White Paper

How to Build a "Lean and Mean" Video Gateway Using 3G-324M-over-IP

by Avi Fisher
Surf Communication Solutions

avi@surf-com.com

Abstract

Video-based applications in the mobile network are one of the main drivers for the deployment of 3G networks. Both the 3GPP and 3GPP2 standards organizations have adopted the 3G-324M protocol as the primary means of transporting conversational video over mobile networks. Traditional video gateways bridge between the representation of video in the mobile network (3G-324M) and in the IP network (video/voice over RTP). They have substantial disadvantages, however. This white paper discusses the role of 3G-324M-to-IP video gateways in mobile video applications and introduces an alternative approach, called the "lean and mean" video gateway. It presents basic implementation guidelines and describes the approach's significant advantages.

1. Introduction: The Need

Video-based applications in the mobile network are one of the main engines driving the deployment of 3G networks. Both the 3GPP and 3GPP2 standards organizations have adopted the 3G-324M protocol as the primary means of transporting conversational video over mobile networks.

The 3G-324M protocol combines voice, video, data and control into a single 64kbps stream of circuit-switched data. It differs considerably from the prevailing protocol for transporting real-time video and voice over IP networks, RTP/RTCP (Real-time Transport Protocol/Real-time Transport Control Protocol). In the latter case, voice, video and control are transported on different streams. Two RTP streams are used for voice and video and an RTCP stream for control. An additional stream is used for the media control: H.245 in H.323 and SDP in SIP. The four streams do not necessarily accumulate into a constant bit rate.

Traditional video gateways bridge between the representation of video in the mobile network (3G-324M) and in the IP network (video/voice over RTP). They perform the following tasks:

- Termination of the 3G-324M protocol toward the mobile network
- Termination of the SIP/H.323 signaling protocol toward the IP network
- Transcoding video from its native representation in the mobile network (typically MPEG4 with a low bit rate) into its other representations in the IP network (H.264 or H.263, typically with much higher bandwidth)
- Transcoding voice from its native representation in the mobile network (typically AMR) into its other representation in the IP network (typically G.729, G.723 or G.711)

This set of substantial tasks introduces four considerable disadvantages when traditional video gateways are used to bridge between the mobile and IP networks:

- **System complexity**: The large number of processes required to terminate the 3G-324M protocol toward the mobile side and the SIP/H.323/RTP protocols toward the IP side, and optionally to perform voice and video transcoding—which include protocol conversion, complex state machine handling and media processing—makes the traditional video gateway significantly more complex than voice gateways.
- **Cost:** If video and voice transcoding is needed, the cost of a video gateway (dollars per port) is extremely high compared with that of voice gateways. High cost is one of the barriers to the mass penetration of mobile and wireless video.
- **Voice and video gateway in two separate systems**: Video gateways require much greater processing resources than do voice gateways. In addition, video and voice transcoding, if needed, introduces technological complexities. Both requirements slow down the convergence of voice and video gateways into a single box.
- **Delay**: Terminating video over 3G-324M toward the mobile side and video over RTP toward the IP side results in an additional end-to-end delay that can be detrimental to the user's overall experience, as does transcoding video and voice, if that is required.

These disadvantages are impeding the move to 3G networks.

An alternative approach to that of the traditional mobile video gateway calls for a "lean and mean" gateway that performs the circuit-switch-to-packet-switch layer 1 and 2 bridging (much as a G.711 VoIP gateway does), and offloading all the other tasks either to a media server located in the core IP network or to the IP videophone itself.

2. Traditional 3G-324M-to-IP Video Gateways

Mobile-to-IP video gateways convert the video and audio signals from their representation in the mobile network into their representation in the IP network.

In the IP network, video content is carried by a set of three separate UDP/IP streams:

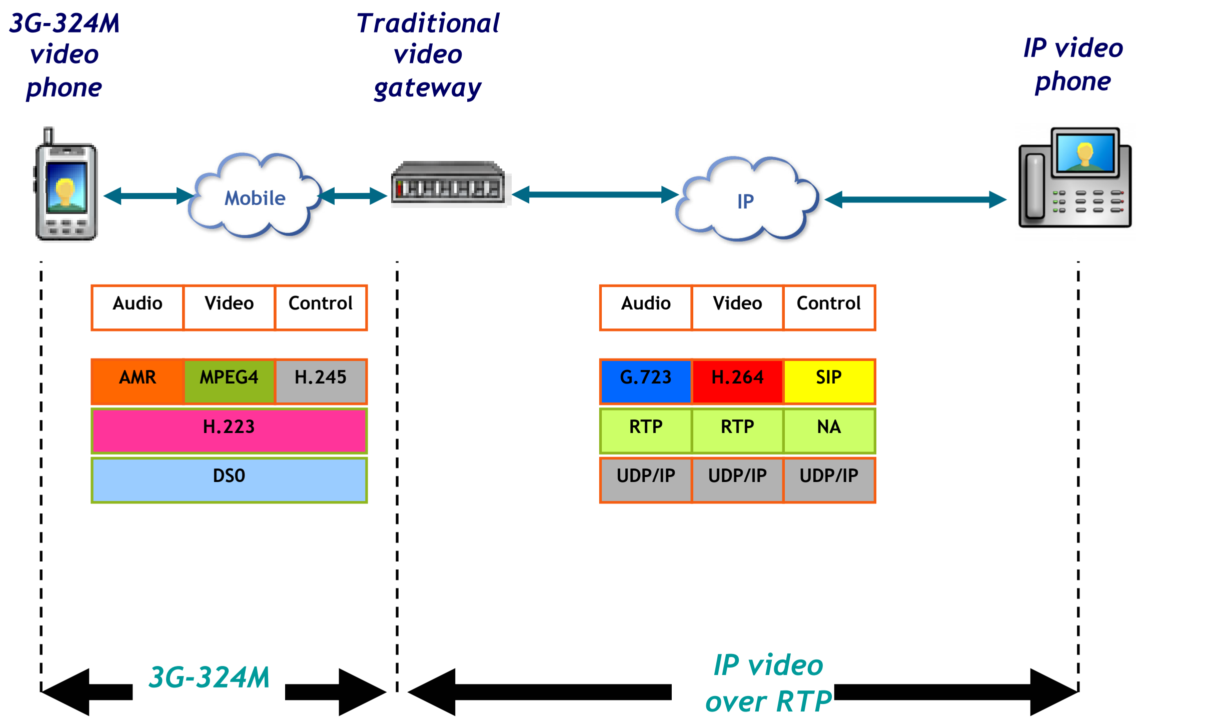
- A compressed video stream (compression using H.264/H.263/H.261 or MPEG4), encapsulated in an RTP stream, which is placed in a UDP/IP packet
- A compressed audio stream (compression using G.711,G.729, G.723 or any other voice compression scheme), encapsulated in an RTP stream, which is placed in a UDP/IP packet
- Signaling and control information, which is carried over the network primarily using SIP or H.323

In the mobile network, 3G-324M–based video content is carried by a single H.223 64kbps stream that multiplexes audio, video, data and control information. In most cases, the video portion of the H.223 protocol is based on MPEG4, whereas the audio portion is based on NB-AMR compression. The control portion of the H.223 stream is based on the H.245 protocol, which is primarily responsible for channel parameter exchange and session control.

Traditional mobile-to-IP video gateways support a wide range of applications. The main ones are:

- **Conversational video** between a mobile user and an IP user: In this application, video and audio are transferred in full-duplex mode.
- **Video messaging:** In this application, video messages and clips, stored in a server, are streamed to the mobile user and vice versa.
- **Videoconferencing:** In this application, several conversational video streams are combined to enable an N-party videoconference. Some of the participants might be located on the mobile network and others located on the IP network. .
- **Video surveillance:** The gateway allows a user with a video-enabled mobile phone to view the video captured by a camera that streams its output to the IP network and vice versa. In this application, video is transferred in half-duplex mode.

The following diagram illustrates the protocol and media conversions performed by the traditional video gateway.



Note: SIP, H.264 and G.723 are just examples of the signaling, video compression and audio compression protocols used in the IP network.

As can be seen, the video gateway performs these tasks:

- Video transcoding
- Audio transcoding
- SIP termination
- H.245 termination
- RTP/UDP/IP termination
- H.223 termination
- SDP-to-H.245 interworking

In the applications above, one or more of four key parameters determine the quality of the IP-to-mobile gateway solution for the end user and the operator:

- Call setup time—conversational video, videoconferencing
- Video quality—video messaging, video surveillance
- Support for high-density systems with a low per-port cost—all four
- End-to-end delay—conversational video, videoconferencing

None of these parameters are handled optimally by the traditional video gateway approach, for several reasons:

- **End-to-end delay:** Since IP packets are converted into an H.223 stream, they must first be accumulated before conversion into a stream. Typically, this results in an additional 30-60ms delay in each direction. Moreover, when the video gateway performs transcoding—for example, from H.264 to MPEG4—an additional 100-150ms delay is incurred because of the video codec's inherent delay and the interprocess communication within the gateway.
- **Call setup time:** Since the video gateway terminates the call toward each network, call setup is accomplished when both legs of the call are set. In some cases, in an error-prone environment, when parameters are relayed between the IP leg toward the mobile leg and SIP messages are converted into H.245 messages, call setup time can double compared with the all-IP or all-mobile cases.
- **Video quality:** When transcoding occurs, video is converted from one compression scheme into another. Because the schemes have different bit rates, transcoding inherently reduces the video quality.
- **High-density system:** As noted, 3G video gateways require significantly more CPU processing power per port than do voice gateways, even ten times as much. This requirement results in the almost impossible task of combining voice and video gateways into a single does-it-all, triple-play gateway.
- **Low per-port cost:** Also as noted earlier, transcoding video and voice, if necessary, is a significant factor in the per-port cost. Eliminating this need will enable gateway cost to drop and density (the number of ports per given footprint of space) to increase dramatically.

3. "Lean and Mean" Video Gateways Using 3G-324M-over-IP

The "lean and mean" video gateway approach is based on three principles:

- The video gateway does not perform IP video protocol termination (3G-324M).
- The video gateway does not perform IP video protocol termination (video/audio over RTP, with SIP/H.323 for control).
- The video gateway does not perform voice and video transcoding. Instead, voice and video transcoding are performed in a separate media server, should the need for transcoding arise.

These principles will dramatically reduce the complexity of the video gateway. In addition, the third significantly reduces the per-port cost, increases video quality and maintains the flexibility needed in case the video source and destination clients are unable to support the same voice and video protocols and bandwidth requirements.

Both of the methods that follow implement the "lean and mean" gateway. The first is used if IP conversion is done by the IP videophone, the second if it is done by the media server.

Method 1

1. Mobile and IP end points implement video sessions using the 3G-324M protocol, which is terminated end-to-end. On the mobile side, video is transmitted as a standard 64kbps stream; on the IP side, the 64kbps stream is encapsulated in UDP/IP packets. The end point on the IP side is implemented in the IP videophone.

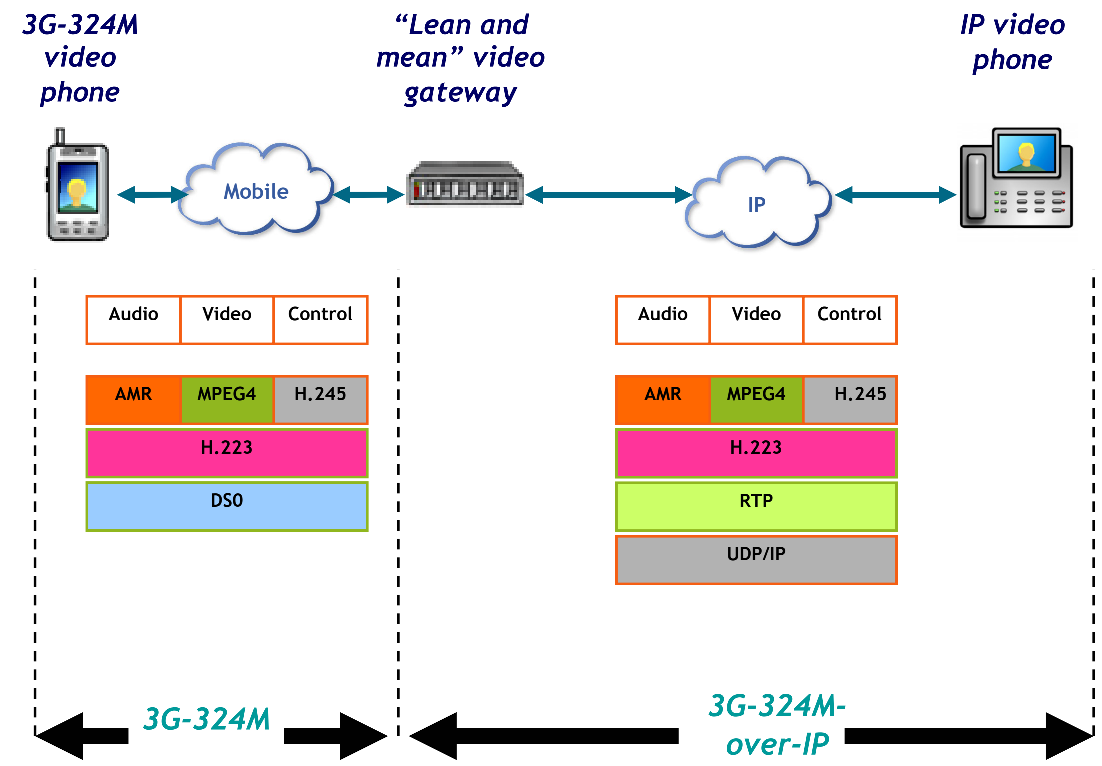
2. The encapsulation of 3G-324M into IP packets is similar to the encapsulation of voiceband data into IP packets. The most common way is to break up the 3G-324M stream into frames of a given size and to add RTP heads with optional redundancy and forward error correction mechanisms, in order to increase transport reliability.
3. Optimally, RTP encryption should be used.
4. In case end-to-end delay is not an important factor (that is, in messaging and video surveillance applications), reliable transports of UDP packets can be used. One example is the SPRT protocol, used in ITU-V.150.1.

Method 2

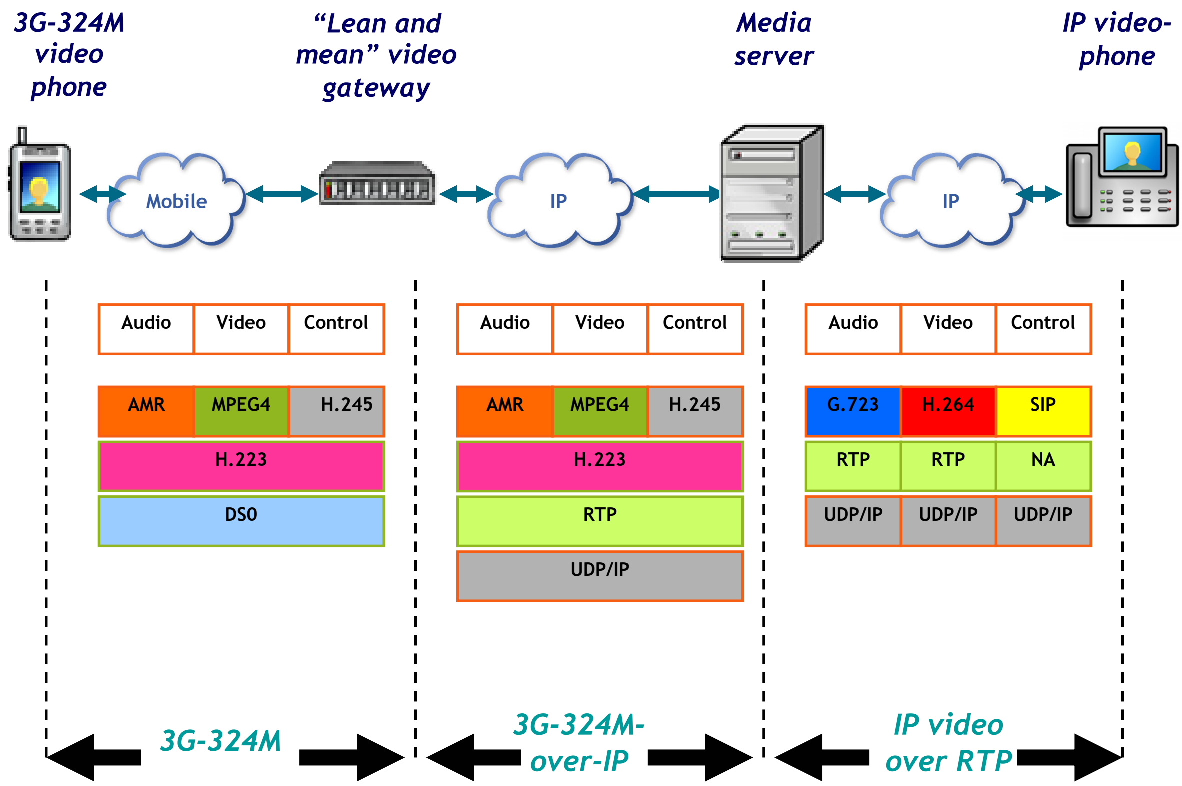
1. The mobile end point implements video sessions using the 3G-324M protocol, and the IP video end point implements video sessions using video/voice over RTP, plus SIP/H.323 for signaling and control. The media server in the IP network converts IP video into an H.223 64kbps stream that is encapsulated over RTP/UDP.
2. The encapsulation of 3G-324M in IP packets is similar to the encapsulation of voiceband data in IP packets. The most common way is to break up the 3G-324M stream into frames of a given size and add RTP heads with optional redundancy and forward error correction (FEC) mechanisms, in order to increase transport reliability.
3. Optimally, RTP encryption should be used.
4. In case end-to-end delay is not an important factor (in messaging and video surveillance applications), a reliable transport of UDP packets can be used, such as the SPRT protocol, used in ITU-V.150.1.

These methods have the name *3G-324M-over-IP*.

The following diagram illustrates the protocol conversions performed by the "lean and mean" video gateway when the IP videophone supports 3G-324M-over-IP (*Method 1*).



The next diagram illustrates the protocol conversions performed by the "lean and mean" video gateway and the media conversion done by the media server if the IP videophone does not support 3G-324M-over-IP (*Method* 2).



Surf Communication Solutions, Ltd. ♦ Tavor Building ♦ P.O. Box 343 ♦ Yokne'am 20692 ♦ Israel
Tel: +972 4 9095 777 ♦ Fax: +972 4 959 4055 ♦ e-Mail: surf@surf-com.com ♦ **www.surf-com.com**

As can be seen, in both cases the "lean and mean" video gateway performs only one task: RTP/UDP/IP termination.

In the second case, the media server performs:

- Voice and video transcoding
- SDP/H.245 interworking
- RTP/UDP/IP termination
- H.223 termination

4. Implementation Guidelines

As indicated above, the "lean and mean" video gateway is no more than a regular VoIP gateway operating in VBD (Voice Band Data) mode. A gateway working in this mode has these requirements and features:

a. The media type is set to transparent pass-through of 64kbps data from the IP/UDP/RTP side to the E1/T1 side and vice versa. In common practice, VoIP gateways enter this mode of operation in one of the following scenarios:
    - The gateway was configured to operate in this mode regardless of incoming media streams.
    - The gateway will enter this mode upon reception of PSTN signaling information—for example, in the user-to-user information field of the ISDN-PRI signaling.
    - The gateway will enter this mode upon reception of IP signaling information—for example, in the MGCP signaling with the media gateway controller.
    - The gateway will enter this mode upon reception of various flavors of modem/fax tones—for example, reception of a 2100 tone with phase reversals. To accommodate this method, the 3G-324M stack of the mobile phone must transmit a modem tone, in order to get the gateway to "think" it is going to handle a modem call and enter the pass-through mode.

b. The jitter buffer of the gateway must be configured to a fixed depth or to adapt very slowly. If the jitter buffer is fixed, the depth should be configured to 80ms. Otherwise, rapid jitter buffer adaptations may cause unwanted loss of frames.

c. The RTP payload type on which the 3G-324M stream is carried can be the same as for the regular G.711 voice.

d. The same redundancy and FEC mechanisms can be used as for modem transmissions over G.711 (VBD mode).

e. The IP payload frame size should be set to the smallest possible value that can be handled by all the other components on the IP side of the system. A recommended value is 10ms.

f. The E1/T1 interface must be a link without attenuation pads and robbed-bit signaling, to ensure that a transparent 64kbps stream can flow from the mobile phone to the IP network.

5. Summary of the Benefits of the "Lean and Mean" Approach

The alternative approach presented here provides the following major advantages over the traditional video gateway approach:

1. Low gateway cost, which helps the mass deployment of video telephony applications
2. An easy path for voice gateway manufacturers to upgrade their products to support video when bridging between the IP and mobile networks
3. An easy path for media server manufacturers to support the 3G-324M protocol without the need to physically support T1/E1 circuit-switched interfaces
4. Decoupling the IP-to-TDM conversion process from the video processing, which makes IP-to-TDM conversion as simple as implementing a G.711 VOIP gateway, and offloading the heavy video processing to dedicated media servers that do not have an interface to the TDM-based networks.
5. Better conversational video experience due to a significantly lower end-to-end delay if IP videophones support *3G-324M-over-IP*.
6. Shorter call setup time due to the elimination of the transport protocol conversion in the gateway (RTP to H.223).

For more information, please contact avi@surf-com.com.