Network Traffic Measurement of MBONE over ATM

Reid van Melle
Carey Williamson
Tim Harrison
Department of Computer Science
University of Saskatchewan

September 24, 1996

Abstract

This paper reports on our measurements of Internet MBONE traffic over an experimental wide area ATM network in Canada called the National Test Network (NTN). In particular, we report measurements for audio and video traffic, as generated by standard MBONE tools. These measurements are part of a workload characterization effort to assess the bandwidth requirements for MBONE traffic, the end-user quality of MBONE traffic, and any potential performance problems for MBONE traffic (over Internet or ATM).

Our measurement results show that the MBONE tools perform very well across the NTN backbone, but that MBONE traffic originating from elsewhere on the Internet is highly disrupted by delay, jitter, and loss effects. The measurements also illustrate the relatively uninteresting constant bit rate nature of audio traffic, and a significant degree of long range dependence in MBONE video traffic.

1 Introduction

The multicast backbone (MBONE) consists of a subset of Internet hosts and routers that are configured with IP multicast capability [4]. The MBONE is used for experimenting with multicast (i.e., multi-site) network applications.

Standard MBONE tools allow MBONE-connected users to do desktop video conferencing between multiple sites, to receive live audio and video feeds from events around the Internet (e.g., space shuttle launches, IETF meetings, conference presentations, seminar series, etc.), and to use shared whiteboard tools in support of collaborative research activities. These tools are available on standard Unix workstations, provided that the workstations (and intermediate routers) have been properly configured to handle IP multicast [2, 3, 13]. These applications provide fairly low quality (e.g., 8 frames per second), low bandwidth (e.g., 200 kilobits per second) audio and video capabilities, as is suitable for most bandwidth-constrained Internet environments.

In the spring of 1996, the CANARIE (Canadian Network for the Advancement of Research, Industry, and Education) program added MBONE capability to Canada's experimental ATM network, called the National Test Network (NTN). This process involved receiving an MBONE feed from the United States, and carrying this feed across the NTN using multicast-capable Cisco routers at selected sites across the country. MBONE traffic is now carried in ATM cell format across this DS-3 (45 Mbps) national backbone, with a maximum bandwidth of 1 Mbps allocated for MBONE traffic.

The addition of MBONE capability to the NTN enables researchers to experiment with the end use of MBONE tools (e.g., distance education course delivery, videoconferencing, computer-supported collaborative work), as well as to experiment with the MBONE protocols themselves, and to conduct network measurements of MBONE traffic. It is the latter activity that is the focus of our work.

The purpose of this paper is to report on our measurements, to date, of MBONE traffic on the NTN ATM network, using the results for workload characterization purposes. Such data will be used to assess the bandwidth requirements and end-user quality of MBONE traffic, and to identify any potential performance problems for MBONE tools, when operating over the Internet or an ATM network.

Our analysis of MBONE traffic proceeds in three stages. In the first stage, we focus on the analysis of outgoing audio traffic (only) from one MBONE-connected SGI Indy workstation at the University of Saskatchewan. This analysis identifies the predominantly constant bit rate nature of the outbound audio traffic, with its distinct framing pattern. The second stage focuses on the analysis of outgoing video traffic (only), which has more interesting traffic characteristics, including a high degree of long range dependence. In the third and final stage, we analyze an incoming combined audio and video stream of a space shuttle interview, originating from the U.S. portion of the Internet. Using complete captures of the ATM cell traffic, cell contents are used to identifiy audio and video portions of the trace, which are then analyzed and compared to the outgoing audio and video, respectively. The analysis of the incoming trace illustrates not only the multiplexing effects of combined audio and video traffic, but also many of the effects (i.e., delay, jitter, loss) encountered by IP packet traffic prior to it reaching the ATM NTN backbone.

The rest of the paper is organized as follows. Section 2 provides some background material on the MBONE and the environment used for our network traffic measurements: the ATM backbone network, the ATM measurement equipment used, and the configuration of the MBONE tools used for the tests. Sections 3, 4, and 5 present the traffic analyses for outgoing audio, outgoing video, and incoming audio/video, respectively. Finally, Section 6 provides a summary of our observations to date, and a brief description of our research plans for the future.

2 Background

2.1 The CANARIE National Test Network

The National Test Network (NTN) is a coast-to-coast experimental ATM network, sponsored in part by the Canadian government CANARIE program. The network, which consists primarily of an east-west backbone network, connects together regional ATM testbeds in several Canadian provinces. The slowest links in the NTN backbone are DS-3 (45 Mbps). The highest speed links, primarily within some of the regional testbeds, are OC-3 (155 Mbps).

The NTN backbone is currently configured as follows: 10 Mbps are reserved for carrying CA*net Internet traffic, 10 Mbps are reserved for distance education course offerings using Motion

JPEG compressed video (or other technologies), and 25 Mbps are for use by Canadian university researchers in high performance computing and high performance networking. MBONE traffic has been allocated 1 Mbps of bandwidth within this portion for research and experimentation. A single VPI/VCI (Virtual Path Identifier/Virtual Channel Identifier) is allocated for carrying all MBONE traffic across the NTN backbone (i.e., between the NTN routers), and additional VPI/VCI's are used to carry MBONE traffic from NTN routers to the individual hosts (i.e., workstations) connected to the NTN.

The only workstation at the University of Saskatchewan that is connected to the MBONE is `joe90.usask.ca`, an SGI Indy, running IRIX 5.3. This workstation has a 174 MHz MIPS R4400/4010 CPU/FPU with a 1 MB secondary cache and 64 MB RAM. This workstation also has an OC-3 (155 Mbps) ATM network interface card (NIC). The precise NIC being used is a Fore GIA-200e, with firmware version 4.0.0, and ForeThought driver version 4.0.1 (1.2). An ATM Permanent Virtual Channel (PVC) is configured so that this workstation can use the NTN directly for ATM traffic, with a maximum outgoing bandwidth of 25 Mbps.

2.2 MBONE Tools and Configurations

Standard MBONE tools such as `vic` (video conferencing), `vat` (voice/audio tool), `nv` (network video), `rat` (real-time audio tool), `sd` (session directory), and `wb` (whiteboard) allow MBONE-connected users to do desktop video conferencing between multiple sites, to receive live audio and video feeds from events around the Internet, and to use shared whiteboard tools in support of collaborative research activities. These tools are available on standard Unix workstations.

The MBONE operates using a UDP/IP protocol stack. UDP (User Datagram Protocol) is a simple datagram-oriented protocol [1, 11, 12]. That is, it is a minimal mechanism protocol, providing no support for flow control, error control, or retransmission of packets. As a result, it requires minimal processing time for transmission and reception of packets, making it suitable for real-time applications. However, it provides no end-to-end guarantee on when (or if) packets get delivered.

MBONE applications communicate using UDP sockets. Every output operation by a process causes one UDP datagram to be sent, which in turn causes one IP datagram to be sent. This means that the size of the IP packets transmitted depends directly on the operation of the audio/video encoder and the block sizes that they work with when writing to the UDP socket.

The audio and video formats used[1] for outgoing traffic can be chosen via the software, but the input format is automatically adjusted to match the incoming stream(s). Since all cells from multiple sources and multiple media arrive on the same VCI, there is some extra RTP protocol overhead required in the UDP payload that the MBONE software uses to distinguish between sessions, and the audio software uses to decide on an input format.

In our tests, the outgoing audio used PCM format, and the software for MBONE video was configured as follows:

- H.261 encoding [6]
- 8 frames per second

[1]The details of how the MBONE tools package audio and video data into UDP packets is quite complicated. For the reader interested in these details, Appendix A of this paper provides a brief description of the common audio and video formats used in the MBONE, as well as a brief description of the Real-Time Protocol (RTP).

- 200 kbps maximum bandwidth

A manual inspection of the contents of the UDP datagrams allowed us to determine that the H.261 encoding was done in the CIF format (see Appendix A). This means that every *picture* consists of 12 GOB's.

2.3 ATM Measurement Device

The network traffic measurements described in this paper were made using a GN Nettest NavTel InterWatch 95000 ATM test set (referred to as the IW95000 in the rest of the paper). This test set was borrowed from TR*Labs* in Winnipeg for one month during the summer of 1996, and used for data collection at TR*Labs* in Saskatoon as well as at the University of Saskatchewan.

The IW95000 allows a user to capture up to 64 Megabytes of complete ATM cell information at OC-3 line rate. This capture buffer size works out to be approximately 1 million ATM cells. At 155 Mbps, this buffer size represents approximately 3 seconds of traffic. For traffic streams with lower bit rates (such as MBONE traffic), the capture time can be significantly longer (minutes or hours). There is a timestamp recorded for each captured cell, with a 1 microsecond resolution.

2.4 Workload Traces Captured

A total of eight traces were captured for MBONE traffic: three for outgoing traffic, and five for incoming traffic. Captures were made separately for incoming and outgoing traffic to help separate the effects of audio traffic, video traffic, and network congestion. In many tests, the incoming MBONE traffic from the Internet was often low bandwidth and unreliable due to network problems.

The following captures were completed:

1. audio only output from the UofS
   - van Melle and Harrison talked to each other, with a couple of dead spots when the microphone was covered for some silence
2. video only output from the UofS
   - camera was focused on van Melle and Harrison talking with occasional pans around the office and periods of static picture
3. audio and video output from the UofS
   - camera was focused on van Melle and Harrison talking, with pans around the office
4. lecture from an IETF meeting
   - combined audio and video traffic stream was captured
   - slow video motion and choppy audio
5. shuttle mission interview

- combined audio and video traffic stream was captured
- slow and jerky video motion and sporadic audio

6. shuttle mission camera panning over the surface of the earth
   - video only traffic
   - fairly smooth video sequence but the scene was quite static
7. lecture from another IETF meeting
   - audio only
   - voice was very choppy and not intelligible
8. another space shuttle trace
   - both audio and video traffic
   - video was a fairly static image but there was lots of audio

In this paper, only traces 1, 2, and 5 are described. The remaining traces are available on tape archive for future analysis, if needed.

2.5 Summary

This section has presented background material on the measurement environment used for collecting data on MBONE over ATM traffic. Details were provided on the network used, the MBONE tools used, the ATM test set used, and the traces collected. The remainder of the paper describes the analysis of three such traces.

3 Measurements for Outgoing Audio Traffic

This section presents the measurement results for outgoing MBONE audio traffic from an SGI Indy workstation at the University of Saskatchewan. The analysis looks at general traffic characteristics, audio framing characteristics, and cell-level traffic characteristics.

3.1 Preamble

The audio input for this trace was general conversation and background noise in a graduate student office at the University of Saskatchewan. The sensitivity of the microphone was controllable, and was adjusted once during the trace to see the effect on cell transmissions.

In total, the trace was 657.8 seconds long and contained 150,517 transmitted cells. Calculations from these numbers give an average rate of 228.8 cells per second, and a raw bit rate of 97.0 kbps.

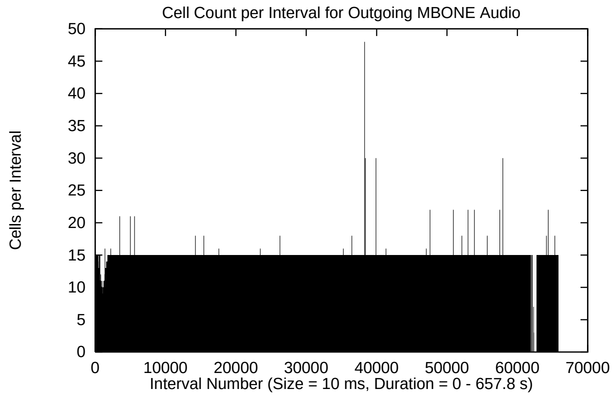
The video software was turned off during this trace, so initially we assumed that no video frames were being sent or received. Upon careful inspection of the trace, however, we found that video frames were being received at a rate of approximately 8 frames per second. All the video frames were very small and contained very little data. It is not known why these frames were being sent. However, it was possible to filter the traffic stream to isolate the audio and video components.

The audio component of the trace contained a total of 117,159 cells giving an average rate of 178.1 cells per second and a raw bit rate of 75.5 kbps. The incoming video contained the remaining 33,355 cells, which translates to 50.7 cells per second and 21.5 kbps. Only the audio portion of this trace is analyzed here. The video component was not analyzed for this trace.

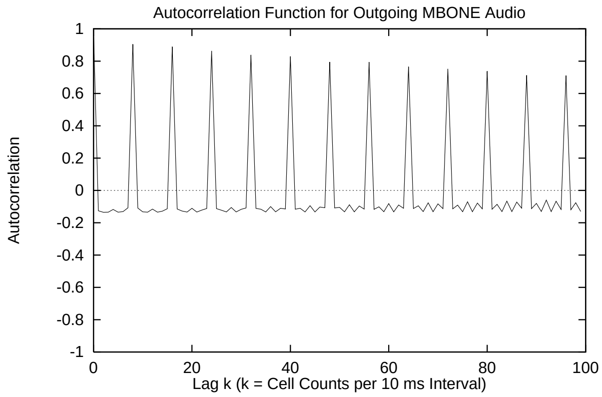
3.2 General Traffic Characteristics

Figure 1 shows our analysis of the outgoing MBONE audio traffic. Figure 1(a) shows the ATM cell counts per 10 ms interval in the trace. The audio traffic is quite predictable and looks very much like constant bit rate (CBR) traffic with 15 cells per interval. Two anomalies occur in the traffic stream: a large spike of 48 cells at approximately 400 seconds and an idle period at 620 seconds. The idle period corresponds to a time when we momentarily decreased the input sensitivity of the microphone to zero. The large spike at 400 seconds is currently unexplained.

Figure 1(b) shows the autocorrelation function for the cell count per interval plot from Figure 1(a). The audio framing format is immediately visible with peaks of positive correlation every 80 ms. This framing format along with the bandwidth used by our audio stream suggests that the audio software was using *pcm4* in order to encode the audio stream (see Appendix A).



(a)

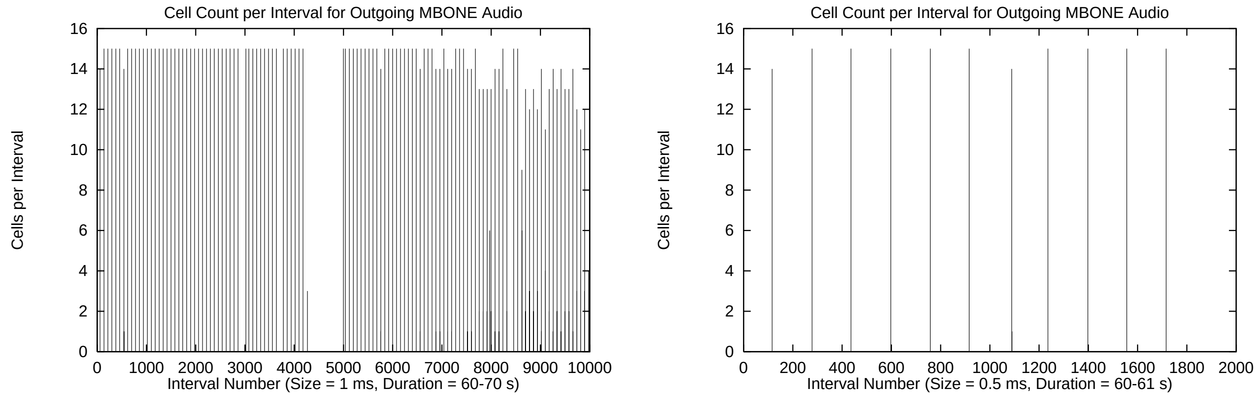


(b)

Figure 1: Traffic Analysis for Outgoing MBONE Audio: (a) Cell Count per 10 ms Interval for Entire Trace; (b) Autocorrelation Function Illustrating 80 ms Audio Framing Structure

Figure 2 zooms in on Figure 1(a) with two different higher resolutions. Figure 2(a) looks at a 10 second portion of the trace (from 60 seconds to 70 seconds) with cell counts per 1 ms interval. The 15 cell transmissions per interval are again evident, though there is an unexplained break of almost one second ocurring at the 64 second mark. This break may have been caused by some interruption at the host computer. Figure 2(b) looks even closer at the region of the trace from 60 to 61 seconds with an interval size of 500 μs. The 80 millisecond framing structure is clearly evident as evenly spaced spikes.

Decoding at higher layers allowed us to determine that the audio information sent every 80 ms almost always consisted of a single 15 cell AAL-5 PDU containing 640 bytes of application data.



(a) (b)

Figure 2: Cell Count per Interval Plots Illustrating Audio Traffic Characteristics: (a) Ten Second Interval of Trace, with an Unexplained Interruption in the Traffic: (b) One Second Interval of Trace, Illustrating Audio Framing Structure

Since every audio sample is 8 bits (one byte), each PDU contains 640 audio samples. Taking 640 samples every 80 ms gives the desired sampling rate of 8000 samples per second, which gives the expected bandwidth of 64 kbps. The difference between the raw bit rate of 75.5 kbps and the data rate of 64 kbps is 11.5 kbps of pure overhead from intermediate protocol layers, including RTP (12 bytes), UDP (8 bytes), IP (20 bytes), AAL-5 (32 bytes in this case: 8 byte trailer and 24 bytes of padding), and ATM (5 byte header per cell $\times$ 15 cells = 75 bytes) layers[2].

3.3 Cell Interarrivals

Our next analysis looks at the cell-level characteristics of the audio trace. The mean cell interarrival time was 5614.6 μs with a standard deviation of 27,644. The gives the same raw bit rate of 75.5 kbps as calculated earlier.

Figure 3(a) shows the cell interarrival time frequency distribution, with a bin size resolution of 100 μs. The large bin size allows us to more easily determine where the peaks occur in the plot. The largest spike, containing 93.2% of all cell interarrival times, occurs at the 50 μs mark (i.e., it encompasses all the interarrival times between 0 and 100 μs). We also see spikes around the 80 ms mark, which is expected based on the audio framing format being used. The largest spikes in this region include an impulse encompassing interarrival times in the 79.7-79.8 ms region (2.97% of all cells) and another impulse encompassing interarrivals in the 79.6-79.7 ms region (1.38% of all cells). These interarrival times are from the end of one audio frame to the beginning of another so the difference between 80 ms and these times represents the time to physically transmit the 15 cell PDU at 25 Mbps.

Figure 3(b) zooms in on the cell interarrival times below 100 μs, with a finer bin size resolution (3 μs). The majority of cells are output at the rate defined by the PVC bandwidth limitation of 25 Mbps on the ATM NIC card (i.e., approximately 20 μs apart). These are all cells sent within an AAL-5 PDU. There were no cells sent at back-to-back OC-3 line rate in this trace.

[2]The math does not work out exactly here because of interruptions in the audio stream, plus some audio frames that use fewer than 15 cells (see Figure 2(b)).

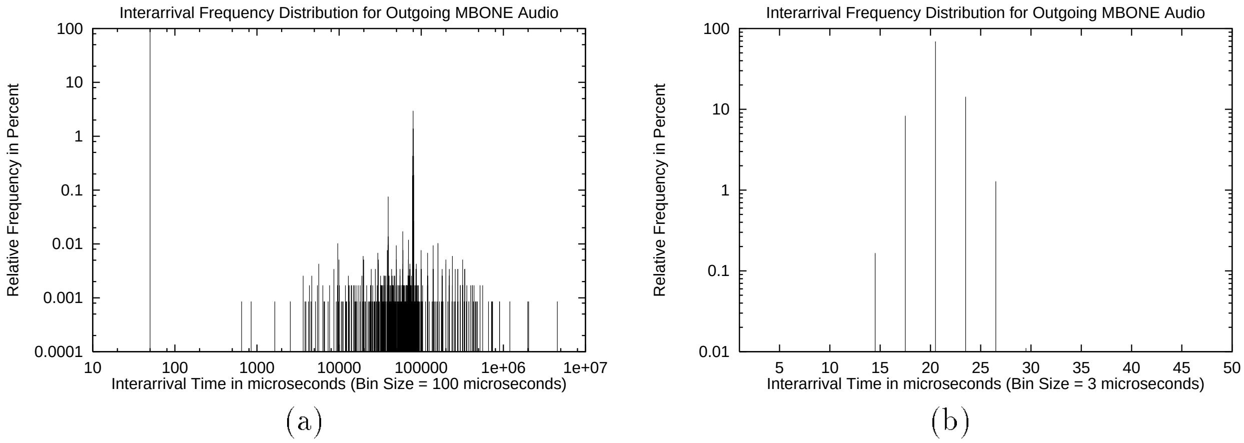


Figure 3: Cell Interarrival Time Frequency Distributions for Outgoing MBONE Audio: (a) Entire Distribution using a Coarse Resolution (100 microseconds); (b) Portion of the Distribution using a Fine Resolution (3 microseconds)

3.4 Summary

This section presented the workload characteristics for outgoing MBONE audio. There were three main observations for this trace. First, the overall traffic stream had an almost constant bit rate profile. Second, a distinct audio framing format was seen, with very strong correlations at periods of 80 milliseconds, though there were occasional unexplained deviations from this behaviour. Third, the cell-level traffic characteristics (e.g., interarrival time distribution) are consistent with periodic UDP packets of approximately 640 bytes being transmitted at a peak rate setting of 25 Mbps.

4 Measurements for Outgoing Video Traffic

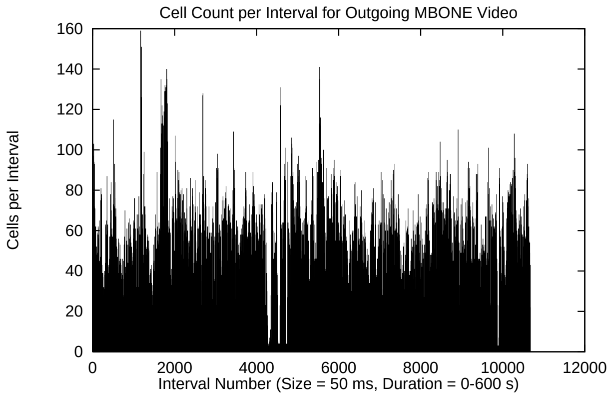
This section presents measurement results for outgoing MBONE video from an SGI Indy workstation at the University of Saskatchewan. The analysis looks at general traffic characteristics, video framing characteristics, long range dependence in video traffic, and cell-level traffic characteristics.

4.1 Preamble

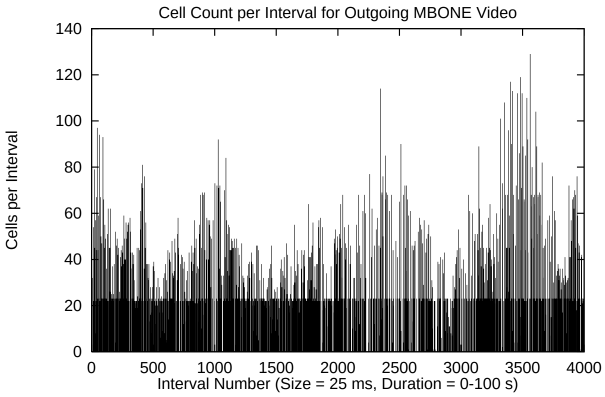
The video input for this trace was general desktop videoconferencing, conversation, and camera panning in a graduate student office at the University of Saskatchewan. Over the course of the 533.8 second trace, a total of 204,526 cells were transmitted, representing an average rate of 383 cells per second and a raw bit rate of 162.5 kbps.

4.2 General Traffic Characteristics

Figure 4 presents the traffic profile for the complete video trace, and also for a 100 second portion of the trace. Figure 4(a) shows cell counts per 50 ms interval over the complete 533.8 second trace. At this scale, the traffic looks very bursty. Figure 4(b) zooms in on the first 100 seconds of the trace and shows cell counts per 25 ms interval. Again, we see the bursty characteristics of the traffic. However, there seems to be a lower limit of approximately one 22 cell packet per 50 ms interval. This background level could correspond to frame updates when there is little new video information.



(a)



(b)

Figure 4: General Traffic Characteristics for Outgoing MBONE Video: (a) Cell Count per 50 ms Interval for Entire Trace (600 seconds); (b) Cell Count per 25 ms Interval for Initial Portion of the Trace (100 Seconds)

Figure 5 takes an even closer look at a period of time from 10 to 20 seconds. The plot shows cell counts per 5 ms interval. The background level of 22 cells per 50 ms interval is seen more clearly here with occasional spikes of 44 to 46 cells. A 22 cell IP packet corresponds to approximately 1 Kbyte worth of data. This could be the size of the data blocks that the encoder creates[3] before writing to the UDP socket for transmission. The taller spikes of 44 cells may occur when the encoder uses two packets in quick succession. Decoding at the IP level allowed us to determine that the largest IP packets sent were 1032 bytes. This corresponds to exactly 1024 bytes of application data once 8 bytes of UDP header overhead are subtracted. Not all IP packets used this 1 Kbyte limit, however, as IP packets varied from 44 to 1032 bytes in size.

4.3 Video Framing

Fortunately, every new H.261 frame and GOB begins with a distinctive bit sequence, based on the RTP header information used to mark frame boundaries. The cells in an IP packet can be inspected for this sequence to see if the IP packet is starting a new frame or GOB. Using this technique, the following observations were made:

- every new frame begins with a new IP packet
- GOB's may be broken across multiple IP packets
- small IP packets (i.e., much less than 1 Kbyte) always contain data to complete the transmission of a frame

[3]Examination of the `vat` source code (`rtp.h`) confirms that RTP is using 1024-byte data blocks.

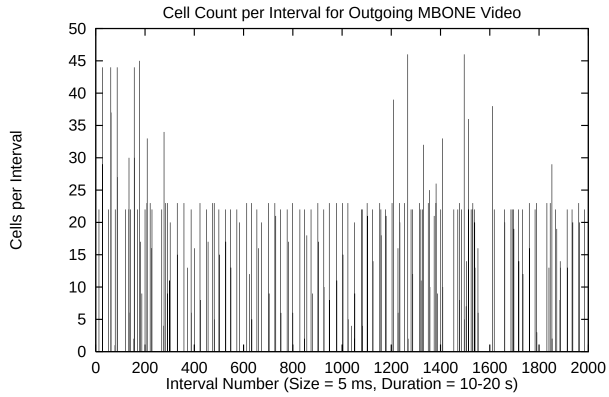


Figure 5: Cell Count per 5 ms Interval for 10 Second Portion of Video Trace

- the delay between frames is highly variable with some frames being transmitted back-to-back and others having considerable delay between them

The same frame boundary information can be used to determine frame sizes. In our trace, the average frame size was 3180 bytes with a standard deviation of 1268. The largest frame was 8696 bytes, while the smallest was 35 bytes. A total of 2835 frames were transmitted over the duration of the trace giving an average frame rate of 5.31 frames per second, which is much lower than expected. Apparently, the bandwidth restriction of 200 kbps did not allow for a frame rate of 8 frames per second.

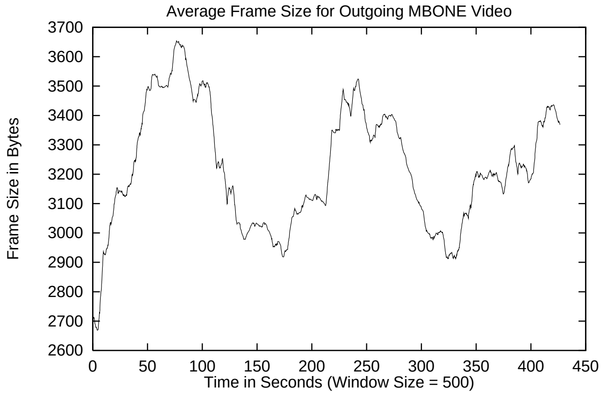
The frame size and the frame rate actually changed quite significantly over the portion of the trace. Figure 6 uses a moving average filter with a sliding window of 500 frames to illustrate this low frequency content of the trace. Figure 6(a) shows the average frame size in bytes at different times in the trace, while Figure 6(b) shows the average frame rate. Clearly, the frame size and the frame rate are inversely related, indicating that the MBONE software adapts its frame rate dynamically based on frame size so as not to exceed its pre-determined bandwidth allocation (200 kbps in this case).

The low frequency content in both plots suggests long range dependence in the video content. Regions with large frames may correspond to periods where the camera was being panned around the office and adjacent frames had little in common.

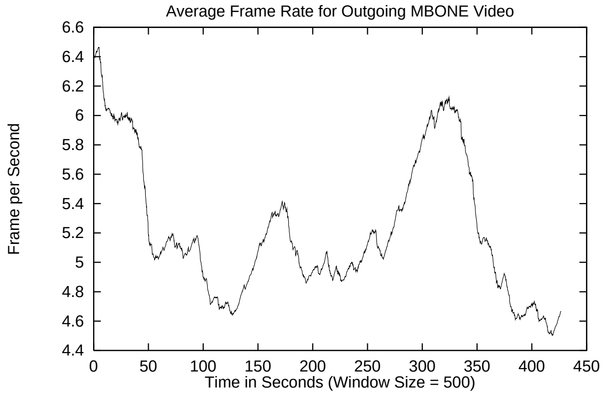
4.4 Long Range Dependence

We now turn our attention to the long range dependence of video traffic hinted at in the foregoing analysis. Figure 7(a) shows the frame sizes over the entire trace. This plot is very similar to the cell count per interval plot of Figure 4(a). The frame sizes clearly vary a great deal depending on the video content that needs to be transferred. Figure 7(b) shows the autocorrelation of the frame sizes up to lags of 50 frames. This correlation structure is strong evidence of long range dependence in the video traffic stream.

A more rigourous way to assess the degree of long range dependence (LRD) and self-similarity in the video trace is with R/S analysis [7]. The Hurst parameter can be read directly off an R/S pox plot as the slope of the line that gives the best fit to the data points. This parameter, which can range from 0.5 to 1.0, measures the degree of self-similarity in the data. If this trace

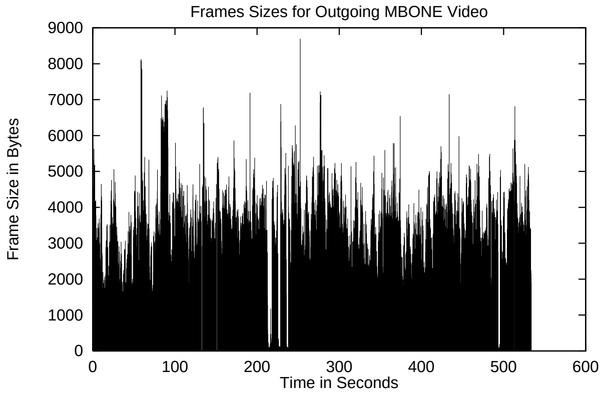


(a)

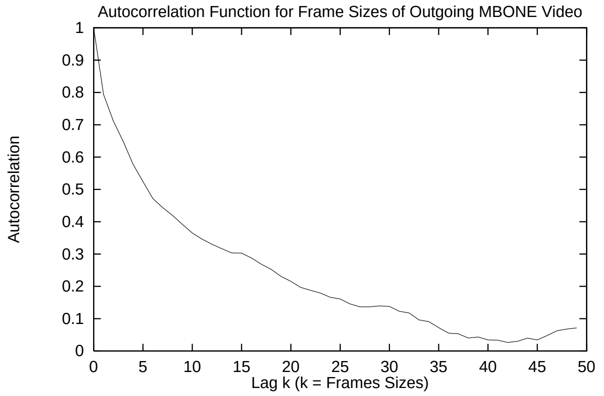


(b)

Figure 6: Variability of Frame Size and Frame Rate in Outgoing MBONE Video: (a) Average Frame Size; (b) Average Frame Rate



(a)

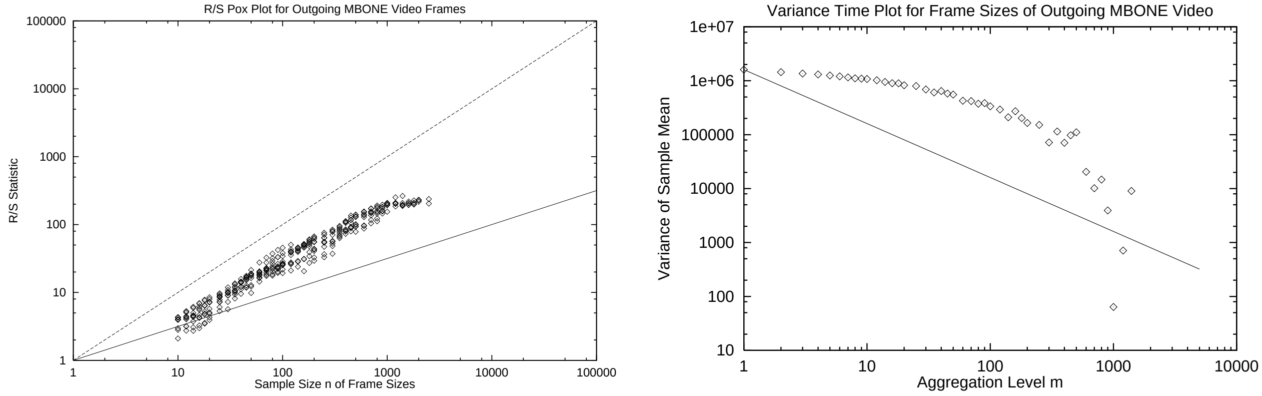


(b)

Figure 7: Analysis of Video Frame Sizes for Outgoing MBONE Video: (a) Time Series Plot of Frame Sizes for Entire Trace; (b) Autocorrelation of Frame Sizes

represents a typical video sequence, then we expect to see a significant degree of self-similarity in the traffic [5].

Figure 8(a) shows the R/S pox plot for the trace based on the frame size data. The slope of the best fitting line gives a value of $H = 0.84$.



(a) (b)

Figure 8: Analysis of Long Range Dependence in Frame Sizes for Outgoing MBONE Video: (a) R/S Pox Plot; (b) Variance-Time Plot

Another method of estimating the Hurst parameter is with the variance-time plot. As the data is aggregated into larger pieces, a slowly decaying variance indicates long range dependence. We saw evidence of this long range dependence earlier as a low frequency component in the frame sizes (see Figure 6(a)).

Figure 8(b) shows the variance-time plot for frame sizes in the MBONE video trace. The variance decays very slowly up to aggregation levels of 200-300 frames. This is significant because there are only 2835 frames in the complete trace. The Hurst parameter is calculated using the rule that $H = 1 - \beta/2$ where β is the (absolute value of the) slope of the line fit to the points in the flat portion of the plot. This approach gives a Hurst parameter estimate of $H = 0.88$.

4.5 Cell Interarrival Times

We conclude our analysis of the video trace with a look at cell-level traffic characteristics. The mean cell interarrival time for the complete trace was 2609.7 microseconds with a standard deviation of 18,698. This gives a raw bit rate of 162.5 kbps and a slightly slower rate once the overhead from the UDP and IP protocol layers is removed.

The cell interarrival time frequency distribution is shown in Figure 9, for two different resolutions. Figure 9(a) uses a bin size of 100 μs, which highlights the peaks better, while Figure 9(b) uses a bin size of 3 μs, which shows higher resolution data.

There are several peaks of interest in these plots. The majority of cells are sent within 100 μs of the last cell. The largest peak occurs for interarrival times of 20 μs and contains 70.6% of the total cells. This is due to the bandwidth limit of 25 Mbps imposed on the ATM NIC card Interestingly, a few cells are transmitted back-to-back at OC-3 line rate in this trace.

A second dominant cluster appears at interarrival times around the 2000 to 4000 μs mark. This cluster of times is the delay between adjacent IP packets that belong to the same frame, and may correspond to the processing time for the software to encode a kilobyte of data. Our speculation concerning the relationship between the digital camera and encoding software is that the encoder first receives a digital frame from the camera, then processes the frame using the H.261 standard, and then writes data to the UDP socket whenever a kilobyte of data is available (or the frame has been completely processed). This guesswork fits the plots in Figure 9, assuming 2 to 4 milliseconds for the encoding process.

Another large cluster appears over the 60 to 100 ms region. These interarrival times correspond to the delay between adjacent frames. These times are spread over a large region due to the high variability in the frame sizes and the frame rate (Figure 6).

A fourth peak is evident in the 200 μs to 300 μs region in Figure 9(b). This peak is probably the time between back-to-back UDP packets, and represents the software and hardware interrupt times necessary to ship application data through the kernel and lower layer protocols.

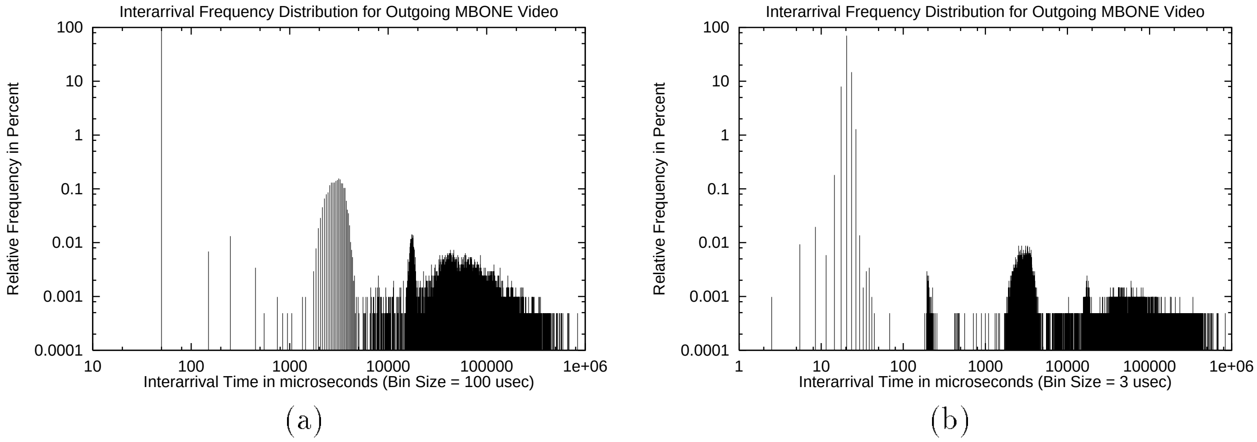


Figure 9: Cell Interarrival Time Distribution for Outgoing MBONE Video: (a) Entire Distribution using a Coarse Resolution (100 microseconds); (b) Entire Distribution using a Fine Resolution (3 microseconds)

4.6 Summary

This section presented the workload characteristics for outgoing video traffic for MBONE over ATM. There were three main observations for this trace. First, the overall traffic stream was highly bursty. Second, the video traffic showed strong evidence of long range dependence, both in its low frequency content, as well as its relatively high Hurst parameter values, indicating self-similarity. Third, the cell-level traffic characteristics (e.g., interarrival time distribution) are consistent with UDP packets of size 1024 bytes, transmitted at a peak rate of 25 Mbps from the ATM NIC on the SGI workstation. Several other dominant peaks in the cell interarrival time distribution are likely due to the processing times for UDP packets, IP packets, and video frames.

5 Measurements for Incoming Audio and Video Traffic

This section presents measurement results for incoming MBONE audio and video traffic, originating from the Internet, and received by an SGI Indy workstation at the University of Saskatchewan. The analysis looks at general traffic characteristics, framing characteristics, long range dependence in the traffic, and cell-level traffic characteristics.

5.1 Preamble

This capture contains both audio and video traffic on the same VPI/VCI. Though more difficult to analyze, this trace is a better representation of real traffic on the MBONE. This trace also contains network effects such as packet loss, delay, and jitter, since the traffic stream originates from the U.S. portion of the Internet. While it is not known what software and encoding format was used at the sending end, these configuration options can be deduced from the trace itself.

5.2 General Traffic Characteristics

The complete trace was 568.4 seconds in length and contained a total of 125,898 cells. This yields a raw bit rate of 93.9 kbps. The trace was terminated when the incoming traffic ceased due to technical difficulties within the MBONE. This is easily seen in Figure 10(a), which shows cell counts per 100 ms interval over the complete trace. The bursty nature of the traffic is immediately visible and a sudden drop in bandwidth occurs just before the 500 second mark. Figure 10(b) zooms in on a 10 second interval of the trace from 10 seconds to 20 seconds, with cell counts per 1 ms interval. We see what looks like a very regular framing pattern, which will be analyzed further in later sections.

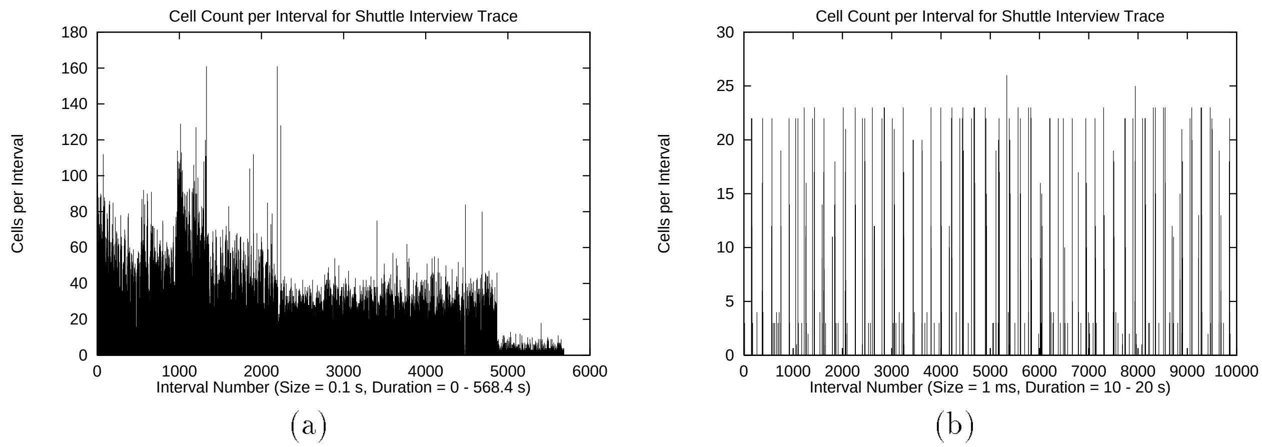
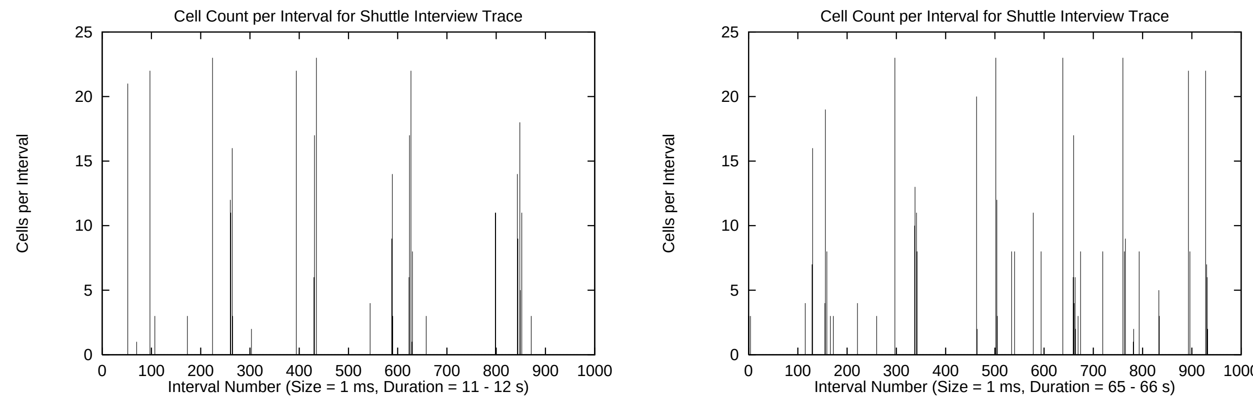


Figure 10: General Traffic Characteristics for Incoming Audio and Video for Shuttle Interview Trace: (a) Cell Count per 100 ms Interval for Entire Trace; (b) Cell Count per 1 ms Interval for Ten Second Portion of Trace (10-20 seconds)

Figure 11(a) zooms in even closer on a one second interval from 11 to 12 seconds, with cell counts per 1 ms interval. Here, individual clusters of cells become visible. Five main clusters appear over the one second interval, with each cluster being composed of several spikes. The largest spikes in this plot contain one complete AAL-5 PDU. As seen in other traces, the largest AAL-5 PDU's contain 1 kilobyte of application data.

The traffic pattern in Figure 11(a) is not always present, however. Figure 11(b) shows a different one second interval from 65 to 66 seconds, which shows between seven and eight bursts. Clearly, the framing behaviour in the traffic changes throughout the trace.



(a) (b)

Figure 11: Cell Counts per 1 ms Interval for One Second Portions of Shuttle Interview Trace: (a) Portion from 11 to 12 Seconds; (b) Portion from 65 to 66 Seconds

5.3 Framing

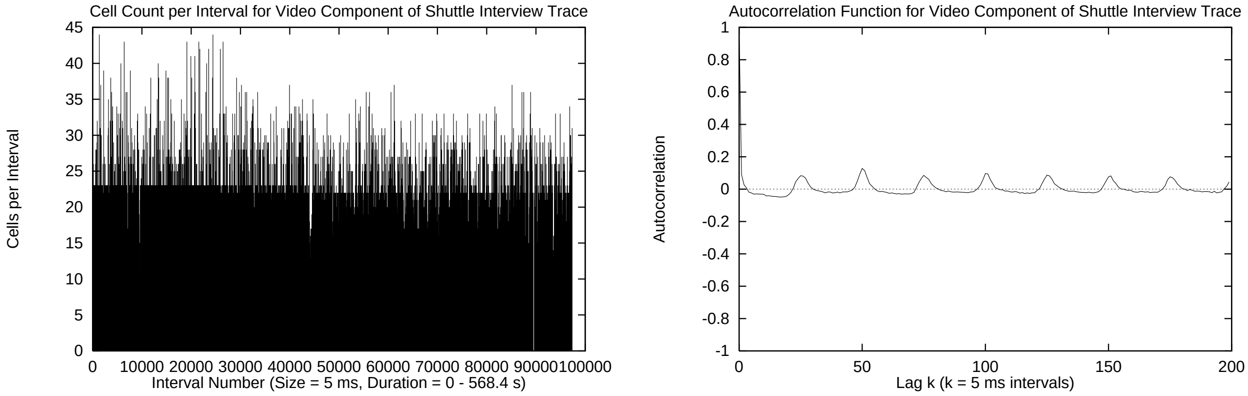
Separating the audio and video traffic (by decoding the AAL-5 PDU's) makes it easier to study the framing structure in the traffic. Figure 12(a) shows the cell counts per 5 ms interval for video traffic only. Figure 12(b) shows the autocorrelation function of this cell arrival process. Figure 13(a) shows the cell counts per 5 ms interval for audio traffic only, with the corresponding autocorrelation function in Figure 13(b).

During the trace capture, the video quality was perceived to be fairly constant. The video traffic contained 106,283 cells, giving the video stream a raw bit rate of 79 kbps, and a data rate somewhere close to 64 kbps after removing overhead. Figure 12 shows that the video framing structure is evident at 8 frames per second.

The audio quality, however, was quite sporadic during the trace capture, and the traffic analysis in Figure 13 confirms this sporadic behavior. The audio framing structure shows positive, but very weak, correlations at lags of 40, 80, 120, ... ms suggesting an encoding format of *pcm2*. The strong audio framing structure that we saw in Section 3 for the outgoing audio stream is not present here due to breaks in the audio feed. The audio traffic contained only 5224 cells contributing a raw bit rate of 3.9 kbps. Whatever audio encoding format was used, a great deal of loss[4] was experienced over the network or at the source.

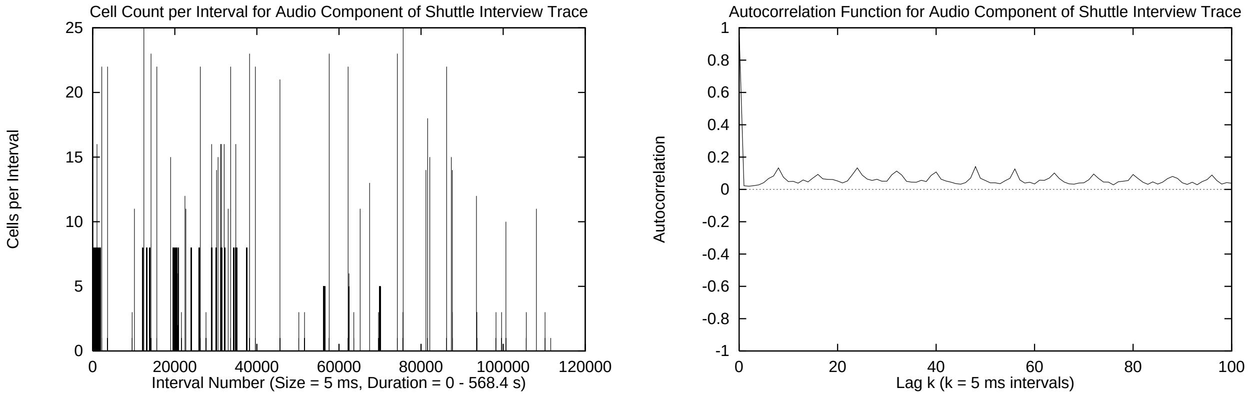
Clearly, the much higher bandwidth video will dominate the characteristics of the combined traffic stream anyway. We thus focus the rest of our analysis on the video traffic only.

[4]We have not yet analyzed the RTP sequence numbers to assess the loss quantitatively, but will likely do so in the near future.



(a) (b)

Figure 12: Video Traffic Analysis for Shuttle Interview Trace: (a) Cell Counts per 5 ms Interval; (b) Autocorrelation of Cell Counts per 5 ms Interval

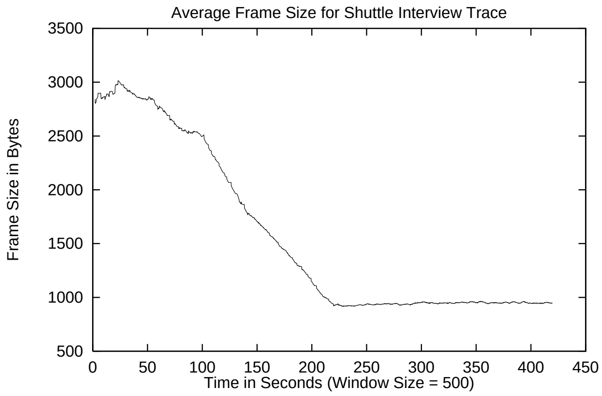


(a) (b)

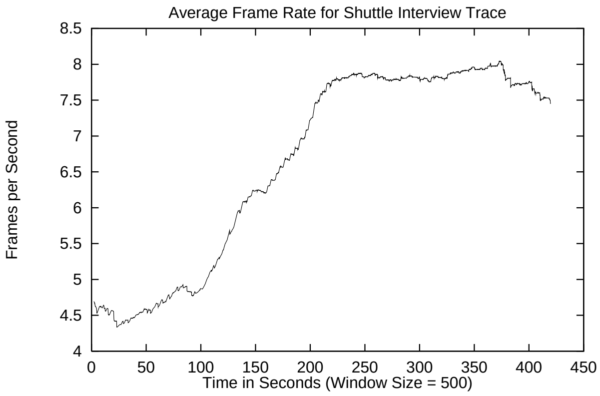
Figure 13: Audio Traffic Analysis for Shuttle Interview Trace: (a) Cell Counts per 5 ms Interval; (b) Autocorrelation of Cell Counts per 5 ms Interval

Figure 14 shows the average frame sizes (Figure 14(a)) and the average frame rate (Figure 14(b)) for the entire trace, averaged with a sliding window of 500 frames. Over the course of the whole trace, the average frame size was 1646 bytes with a standard deviation of 1789 bytes. The largest frame contained 25,876 bytes while the smallest was 40 bytes.

Figure 14 shows that as the average frame size decreases, the frame rate increases, indicating a traffic shaping mechanism in the encoding software (as observed in Section 4). The average frame rate climbs towards a maximum of 8 frames per second and remains there for a lengthy period. This behaviour is consistent with the autocorrelation function in Figure 12(b), which also indicates a framing structure of 8 frames per second. The low frequency content evident in Figure 14(a) and Figure 14(b) again suggests long range dependence in the video traffic.



(a)



(b)

Figure 14: Variability of Frame Size and Frame Rate in Shuttle Interview Trace: (a) Average Frame Size; (b) Average Frame Rate

5.4 Long Range Dependence

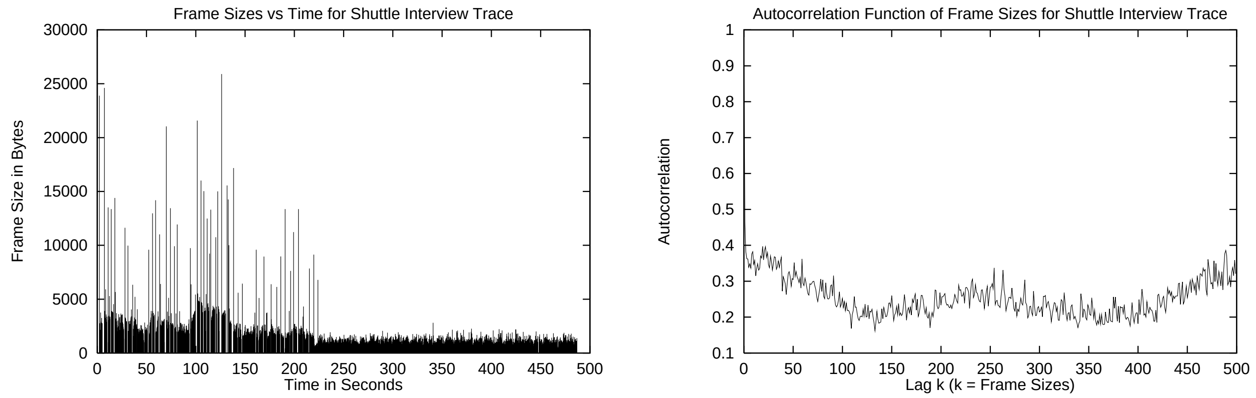
Figure 15(a) shows the video frame sizes plotted as a function of time. There are some very large frames transmitted in the first half of the trace, and some consistently smaller ones transmitted in the last half of the trace. Figure 15(b) shows the autocorrelation function of these frames sizes. Even at very large lags, the frame sizes show significant correlation, indicating long range dependence in the video sequence.

Figure 16 shows the R/S pox plot for frame sizes. The Hurst parameter is estimated for video frame sizes over the complete trace. The slope of this plot indicates some degree of self-similarity[5] in the video sequence (e.g., $H = 0.78$).

5.5 Cell Interarrivals

We close our traffic analysis with a look at cell interarrival time characteristics for the incoming audio and video trace. Note here that the MBONE feed is being carried across Canada on a 45 Mbps DS-3 physical link, which gives a minimum interarrival time for back-to-back cells of approximately 8 μs. However, once the MBONE traffic reaches the University of Saskatchewan,

[5]Note, however, that the assumption that the process in Figure 10(a) is stationary is questionable.



(a) (b)

Figure 15: Analysis of Video Frame Sizes for Shuttle Interview Trace: (a) Time Series Plot of Frame Sizes for Entire Trace; (b) Autocorrelation of Frame Sizes

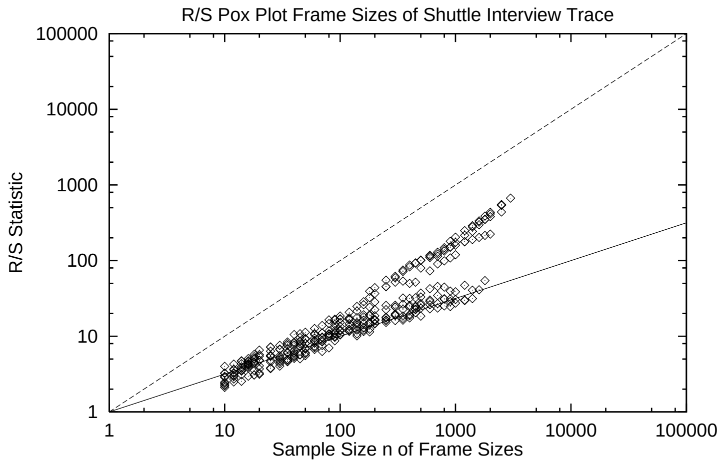
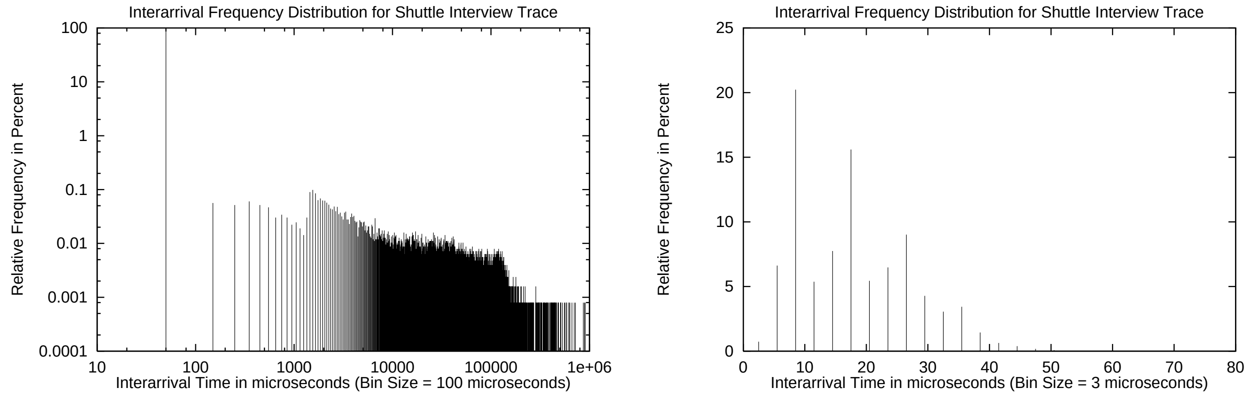


Figure 16: R/S Pox Plot for Video Frame Sizes for Shuttle Interview Trace

it is delivered over a 155 Mbps OC-3 link, which allows for back-to-back cells separated by 3 μs. The mean cell interarrival time for the complete trace was 4515 μs with a standard deviation of 22,934. The largest gap was a period of silence just over one second in duration and the shortest interarrivals times were for back-to-back cells at OC-3 line rate.

Figure 17(a) shows the cell interarrival time frequency distribution using a bin size of 100 μs. The vast majority (over 90%) of cells arrive within 100 μs of each other as indicated by the spike at 50 μs, which falls in the middle of the first bin. The interleaving of audio, video, and other MBONE cells results in a blur of larger interarrival times, without any clear indication of framing.

Figure 17(b) shows the range of interarrival times up to 70 μs with a bin size of 3 μs. The largest spike is at 8 μs, which corresponds to back-to-back cells over the 45 Mbps DS-3 link, while the next two largest spikes indicating MBONE cells separated by one and two idle cells, respectively, at DS-3 rate. We also see some back-to-back cells at OC-3 line rate indicating some buffering within the router or switches en route to the University of Saskatchewan. The wider spread of interarrival times in Figure 17(b) are probably due to delay variation in the network and switches across the MBONE.



(a) (b)

Figure 17: Cell Interarrival Time Frequency Distributions for Shuttle Interview Trace: (a) Entire Distribution using a Coarse Resolution (100 microseconds); (b) Portion of Distribution using a Fine Resolution (3 microseconds)

5.6 Summary

This section presented the workload characteristics for incoming audio and video traffic for MBONE over ATM, with the traffic streams originating from elsewhere on the Internet. There were three main observations for this trace. First, the overall traffic stream was highly bursty. Second, the audio and video framing in the trace was much less distinct than for the outgoing audio and video traces. The video frame rate is highly variable during the course of the trace, and the audio traffic stream in particular is highly disrupted by packet delay, jitter, and loss effects in the Internet. Third, the video traffic showed some evidence of long range dependence, both in its low frequency content, as well as its Hurst parameter values.

6 Summary and Conclusions

Our efforts to date have proven to be an interesting (and hopefully useful) exercise in network traffic measurement and workload characterization for MBONE over ATM traffic. Our ATM network traces illustrate the general characteristics of such audio and video traffic, and also demonstrate the process of taking cell-level measurements and decoding them up the protocol stack to understand the traffic characteristics of interest.

The main observations that we draw from our work are:

- audio traffic on the MBONE is so CBR-like that it is uninteresting to analyze
- video traffic on the MBONE has a significant degree of long range dependence
- the framing structure of outgoing audio and video traffic is very distinct, at least on the ATM portion of its travels
- the framing structure of incoming audio and video traffic is very weak, due to the (not surprising) Internet effects of packet loss, delay, and delay variation
- the MBONE tools (as configured in our experiments) behave quite well, producing traffic streams with very low bandwidth, and posing no threat to the available bandwidth on the NTN ATM backbone (or the Internet)
- the MBONE tools are surprisingly robust, being able to adapt and provide reasonable audio and video delivery across the harsh environment of the Internet

The CANARIE Multicast Test Project continues to experiment with MBONE over ATM. The MBONE has been used for several videoconference meetings (e.g., seven sites) across the NTN, for delivery of several conference talks (e.g., INET'96) to the Canadian research community, and for delivery of seminars (e.g., at UBC) to multiple remote sites. Plans are underway to experiment further with MBONE over ATM (e.g., for distance education course delivery), as well as with IPv6 on this same NTN test environment.

Acknowledgements

The traffic measurements presented in this paper would not have been possible without the NavTel IW95000 ATM test set borrowed from TR*Labs* in Winnipeg. The authors gratefully acknowledge the cooperation and assistance provided by Clint Gibler, Len Dacombe, Dave Blight, Jeff Diamond, Yair Bourlas, Martin Meier, and others at TR*Labs* Winnipeg.

The SGI Indy at the University of Saskatchewan, and the connection to the MBONE, were provided by CANARIE through the Multicast Test Project, led by Ken Hewitt at the University of Calgary. The project participants are grateful to Steve Ladouceur and others at Bell Canada for making the MBONE work on the National Test Network.

Financial support for this research was provided by TR*Labs* in Saskatoon and by the Natural Sciences and Engineering Research Council (NSERC) of Canada, through research grants OGP0120969, IOR152668, and through an Industrial NSERC Summer Undergraduate Research Award.

A. Appendix

This appendix provides a brief description of the audio and video formats used in the MBONE, as well as the RTP protocol used by the MBONE tools.

A.1 Audio Formats

Several audio teleconferencing tools are available for use over the Internet and MBONE. The most popular of these are `rat` and `vat`. Both of these packages allow several different audio formats to be used including [9]:

pcm: 78 kbps 8-bit mu-law encoded 8KHz PCM (20 ms frames)

pcm2: 71 kbps 8-bit mu-law encoded 8KHz PCM (40 ms frames)

pcm4: 68 kbps 8-bit mu-law encoded 8KHz PCM (80 ms frames)

dvi: 46 kbps Intel DVI ADPCM (20 ms frames)

dvi2: 39 kbps Intel DVI ADPCM (40 ms frames)

dvi4: 36 kbp/s Intel DVI ADPCM (80 ms frames)

gsm: 17 kbps GSM (80 ms frames)

lpc4: 9 kbps Linear Predictive Coder (80 ms frames)

The *pcm* (pulse code modulation) format with 8-bit mu-law encoded 8 kHz PCM is standard for telephony and generates a raw data rate of 8 bits $\times$ 8 kHz = 64 kbps. The difference between the three *pcm* formats above is in the number of audio samples that are grouped together for transmission. Aggregating more samples minimizes the overhead and required bit rate but may cause longer delays in the conversation as is the case with *pcm4*. On the other hand, *pcm* sends out audio frames at a faster pace but this requires an extra 10 kbps.

A.2 Video Formats

The MBONE software allows the user to manually configure the video encoding format to be used for transmission. Video received over the MBONE is automatically decoded via the correct decoding scheme. One of the most popular MBONE video tools is `vic`, which is the software we used at the University of Saskatchewan for transmission. The `vic` software supports six different video transmission schemes [8]:

- Motion JPEG
- MPEG
- CCITT H.261
- Xerox PARC Network Video
- Sun CellB
- Berkeley Video Codec

All video transmissions done from the University of Saskatchewan used the default H.261 scheme. It is assumed that most MBONE video is sent via this format.

A.3 Overview of the H.261 Video Standard

The H.261 video standard [6] was developed for use over one or more 64 kbps ISDN lines. The source coder operates on a 30 frames per second non-interlaced picture. However, depending on the bandwidth available, the source coder may skip up to three intermediate frames. Two picture scanning formats are defined:

1. **CIF** (Common Intermediate Format)
   - luminance sampling (Y) is 352 pels per line, 288 lines per picture
   - two color difference components (C_B and C_R) are sampled at 176 pels per line, 144 lines per picture
   - maximum number of bits allowed per frame (sub-image) is 256 kbits
2. **QCIF** (Quarter-CIF)
   - luminance sampling is 176 pels per line, 144 lines per picture
   - the two color difference components are sampled at 88 pels per lines, 72 lines per picture
   - maximum number of bits allowed per frame (sub-image) is 64 kbits

Motion compensation intraframe prediction and adaptive discrete cosine transform (DCT) coding are used to reduce temporal and spatial redundancies. The video multiplex is arranged in a hierarchical structure with the following layers:

- Picture
- Group of blocks (GOB)
- Macroblock (MB)
- Block

A GOB consists of 176 pels by 48 lines of Y and the corresponding 88 pels by 24 lines of C_B and C_R. Therefore, a picture is composed of 12 GOB's in the CIF format and 3 GOB's in the QCIF format. Each GOB is further divided into 33 macroblocks, each of which contains 16 pels by 16 lines of Y and the corresponding 8 pels by 8 lines of C_B and C_R. Macroblocks are not transmitted when they contain no information for that part of the picture. Each transmitted macroblock may contain information on one or more of the four luminance blocks and two color difference blocks that it comprises.

A.4 Real-Time Protocol (RTP)

The MBONE tools make use of a protocol called RTP (Real-Time Protocol) to multiplex and demultiplex audio and video traffic using a single UDP/IP packet stream. This protocol adds 12 bytes of overhead to every UDP/IP packet.

Investigation of the source code for `vat` identified the format of the RTP header that accompanies all MBONE packets, as well as other useful information [10]. The 12 byte header is composed of the following fields:

2 bytes header flags

2 bits version number

- **1 bit** padding is present
- **1 bit** extension header is present
- **4 bits** source description information
- **1 bit** marker for significant event such as a frame boundary
- **7 bits** payload type

2 bytes sequence number of packet

4 bytes media-specific time stamp

4 bytes synchronization source identification

The source code also identifies the payload type indicators used to identify the encoding format. The significant indicators for our purposes are:

31 ITU H.261

0 PCM Audio Encoding

This RTP protocol information is used to decode audio and video streams in our cell-level traces of network traffic.

References

[1] D. Comer, *Internetworking with TCP/IP, Volume 1: Principles, Protocols, and Architecture*, Second Edition, Prentice-Hall, Englewood Cliffs, NJ, 1991.

[2] S. Deering, *Multicast Routing in a Datagram Internetwork*, Ph.D. Thesis, Stanford University, 1991.

[3] S. Deering, D. Estrin, D. Farinacci, and V. Jacobson, "An Architecture for Wide-Area Multicast Routing", *Proceedings of the 1994 ACM SIGCOMM Conference*, London, UK, pp. 126-135, August 1994.

[4] H. Eriksson, "MBONE: The Multicast Backbone", *Communications of the ACM*, Vol. 37, No. 8, pp. 54-60, August 1994.

[5] M. Garrett and W. Willinger, "Analysis, Modeling and Generation of Self-Similar VBR Video Traffic", *Proceedings of ACM SIGCOMM '94*, London, UK, pp. 269-280, August 1994.

[6] International Telecommunication Union, "Line Transmission of Non-Telephone Signals: Video Code for Audiovisual Services at p x 64 kbits", ITU-T Recommendation H.261, March 1993.

[7] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the Self-Similar Nature of Ethernet Traffic (Extended Version)", *IEEE/ACM Transactions on Networking*, Vol. 2, No. 1, pp. 1-15, February 1994.

[8] Steven McCanne. 1995 `vic` *man page.* Berkeley, CA.

[9] S. McCanne & V. Jacobson. `vat` *man page.* Berkeley, CA.

[10] S. McCanne & V. Jacobson. `vat` source code. Berkeley, CA.

[11] W. Stevens, *TCP/IP Illustrated, Volume 1*, Addison-Wesley, Reading, Massachusetts, 1993.

[12] G. Wright and W. Stevens, *TCP/IP Illustrated, Volume 2: The Implementation.* Addison-Wesley, Reading, Massachusetts, 1995.

[13] D. Waitzman, C. Partridge, and S. Deering (editors), *Distance Vector Multicast Routing Protocol*, RFC 1075, BBN STC and Stanford University, November 1988.