986	15888	0	0.888889	0
10	1	112	896	2075	16600	0	1	0

Table 4. Video Frame Statistics

Table 5 shows the .csv file's summary information. This information is used for the examples in the *Analyzing the Data for Conformance* section below.

Frame Rate	Peak Frame Size (Bytes)	Peak Frame Size (Bits)	Peak Bytes Per Sec	Peak Bits Per Sec	Average Bits Per Sec
10	4266	34128	12339	98712	32318

Table 5. Video Summary Information

Analyzing the Data for Conformance

To analyze the data for conformance:

- Review the summary section to get a quick estimate of video file conformance over the entire file. Pay close attention to the data in the Peak Frame Size, Peak Bits Per Sec, and Average Bits per Sec fields.
- If the data requires further analysis, plot the data to show the video frame statistics over time, including the bitrate profile and I-frame frequency. This analysis can show potential problem areas within the file.

The fields described here provide useful statistics for analyzing conformance.

- **Peak Frame Size** — provides the maximum size of the largest video frame in the file. As discussed in *Endpoint Capabilities*, earlier in this white paper, the peak frame size can be an important measure of how much data needs to be processed at the endpoint. It can also indirectly reflect the amount of time an endpoint must wait to receive a complete frame. In this example, a peak frame size of ~34 kbps means that the largest video frame will take over ½ second to be sent across the 3G-324M network.
- **Peak Bits Per Second** — provides the maximum number of bits required to transmit the video over a period of one second. This value can help to quickly show whether the largest spike in bitrate is too much for the network to handle. In this example, the Peak Bits Per Second is ~98 kbps, which is much too high to fit into the 3G-324M 64 kbps data pipe.
- **Average Bits Per Second** — provides the median for a one-second rolling average over the entire video, and provides a good estimate to determine if the video has the proper target bitrate. (Do not confuse average bitrate with bps transmitted over the duration of the complete file.) As noted earlier, for 3G-324M networks, the video bandwidth should be less than 42 kbps to conform to network constraints. In this example, the Average Bits Per Sec is ~32 kbps, which is well within the 3G-324M network constraint.

Reviewing the Bitrate Profile

To review the bitrate profile of the video, plot the Average Bits Per Sec data from the .csv file against the frame index to show the bitrate conformance over time. The hmp3gp utility calculates Average Bits Per Sec as the average number of bits transmitted over the last one second rolling window. The bitrate conformance graph is very important, because it shows the areas of the video where there may be spikes in bitrate over a few frames. It also shows the peak data rate that occurs over the one second window with respect to the video frame.

Figure 2 shows what an ideal bitrate plot looks like. The video conforms well to target bitrate characteristics and does not have large spikes in bandwidth that exceed the channel capacity of the 3G-324M network (42 kbps).

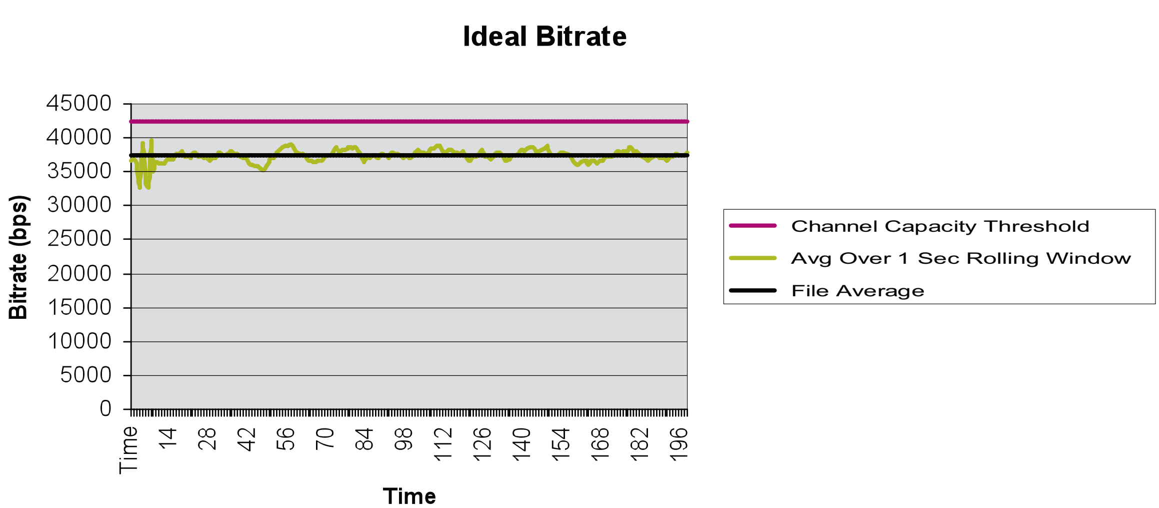

Figure 2. Ideal Bitrate Graph

Figure 3 shows the video bitrate profile for a second video clip destined for a 3G-324M network. In this video clip, the average bitrate across the file is approximately 37 kbps and stays below the 42 kbps channel capacity of the 3G-324M network. However, there are peaks in the rolling average (circled) that signal areas of the file that may experience poor video quality, A/V synchronization problems, or video anomalies, as the data is delayed or dropped through the network.

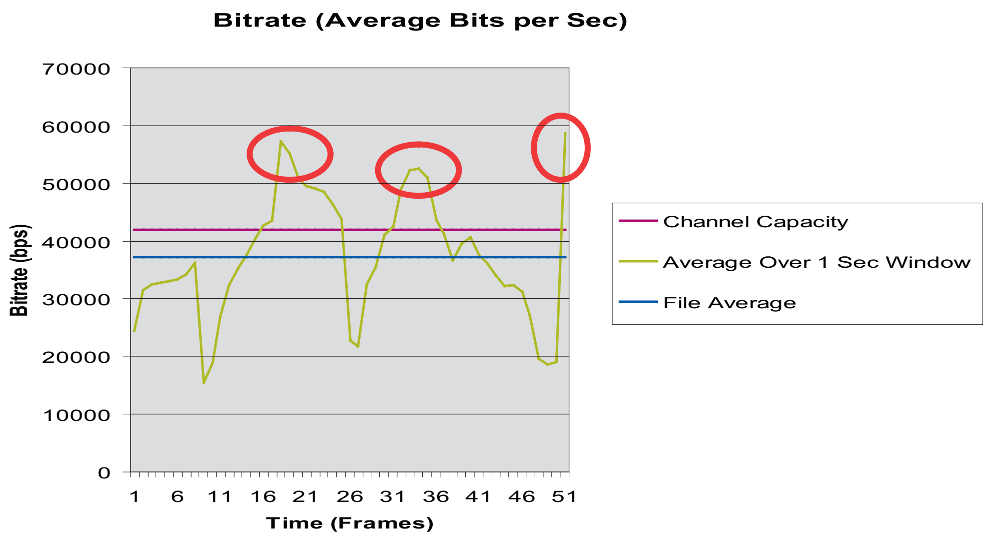

Figure 3. Video Bitrate Profile Graph

Graphing the I-Frame Frequency

Different video encoder tools present the option of I-frames generation differently. Some tools allow periodic or automatic generation of I-frames at specific intervals. Other tools may seemingly generate I-frames haphazardly; from an I-frame only at the start of a file to an I-frame for every minor scene change. In an error-prone data-transmission environment, such as 3G-324M or mobile networks, having a single I-frame in the bitstream is not recommended. This is especially true for native file streaming, where I-frames cannot be generated on demand. Likewise, a new I-frame for every minor scene change is not desirable, because this will likely drive up the bandwidth usage.

Figure 4 shows the presence of I-frames in a video bitstream.

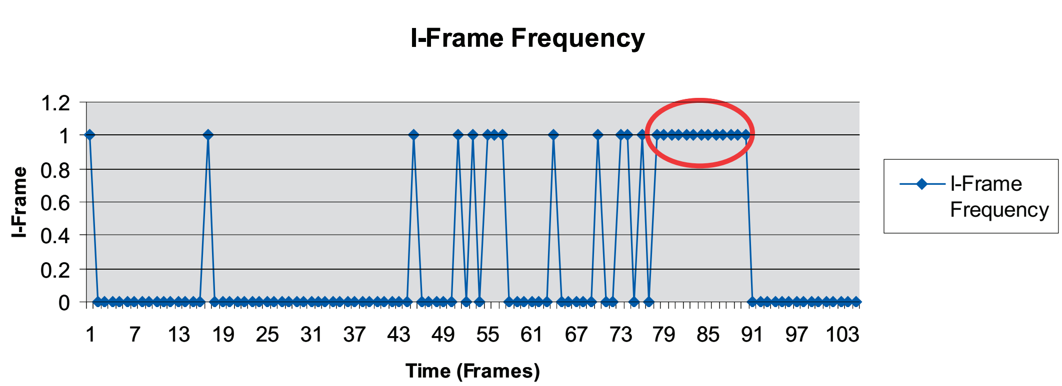

Figure 4. I-Frame Frequency Graph

In this graph, there is an I-frame present at the start of the bitstream (which is true for all bitstreams). Toward the end of this video clip, the I-frames arrive more frequently, perhaps due to the complexity of the subsequent scenes or too many scene changes. If this file is destined for the 3G-324M network, there will likely be a bandwidth issue where the I-frames arrive frequently. In this situation, consider setting a greater I-frame distance to force the video encoder to generate fewer I-frames in the file.

Dialogic® Products That Support Video Streaming

Dialogic delivers highly scalable multimedia processing platforms that can save development time and reduce costs and time-to-market for solution providers who are developing multimedia and video applications. Below is a partial list of Dialogic® products that support multimedia and that can be used in 3G mobile service solutions that include end-to-end video streaming, text overlay, Multimedia Ring Back Tone (MRBT), mobile advertising, video share, and mobile social networking:

- **Dialogic® Host Media Processing Software** — media processing software that provides services for flexible, scalable, and cost-effective IP and 3G-324M multimedia platforms
- **Dialogic® Vision™ VX Integrated Media Platform** — a standards-based integrated media and signaling platform supporting VoiceXML and CCXML for IP and 3G-324M multimedia solutions
- **Dialogic® Vision™ CX Video Gateway** — a gateway designed to connect SIP-based video and multimedia services to both PSTN endpoints and 3G-324M mobile video-enabled phones
- **Dialogic® IP Media Server** — a software-based multimedia server that leverages the simplicity, openness, and flexibility of SIP, VoiceXML, and MSCML, and that can provide Media Resource Function (MRF) capability in IMS environments

Obtaining Third-Party Licenses

Using the AMR-NB resource in connection with Dialogic HMP Software 4.1 does not grant the right to practice the AMR-NB standard. To seek a patent license agreement to practice the standard, contact the VoiceAge Corporation at http://www.voiceage.com/licensing.php.

Acronyms

3GPP	Third Generation Partnership Project
4CIF	4 Times Common Intermediate Format
CBR	Constant Bitrate
CIF	Common Intermediate Format
DCI	Decoder Configuration Information
FPS	Frames Per Second
HMP	Host Media Processing
Kbps	Kilobits Per Second
LTE	Long Term Evolution
Mbps	Megabits Per Second
MPEG	Motion Picture Experts Group
MRBT	Multimedia Ring Back Tone
MRF	Media Resource Function
PSTN	Public Switched Telephone Network
QCIF	Quarter Common Intermediate Format
QOS	Quality Of Service
SIP	Session Initiation Protocol
Sub-QCIF	Sub-Quarter Common Intermediate Format
VBR	Variable Bitrate
VFU	Video Fast Update
WiMax	Worldwide Interoperability For Microwave Access

For More Information

Ron Garrison, "Producing Content for Mobile Delivery," Feb 15, 2008 available at http://www.streamingmedia.com/article.asp?id=10093&page=1&c=31

Kumar, Amitabh. *Mobile TV: DVB-H, DMB, 3G Systems and Rich Media Applications*. Oxford: Elsevier Inc., 2007.

Richardson, Iain E.G. *H.264 and MPEG-4 Video Compression: Video Coding for Next Generation Multimedia*. Chichester: John Wiley & Sons Ltd., 2003

"Optimizing A/V Content For Mobile Delivery: Media Encoding using Helix Mobile Producer 11.0", November 3, 2005 available at http://docs.real.com/docs/mobileproducer11/HMP11_WhitePaper.pdf

3GPP TS 26.210: "*Codec for circuit switched multimedia telephony service; General Description*".

3GPP TS 26.211: "*Codec for circuit switched multimedia telephony service; Modifications to H.324*".

3GPP TS 26.244: "*Transparent end-to-end packet switched streaming service (PSS); 3GPP File Format*".

Industry associations are also important sources of information:

- 3GPP (Third Generation Partnership Project) — Association that provides access to all 3GPP network and technology standards and posts industry and partner news
- NGMN (Next Generation Mobile Networks Alliance) — A mobile network operator association formed to support evolution to packet-based mobile broadband networks
- 3G Americas — Association of mobile operators and manufacturers in the Americas whose products and services relate to 3GPP standards

Dialogic
Making Innovation Thrive™

www.dialogic.com

Dialogic Corporation
9800 Cavendish Blvd., 5th floor
Montreal, Quebec
CANADA H4M 2V9

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH PRODUCTS OF DIALOGIC CORPORATION OR ITS SUBSIDIARIES ("DIALOGIC"). NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN A SIGNED AGREEMENT BETWEEN YOU AND DIALOGIC, DIALOGIC ASSUMES NO LIABILITY WHATSOEVER, AND DIALOGIC DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF DIALOGIC PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY INTELLECTUAL PROPERTY RIGHT OF A THIRD PARTY.

Dialogic products are not intended for use in medical, life saving, life sustaining, critical control or safety systems, or in nuclear facility applications.

Dialogic may make changes to specifications, product descriptions, and plans at any time, without notice.

Dialogic and Vision are either registered trademarks or trademarks of Dialogic Corporation or its subsidiaries. Dialogic's trademarks may be used publicly only with permission from Dialogic. Such permission may only be granted by Dialogic's legal department at 9800 Cavendish Blvd., 5th Floor, Montreal, Quebec, Canada H4M 2V9. Any authorized use of Dialogic's trademarks will be subject to full respect of the trademark guidelines published by Dialogic from time to time and any use of Dialogic's trademarks requires proper acknowledgement.

Microsoft and Excel are registered trademarks of Microsoft Corporation in the United States and/or other countries. Other names of actual companies and products mentioned herein are the trademarks of their respective owners. Dialogic encourages all users of its products to procure all necessary intellectual property licenses required to implement their concepts or applications, which licenses may vary from country to country.

Any use case(s) shown and/or described herein represent one or more examples of the various ways, scenarios or environments in which Dialogic® products can be used. Such use case(s) are non-limiting and do not represent recommendations of Dialogic as to whether or how to use Dialogic products.

Copyright © 2009 Dialogic Corporation All rights reserved.

09/09 11489-01