Video Skew Factor and MDI Scoring

A white paper describing a practical method of simulating video phones for determining the delay (Skew Factor) between video and audio packets in an IP network

askzeke@ameritec.com

Video Skew Factor and Media Delivery Index (MDI) Scoring

A white paper describing a practical method of simulating video phones for determining the delay (Skew Factor) between video and audio packets in an IP network

Contents

I. Introduction

Since 1998 Ameritec has been a provider of Voice Over Packet test solutions for companies whose business focus is the development of voice over packet networks and network components. The solution Ameritec has adopted for testing packet-switched networks is unique in the market segment and one that has become a standard for most companies. The great challenge for a developer has been to determine how to improve the performance of their product without metrics that highlight the potential sources of the problem. This is why Ameritec's focus for a VoP test product has been based on stimulating the network with stimuli of known characteristics and then reporting results based on objective measurements. The results of these tests provide developers a broad range of measurements that allow them to evaluate the product's performance based on "*real*" numbers. In a white paper authored and published by Ameritec in August 2000, "*Filling the VoID in VoIP Testing*", we provided developers with insight into the correlation of impairments with specific elements of a packet-switched network.

In 2004, a whitepaper was written describing Ameritec's development of an implementation of specific E-Model metrics comprised of a Conductor™ software tool named MQoS™ and bit and block error rate testing as a means of characterizing fax/data throughput in a digital network. Because bit and block error rate testing is not possible on an analog call, a different approach was undertaken to ensure that a fax/data modem call would be recognized by a device under test and consequently, the vocoding scheme used would be switched over to G.711 operation and thereby provide a path for the call to complete. Utilizing these fax/data modem testing methods will allow the user to test end to end fax/data modem call simulation in a high capacity environment without incurring significant development costs.

In 2006, Ameritec again has risen to the challenge by developing a test methodology for determining the impact of network delay on video and audio packets with its Skew Factor test. Using a SIP based network call generator, Ameritec's Skew Factor provides the developer of video telephones or IP infrastructure equipment with the tools needed to determine buffer sizes required to ensure that video and audio packets arrive within a time span that does not diminish the quality of experience of the end user.

II. Why Audio/Visual Synchronization is Critical

Although Audio-Video out of sync, also known as A/V out of sync, A/V sync, lip-sync, etc, has strong personal subjective bias, when the problem is too obvious, say audio comes before the video by 1 second, or lags behind the video by 3 seconds, almost everyone will be disturbed. We know that because the speed of sound in air is much slower than the speed of light, it is natural to observe audio lagging video. The farther away from the observer a sound is made, the longer it takes the sound to reach the observer, while the accompanying visual stimulus will always appear to be instantaneous. We learned about this in high school science, illustrated with the example of a distant hammerer: The farther the observer from the hammerer, the longer the delay between seeing the blow and hearing it. Conversely, it is completely unnatural to hear the sound of an event before we see it happen. We thus have a greater tolerance for audio-lagging video than for video-lagging audio. This is unfortunate, as we also know that the delays imposed by digital systems on video are typically greater than those imposed on audio, and this results in a tendency for audio signals to precede associated video signals.

Studies in the 1940s by Bell Laboratories concluded that when audio led video by more than 35 ms or lagged video by more than 100 ms, audio/video out of sync will be detected. In a more recent document, the 1998 ITU-R BT.1359-1, average audio/video out of sync **detectability** threshold for ordinary people was set at about 45 ms for audio leading video and 125 ms for audio lagging video, and audio/video out of sync **acceptability** thresholds – meaning passing that people will reject the video, are set at 90 ms for audio leading video and 185 ms for audio lagging video.

III. Characteristics of a Videophone Call

Using SIP as the medium for IP video traffic, a SIP call is initiated for a videophone whereby the SDP in the INVITE and 200 OK messages must contain a line that describes the video stream. The video media line will describe the RTP port number and the video profile that will be used. Audio and video data are sent in both directions in separate RTP streams. Each media stream actually uses two ports; the RTP port for data and the RTPC port for control messages. See Figure 1 for more detailed call flow diagram.

The two media streams can take different paths through the network which may result in one of the streams having a longer transit time than the other. If this happens a “lip sync” issue may be seen on playback unless an appropriately sized jitter buffer is used.

The Skew Factor test is designed to measure the difference in delay between the audio and video streams in order to characterize the network and help determine the required jitter buffer size.

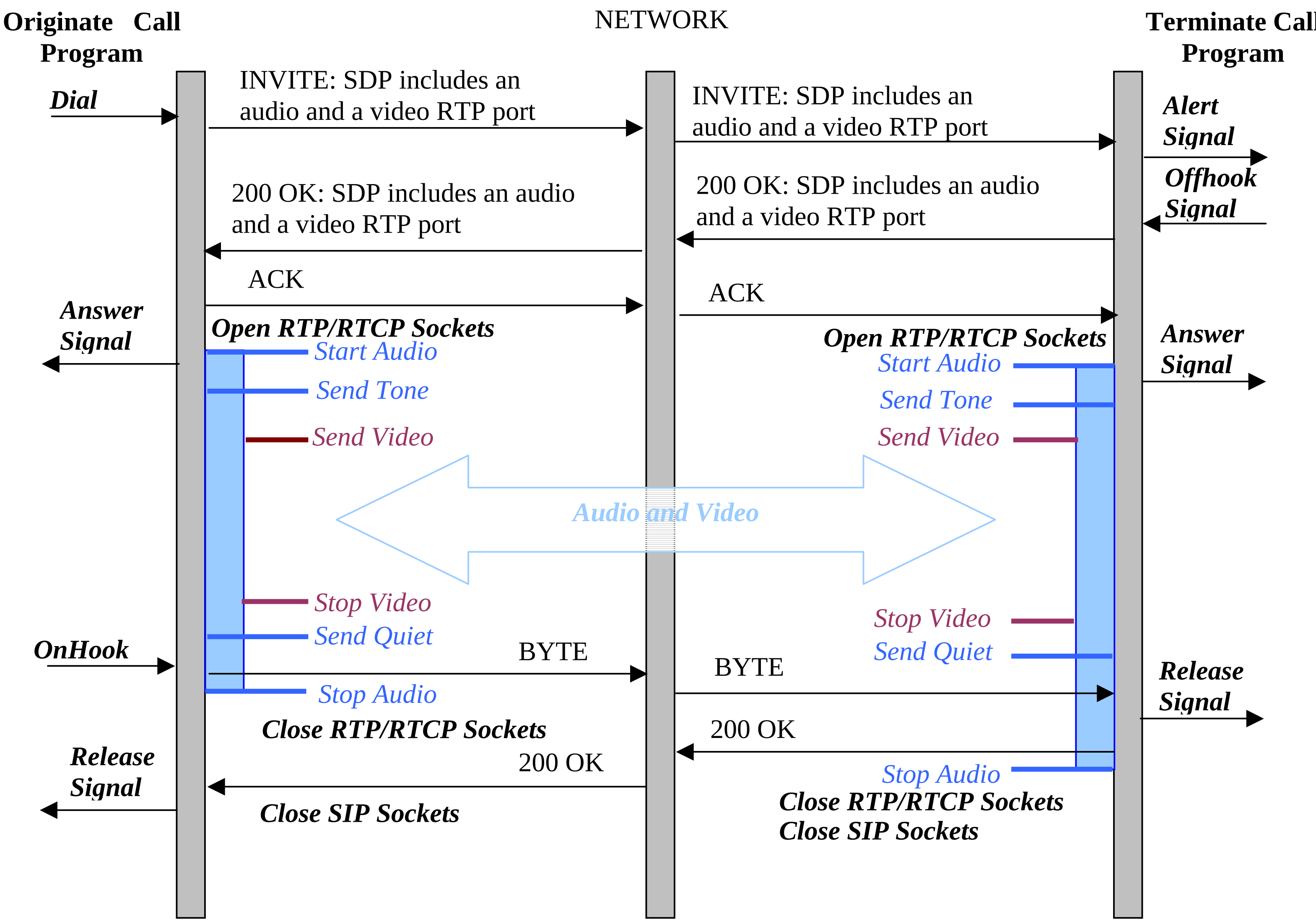


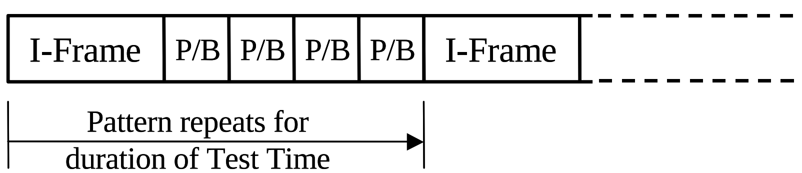
Figure 1
Basic Call Flow

IV. Video Stream Characteristics

There are two types of frames in a video call: I-Frames and P/B-Frames. I-Frames are larger and contain an entire screen shot. P/B-Frames only contain update information and require the previous I-Frame to create a new screen. Therefore, a lost or missing I-Frame is more significant than a lost P/B-Frame.

I-Frames also occur less frequently where one I-Frame will be followed by several P/B-Frames. I-Frames can be larger than the maximum allowed UDP packet size (1500 bytes), so I-Frames will be sent in multiple UDP packets. Any lost I-Frame packet will cause the I-Frame to be counted as lost. P/B-Frames are typically smaller and fit in one UDP packet.

For simulating videophone traffic, the Ameritec Fortissimo product will therefore utilize a repeat cycle of 5 whereby, 4 P/B Frames will follow one I-Frame.



Notes:
- I-Frame can be 1 of more packets (see table for the number of packets in the I-Frame)
- P/B frame is 1 packet
- Maximum packet size is 1500 bytes

V. Skew Factor Calculation

A Skew Factor (SF) calculation is performed on each packet received by evaluating the timestamps of the audio and video packets. The Skew Factor (SF) is the variance between the arrival times of the audio and video packets measured in milliseconds. A histogram containing 5 bins is maintained along with the Maximum and Average SF and the maximum lost I-Frame count. Lost packets will be ignored and do not affect the skew results (lost packets will be counted as lost).

The Fortissimo product will keep statistics on the Maximum and Average Skew Factor measured during the call for both leading and lagging skew. It will also count the number of Skew Factor values within seven different ranges to generate a histogram.

These seven bins include a center bin and three leading bins and three lagging bins. The user can select the bin ranges from a value of 1ms to 8000 ms.

Max Leading SkewFactor (ms)	Avg Leading SkewFactor (ms)	Max Lagging SkewFactor (ms)	Avg Lagging SkewFactor (ms)	Leading SkewFactor Bin 3 (cnt)	Leading SkewFactor Bin 2 (cnt)	Leading SkewFactor Bin 1 (cnt)	SkewFactor Bin 0 (cnt)	Lagging SkewFactor Bin 1 (cnt)	Lagging SkewFactor Bin 2 (cnt)	Lagging SkewFactor Bin 3 (cnt)
13	1.1	13	1.1	75	4671	67021	1371561	110853	4643	122
6	0.4	7	0.7	0	1	4331	1072290	32000	11	0
13	0.7	13	0.9	75	4672	71352	2443851	142853	4654	122

Figure 2. Sample Skew Factor Report

VI. Media Delivery Index (MDI) RFC 4445

The Media Delivery Index is a measurement that can be used as a quality indicator for video, streaming media or information that is sensitive to delay or packet loss. It is comprised of two elements, the DF (Delay Factor) and the MLR (Media Loss Rate). The Delay factor is the maximum difference between the arrival of a packet and its playback. It indicates the length of time a packet stream must be buffered to prevent packet loss. A perfect DF score is the packet size. Ameritec provides selections of 10, 20 or 30ms for the audio stream and 33 or 66ms for the video stream depending on the Frame Rate. The Media Loss Rate is the count of lost or out-of-order packets over a time interval. A perfect Media Loss Rate is 0% packet loss.

Both the Delay Factor and Media Loss Rate measurement interval is one second and is repeated for the duration of the call. These MDI measurements are calculated independently for the audio and the video streams.

Audio MDI DF Max (ms)	Audio MDI DF Avg (ms)	Audio MDI MLR Max (%)	Audio MDI MLR Avg (%)	Audio Packet Loss Count	Video MDI DF Max (ms)	Video MDI DF Avg (ms)	Video MDI MLR Max (%)	Video MDI MLR Avg (%)	Video Frame Loss Count	I-Frame Loss Count
33	24	0	0	0	48	36.1	0	0	0	0
26	21.8	0	0	0	38	34.1	0	0	0	0
33	22.9	0	0	0	48	35.1	0	0	0	0

Figure 3. Sample MDI Report

VII. Video Data and Bandwidth Settings

The AVDATA RATE parameter controls the bandwidth of the video stream, video frame rate and thus video frame size. The relationship between the different AVDATA RATE settings and video frame size is shown in the tables to follow. Default AVDATA RATE setting is 9 (shaded), a 256Kbps video rate and 30 fps frame rate. The default RTP packet size is 20ms.

Table Definitions:

1. AVRate – defines the traffic rate (128 Kbps to 768 Kbps) and the associated frame rate (15 frames per second or 30 frames per second)
2. The Psize defines the number of utilized bytes in each packet
3. The Isize defines the number of bytes utilized in each packet and the IFRM defines the number of packets in the I-Frame.

The default call duration (TEST TIME) is 30 seconds. The MDI measurement requires a minimum duration TEST TIME of 1 second (defined in the RFC). The application will support TEST TIME as short as 1 second, but the instantaneous load on the network could distort the jitter at the beginning of the test period and cause unreliable measurements. A minimum TEST TIME of 10 seconds is suggested. A more typical TEST TIME would be several minutes.

For RTP Packet Size settings of 20ms and 30ms (selected from the Summary Editor Traffic Port tabs in Conductor) all 192 Fortissimo IP channels may be configured for the Skew Factor/Media Delivery Index test.

When the RTP Packet is set at 10ms, the maximum channel utilization is reduced by half to 96 channels (48 channels per side) and the test must be assigned to alternating channels (all odd or all even numbered channels).*

G.711, 10ms packet size*

AVRate Selection	Traffic Rate KBit Per Sec	Frame Rate Frames Per Sec	P/B Frame Psize (BYTES)	I-Frame Isize (BYTES)	IFRM Count (# of packets)
0	128 Kbps	15	108	432	1
1	192 Kbps	15	392	980	2
2	256 Kbps	15	620	1240	3
3	320 Kbps	15	806	1410	4
4	384 Kbps	15	960	1280	6
5	448 Kbps	15	1092	1404	7
6	512 Kbps	15	1204	1336	9
7	128 Kbps	30	86	86	1
8	192 Kbps	30	294	588	1
9	256 Kbps	30	442	1326	1
10	320 Kbps	30	554	1108	2
11	384 Kbps	30	640	1066	3
12	448 Kbps	30	710	1420	3
13	512 Kbps	30	766	1340	4
14	576 Kbps	30	812	1298	5
15	640 Kbps	30	924	1232	6
16	704 Kbps	30	1036	1380	6
17	768 Kbps	30	1146	1308	7

G.711, 20ms packet size

AVRate Selection	Traffic Rate KBit Per Sec	Frame Rate Frames Per Sec	P/B Frame Psize (BYTES)	I-Frame Isize (BYTES)	IFRM Count (# of packets)
0	128 Kbps	15	220	880	1
1	192 Kbps	15	492	1230	2
2	256 Kbps	15	710	1420	3
3	320 Kbps	15	886	1240	5
4	384 Kbps	15	1036	1380	6
5	448 Kbps	15	1160	1304	8
6	512 Kbps	15	1268	1408	9
7	128 Kbps	30	176	176	1
8	192 Kbps	30	368	736	1
9	256 Kbps	30	506	758	2
10	320 Kbps	30	610	1220	2
11	384 Kbps	30	690	1150	3
12	448 Kbps	30	754	1130	4
13	512 Kbps	30	806	1410	4
14	576 Kbps	30	850	1360	5
15	640 Kbps	30	962	1282	6
16	704 Kbps	30	1072	1428	6
17	768 Kbps	30	1184	1352	7

G.711, 30ms packet size

AVRate Selection	Traffic Rate KBit Per Sec	Frame Rate Frames Per Sec	P/B Frame Psize (BYTES)	I-Frame Isize (BYTES)	IFRM Count (# of packets)
0	128 Kbps	15	258	1032	1
1	192 Kbps	15	526	1314	2
2	256 Kbps	15	740	1110	4
3	320 Kbps	15	914	1278	5
4	384 Kbps	15	1060	1412	6
5	448 Kbps	15	1184	1332	8
6	512 Kbps	15	1290	1432	9
7	128 Kbps	30	206	206	1
8	192 Kbps	30	394	788	1
9	256 Kbps	30	528	792	2
10	320 Kbps	30	628	1256	2
11	384 Kbps	30	706	1176	3
12	448 Kbps	30	770	1154	4
13	512 Kbps	30	820	1434	4
14	576 Kbps	30	862	1378	5
15	640 Kbps	30	974	1298	6
16	704 Kbps	30	1086	1448	6
17	768 Kbps	30	1196	1366	7

G.729, 10ms packet size*

AVRate Selection	Traffic Rate KBit Per Sec	Frame Rate Frames Per Sec	P/B Frame Psize (BYTES)	I-Frame Isize (BYTES)	IFRM Count (# of packets)
0	128 Kbps	15	400	800	2
1	192 Kbps	15	650	1082	3
2	256 Kbps	15	852	1278	4
3	320 Kbps	15	1018	1424	5
4	384 Kbps	15	1154	1318	7
5	448 Kbps	15	1270	1428	8
6	512 Kbps	15	1370	1370	10
7	128 Kbps	30	320	320	1
8	192 Kbps	30	488	976	1
9	256 Kbps	30	608	912	2
10	320 Kbps	30	700	1400	2
11	384 Kbps	30	770	1282	3
12	448 Kbps	30	826	1238	4
13	512 Kbps	30	872	1220	5
14	576 Kbps	30	910	1456	5
15	640 Kbps	30	1022	1362	6
16	704 Kbps	30	1132	1292	7
17	768 Kbps	30	1244	1420	7

G.729, 20ms packet size

AVRate Selection	Traffic Rate KBit Per Sec	Frame Rate Frames Per Sec	P/B Frame Psize (BYTES)	I-Frame Isize (BYTES)	IFRM Count (# of packets)
0	128 Kbps	15	512	1024	2
1	192 Kbps	15	750	1250	3
2	256 Kbps	15	942	1412	4
3	320 Kbps	15	1100	1282	6
4	384 Kbps	15	1230	1404	7
5	448 Kbps	15	1340	1340	9
6	512 Kbps	15	1434	1434	10
7	128 Kbps	30	410	410	1
8	192 Kbps	30	562	1124	1
9	256 Kbps	30	672	1008	2
10	320 Kbps	30	756	1008	3
11	384 Kbps	30	820	1366	3
12	448 Kbps	30	870	1304	4
13	512 Kbps	30	912	1276	5
14	576 Kbps	30	948	1264	6
15	640 Kbps	30	1058	1410	6
16	704 Kbps	30	1170	1336	7
17	768 Kbps	30	1280	1280	8

G.729, 30ms packet size

AVRate Selection	Traffic Rate KBit Per Sec	Frame Rate Frames Per Sec	P/B Frame Psize (BYTES)	I-Frame Isize (BYTES)	IFRM Count (# of packets)
0	128 Kbps	15	550	1100	2
1	192 Kbps	15	784	1306	3
2	256 Kbps	15	972	1458	4
3	320 Kbps	15	1126	1312	6
4	384 Kbps	15	1254	1432	7
5	448 Kbps	15	1364	1364	9
6	512 Kbps	15	1456	1456	10
7	128 Kbps	30	440	440	1
8	192 Kbps	30	588	1176	1
9	256 Kbps	30	694	1040	2
10	320 Kbps	30	774	1032	3
11	384 Kbps	30	836	1392	3
12	448 Kbps	30	886	1328	4
13	512 Kbps	30	926	1296	5
14	576 Kbps	30	960	1280	6
15	640 Kbps	30	1072	1428	6
16	704 Kbps	30	1182	1350	7
17	768 Kbps	30	1294	1294	8

VIII. Summary

What Ameritec Call Generators offer are performance, accuracy, capability and scalability. Ameritec Call Generators are designed to make precise impairment measurements, and now with videophone simulation, the developer has a complete testing suite to fit his or her call generation needs. All the functions are done simultaneously on every channel in the call generator without impacting the performance of the test equipment – a single tool that is capable of satisfying the needs of every department in your organization.

Ameritec Call Generators are available in a wide range of physical interfaces -- Analog, T1/E1 CAS, ISDN-PRI, SS7, DS3, SIP, OC3 and STM-1 – and provide interworking between the different interfaces. No longer is your testing confined to one piece of equipment in one location but with a GPS clocking source testing can also be done over a wide area duplicating the characteristics that are representative of actual deployed products.

You can count on Ameritec to provide the tools, the resources and the support to make you a success. If you have questions on testing your application: askzeke@ameritec.com

Sources:
http://www.inventa.com.au/Audio-VideoOutofSync What is Audio/Video Out of Sync?

http://tvtechnology.com/features/tech-corner, A/V Synchronization: How Bad is Bad?