The Effects of High-Speed Networks on Multimedia Jitter

Mark Claypool
`claypool@cs.wpi.edu`
Worcester Polytechnic Institute
Computer Science Department
Worcester, MA 01609, USA

John Riedl
`riedl@cs.umn.edu`
University of Minnesota
Computer Science Department
Minneapolis, MN, USA

Abstract

Jitter can cause silent gaps in the playout of an audio stream such as in an audioconference, or a choppy appearance to a video stream for a videoconference, especially under conditions of high processor load. We experimentally measure the effects of high-speed networks on jitter in a multimedia stream. We incorporate our jitter measurements into a general model for multimedia application quality. Our model allows us to explore how advances network speeds will improve application quality compared with other jitter reduction techniques. As an example, we apply our model to a videoconference. We find high-speed networks significantly reduce jitter under conditions of heavy processor load, but tend to affect overall quality less than processor improvements.

1 Introduction

There are many exciting new distributed multimedia applications. Today, two to tens of users can communicate through a computer audioconference. Tomorrow, tens to hundreds of neuroscientists will explore and contribute to a distributed brain database [CRG$^+$94]. Soon, tens, hundreds and perhaps even thousands of soldiers will train for combat in a distributed interactive simulation [Com94].

In order to build systems that will adequately support such applications, it is important to predict application performance as the number of users increases and evaluate performance and cost tradeoffs for different system designs. One indication of the performance of an entire computer system is the user's opinion on the *multimedia quality* of the applications they run. Multimedia quality is a measure of how closely the performance of a multimedia application meets the requirements expected by the user. If the user performance requirements are met, application quality will be acceptable. If the user performance requirements are not met, application quality will be unacceptable. We are developing a quality planning model to aid in designing systems that meet users' quality requirements for multimedia applications in the future.

Although we often think of a multimedia application as a continuous stream of data, computer systems handle multimedia in discrete events. An event may be receiving an update packet or displaying a rendered frame on the screen. The quantity and timing of these events give us measures that affect application quality. We have identified three measures that determine quality for most distributed multimedia applications [CR98]:

- *Latency.* The time it takes information to move from the server through the client to the user we call *latency.* Latency decreases the effectiveness of applications by making them less like real-life interaction [IKK93, Roy94, DCJ93].
- *Data Loss.* Any data less than the amount determined by the user requirements we call *data loss.* Data loss takes many forms such as reduced bits of color, jumbo pixels, smaller images, dropped frames and lossy compression [AFKN95, MS94, SW93]. Data loss may be done voluntarily by either the client or the server in order to reduce load or to reduce jitter and/or latency.
- *Jitter.* In a distributed multimedia application, a multimedia stream is generated at the sending workstation and sent over a network to the receiving workstation. The data is generated and sent in fixed-sized quantities called frames. The end-to-end frame delay is the time between a frame's generation on the sender and the time it is processed by the receiver. Variation in this end-to-end delay we call *jitter.*

How does jitter affect a multimedia stream? In the absence of jitter, the frames can be played as they are received, resulting in a smooth playout, depicted by Figure 1. However, in the presence of jitter, interarrival times will vary, as depicted in Figure 2. In Figure 2 the third frame arrives late at r*2*. In the case of audio speech, the listener would experience an annoying pause during this period. In the case of video, the viewer would see the frozen image of the most recently delivered frame.

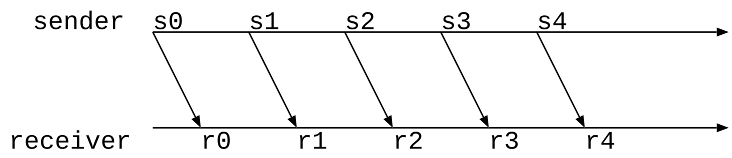

Figure 1: A Jitter-Free Stream. The above figure is a model of a jitter-free stream. Each si is the time at which the send process initiates the transmission of frame i. Each ri is the time at which the receiving process plays frame i.

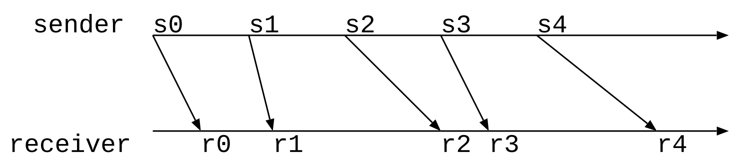

Figure 2: A Stream with Jitter. The above figure is a model of a stream with jitter. Each si is the time at which the send process initiates the transmission of frame i. Each ri is the time at which the receiving process plays frame i.

High-performance processors have higher throughput and a faster context switch time than typical processors resulting in better application response time. High-speed networks deliver frames from the sender to the receiver faster than typical networks, reducing the network transmission time. Together, high-performance processors and high-speed networks networks will reduce application latency. The reduced latency should be accompanied by a reduction in latency variation, or jitter. We run experiments on SGI Challenge workstation clusters and Fibre Channel and HIPPI networks under both light and heavy load to determine the effects that hardware improvements have on jitter.

The effects of latency on a user's perception of an application is well-understood and well-researched [IKK93, Roy94, DCJ93]. Similarly, there is a clear relationship between data loss and application quality deterioration [AFKN95, MS94, SW93, HSK98, PHH98, GKL+98]. Methods to ameliorate the effects of jitter have been explored by many researchers [SJ95, RS94, Fer92]. The tradeoff between buffering and jitter has also been explored [KN82]. There has also been work done on multimedia quality assessment [WS98]. To the best of our knowledge, the effects of high-performance workstations and real-time priorities has not been thoroughly investigated. Moreover, the effects of jitter on the overall quality of a multimedia application has not been established.

The major contributions of this paper are:

- An experimentally-based study of high-speed networks on jitter.
- A multimedia application quality model that enables the prediction of application performance and evaluation of system design tradeoffs.
- Performance predictions for videoconferences using with high-speed networks.

The rest of this paper is laid out as follows: Section 2 details the experiments to measure the effects of high spped networks on jitter. Section 3 describes how jitter fits into our a model for application quality. Section 4 applies our quality model to a videoconference. Section 5 summarizes our conclusions. Finally, Section 6 lists possible future work.

2 Experiments

We ran experiments to measure the effects of high-speed networks on jitter. Our experiments were all conducted on a single-hop LAN. Each experiment simulated the transmission of a multimedia stream under various conditions and measured the amount of jitter. We used pairs of user-level processes that sent and received UDP datagrams using Berkeley socket I/O. The `send` process used an interval timer to initiate frame delivery. The `receive` process took timestamps using the `gettimeofday()` system call. These timestamps are used to compute a measure the amount of jitter. We choose variance as our measure of jitter because it is easy to understand and compute and has been used by other researchers to measure jitter [JVS91, BP80].

We used 4 processor SGI Challenge L workstations. Each workstation was connected to three networks: a 10 Mbit/s Ethernet; a 266 Mbit/s Fibre Channel; and a 800 Mbit/s HIPPI.

We induced network load using `netperf` [Jon95]. Netperf is a benchmark that can be used to measure the performance of many different types of networks. It provides tests for both uni-directional throughput and end-to-end latency.

Both HIPPI and Fibre Channel are point-to-point networks, unlike Ethernet which is a shared medium network. Inducing network load between two workstations that were not the sender and receiver did not increase jitter because they did not share the same physical wire. Instead, we ran the `netperf` and `netserver` processes on the same workstations as the sender and receiver, respectively. We ran the netperf processes on different processors than the sender and receiver by using the `runon` command. This minimized the jitter that might be contributed by the processor load induced by the `netperf` processes.

2.1 Analysis

Figure 3 depicts the experiments results that show the interarrival times for the three networks under two different conditions of network load. The "no load" runs are the experiments on a quiet network. The "load" runs are the experiments run with `netperf`. The three "no load" lines are almost indistinguishable, indicating no correlation between jitter and network bandwidth. However, under high network load there is a noticeable difference in the interarrival times for the three networks, indicating there may be a correlation between jitter and network bandwidth.

Figure 4 shows the relationship between jitter and network bandwidth for loaded networks. We expect that jitter decreases as network bandwidth increases, so we graphed the theoretical network bandwidth versus 1/variance. There are three data points plotted, one for each of Ethernet, Fibre Channel and HIPPI. The line represents a least squares line fit through the three data points. The correlation coefficient is 0.98, indicating that there is a strong inverse relationship between jitter and network bandwidth. In other words, as network bandwidth increases, jitter decreases. Note that although the correlation coefficient is a high 0.98, the fact that there are only three data points is evident in the width of the 95% confidence intervals around the least squares line fit. It would be nice to have more data points in order to strengthen the statistical significance of our results. To do this, however, we need to have new networks on which to experiment.

2.2 Summary

Under low network loads, high-speed networks do not significantly reduce jitter. Under low loads, the network contributes little variation to the interarrival times for the multimedia frames. Most of the variance is caused outside the network, rendering any jitter reduction from the high-speed networks insignificant.

However, under heavy network loads, high-speed networks significantly reduce jitter. Under heavy loads, interfering network traffic causes increased variation in the interarrival times for the multimedia frames. High-speed networks reduce the amount of time to deliver the interfering network traffic, decreasing the variation in the multimedia interarrival times.

The amount of jitter reduction from high-speed networks depends upon the network load. We use our model in Section 3 to determine the network load from multimedia applications as the number of users increases. Then, using the results from this section, we can determine what affect jitter reduction from high-speed networks has on the application quality.

3 Quality

Ideally, we would like there to be no latency, jitter or data loss. Unfortunately, on a variable delay network and non-dedicated computer this can never be achieved. To compute the application quality, we use the above quality components in a process depicted by Figure 5. The user requirements for the application define the acceptable latency, jitter and data loss. The system determines the predicted latency, jitter and data loss. Acceptable and projected data are fed into a *quality metric* for the application. The quality metric is a function, based on the acceptable components and dependent upon the projected components, that computes the application quality.

In order to quantitatively compare application quality for different system configurations, we need a reasonable quality metric. To form our quality metric, we build upon the work of Naylor and Kleinrock [KN82]. Naylor and Kleinrock developed a model for measuring the quality of an audioconference based on the probability of playout gaps and end-to-end delay. The quality of the audioconference was computed by taking the normalized distance of the audioconference's delay and gaps from the origin in the delay-gap plane. We extend this model by using latency, jitter and data loss as axes, creating a multi-dimensional quality space. We place the best quality value for each axis at the origin and normalize each axis so that the user-defined minimum acceptable values have equal weights. The user-defined minimum depend upon the type of media (say, audio or video) and the use of the application (say, broadcast or interactive). An instantiation of the application lies at one point in this space. The location of the point is determined by our predictions of the amount of latency, jitter an data loss that would occur with the given system configuration. In order to satisfy the mathematical properties of a metric, we compute the application quality by taking the Euclidean distance from the point to the origin. All points inside the region defined by the user-defined minimums have acceptable quality while points outside do not.

Figure 6 depicts a 3-d quality space for multimedia applications. The user requirements determine a region of acceptable application quality, depicted by the shaded region. Each instantiation of the application and the underlying computer system is a point in this space.

There can be many possible quality metrics for a given application. In fact, there may be many quality metrics that agree with a user's perception of the application. However, the rest of our method is independent of the quality metric chosen. If new metrics are developed and validated with user testing, they can be used in place of our quality metric.

4 An Example: Videoconference Quality

In this section, we apply the predictive powers of our quality model presented in Section 3 to a videoconference. In order to apply our quality model to a videoconference under various system configurations, we must: 1) determine the region of acceptable videoconference quality; 2) predict jitter; 3) predict latency; and 4) predict data loss.

4.1 The Region of Acceptable Videoconference Quality

To determine the region of acceptable videoconference quality, we need to define acceptable limits for audioconferences along each of the latency, jitter and data loss axes. According to Jeffay and Stone, delays of 230 milliseconds or under are acceptable for a videoconference [JSS92]. For data loss, research in remote teleoperator performance has found that task performance is virtually impossible below a threshold of 3 frames per second [MS94]. We use 3 frames per second as the minimum acceptable frame rate.

The presence of jitter often presents an opportunity for a tradeoff among latency and data loss. Buffering, an application-level technique for ameliorating the effects of jitter, can compensate for jitter at the expense of latency. Transmitted frames are buffered in memory by the receiver for a period of time. Then, the receiver plays out each frame with a constant latency, achieving a steady stream. If the buffer is made sufficiently large so that it can hold all arriving data for a period of time as long as the tardiest frame, then the user receives a complete, steady stream. However, the added latency from buffering can be disturbing, so minimizing the amount of delay compensation is desirable.

Another buffering technique to compensate for jitter is to discard any late frame at the expense of data loss. Discarding frames causes a temporal gap in the play-out of the stream. Discarding frames can keep play-out latency low

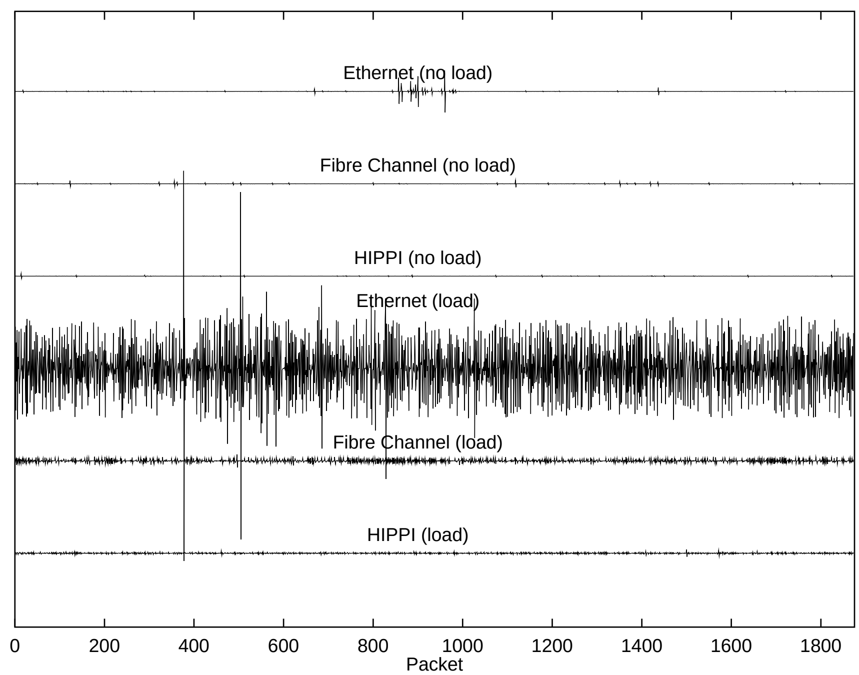

Figure 3: Interarrival Times for Different Networks and Network Loads. There are 6 multimedia streams represented in this picture, each by by a horizontal line depicting the interarrival times on the indicated network. The network was either either loaded or unloaded. Each multimedia stream is off-set from the one below it by 750,000 microseconds. The horizontal axis is the packet number.

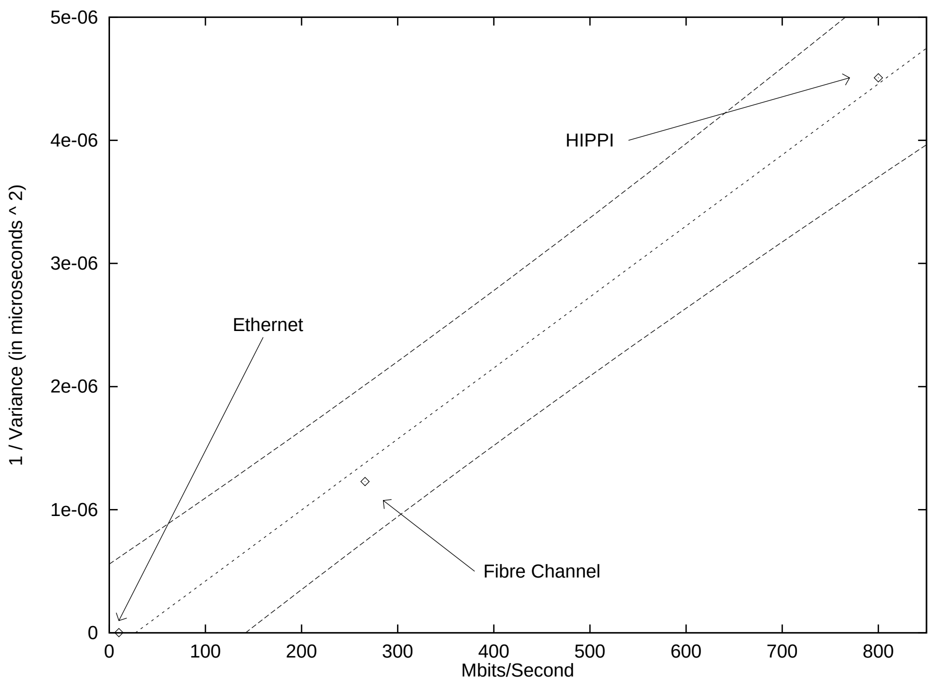

Figure 4: Jitter versus Network Bandwidth. The horizontal axis is the network bandwidth. The vertical axis is 1 / variance. The line is a least squares line fit of jitter versus theoretical network bandwidth for three networks: Ethernet, Fibre Channel and HIPPI. The correlation coefficient is 0.98.

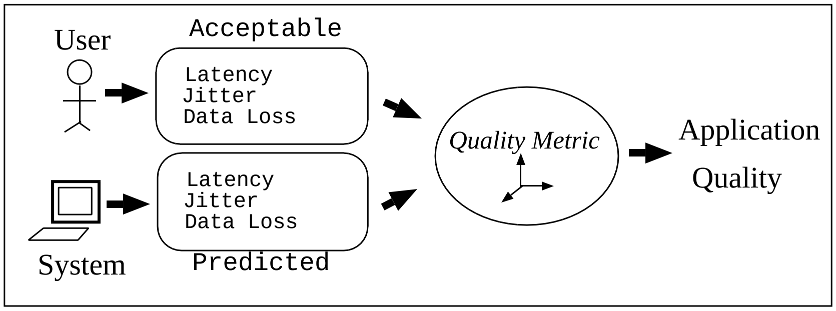

Figure 5: The Process for Computing Application Quality. The user defines the acceptable latency, jitter and data loss and the system determines the actual values. Based on the acceptable values specified in the user requirements, a quality metric computes the application quality from the actual values.

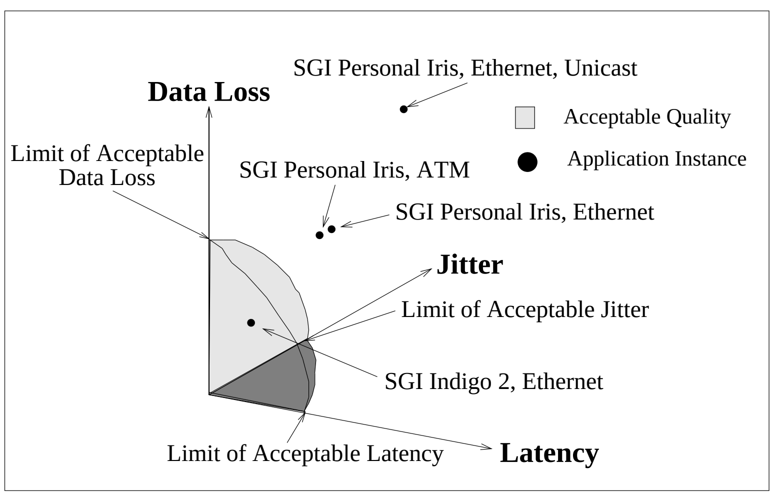

Figure 6: Multimedia Application Quality Space. The user defines the acceptable latency, jitter and data loss. These values determine a region of acceptable application quality, depicted by the shaded region. All points inside the shaded region have acceptable quality, while those outside the region do not. An instantiation of the application and the underlying computer system lies at one point in this space. Four application configuration instantiations are shown.

and constant, in the case of audio speech, the listener would experience an annoying pause during this period. In the case of video, the viewer would see the frozen image of the most recently delivered frame.

4.2 Predicting Jitter

Our experimental results in Section 2 give us the relationship between network bandwidth and jitter for loaded and unloaded networks. We use these results as the basis for determining the jitter in the videoconference under various system configurations.

4.3 Predicting Latency

We can predict the amount of latency from the jitter compensation buffer by using predictions on the amount of jitter. In addition to the buffering latency, there is the additional latency from the sender processing, the network transmitting and the receiver processing. We can compute the latency from the network based on the frame size and network bandwidth. To predict the total latency, we add the latencies from: recording the video frame; sending the video frame to the client; receiving the video frame from the receiver; buffering in the jitter compensation curve; and playing the video frame.

4.4 Predicting Data Loss

In order to predict data loss, we need to identify what form data loss may take and when data loss may occur. In general, data loss can take many forms such as reduced bits of color, jumbo pixels, smaller images, dropped frames and lossy compression. For a videoconference, we assume data loss only in the form of dropped frames, when an application chooses to discard late frames in order to keep playout latency low and constant, or when either the network or the processor do not have sufficient capacity to transmit data at the required frame rate.

4.5 Predicting Quality

We compute quality under two different scenarios. In the first, processor load remains constant while the network bandwidth increases. In the second, network bandwidth remains constant while processor power increases. Figure 7 shows these predictions. For five users, increasing the processor power to a SPECint92 of 40 or greater results in acceptable videoconference quality. At no time does increasing the network bandwidth result in an acceptable quality. In this scenario, we conclude that processor power influences videoconference quality more than does network bandwidth.

While today's computer systems may struggle to support even five videoconference participants, tomorrow's processor improvements promise to support more and more users. But how many more? How do more and more videoconference users affect application quality? Figure 8 depicts the predicted effects of increasing users on videoconference quality. We predict videoconference quality for three different videoconference configurations: a low-end workstation with a typical network (Sun IPX and Ethernet), a mid-range workstation with a fast network (Sun Sparc 5 and Fibre Channel), and a high-performance workstation with a high-speed network (DEC Alpha and HIPPI). Powerful workstations such as Sun Sparc 5s connected by fast networks such as a Fibre Channel can support up to 10 users. Very high-performance workstations such as a DEC Alpha connected by high-speed networks such as a HIPPI can support over 50 users.

5 Conclusions

Jitter hampers computer support for multimedia. Jitter is the variation in the end-to-end delay for sending data from one user to another. Jitter can cause gaps in the playout of a stream such as in an audioconference, or a choppy appearance to a video display for a videoconference. There are several techniques that can be used to reduce jitter. In this work, we have experimentally measured the effects of high-speed networks on jitter.

We find high-speed networks significantly reduce jitter under conditions of heavy load. Multimedia applications, tending to be resource intensive, are likely to push processors capacities to the limit, making conditions of heavy load likely. As the growth in distributed collaborative applications continues, multimedia applications will push network bandwidths to the limits, also. Thus, high-performance processors and real-time priorities will be necessary to reduce jitter.

However, improving jitter alone is not sufficient to guarantee improving application quality. In addition to jitter, the application quality also depends upon *latency*, which decreases the application realism, and *data loss*, which reduces the application resolution. We have developed a model that allows us to evaluate the effects of jitter reduction in the larger context of a user's perception of the multimedia application. Our model allows us to show how advances in networks and processors will improve application quality without tuning the operating system or the application.

6 Future Work

Another possible potential jitter reduction technique would be a real-time network protocol. Future work includes experiments to measure the effects of real-time network protocol on jitter. However, network delivery of data within strict jitter bounds does not significantly help the sender and receiver. They must still worry about internal jitter due to queuing in the operating system. Stamping out jitter in the network does not eliminate the need for jitter management code in the hosts.

In our analysis of a videoconference, we assumed a constant bit rate. Video is almost invariably transmitted in a compressed form for the simple reason that compression is so effective. Commercial compression algorithms can achieve 25-to-1 or better reductions in the average number of bits transmitted. However, the amount of data that must be sent for each video frame can vary widely, depending upon the amount changes from frame to frame. Transmission schemes that generate variable amounts of data for each frame are termed *variable bit-rate* (VBR) video schemes, such as MPEG [PSR93]. MPEG codes video using three frame formats. The effects of losses in the MPEG data stream depend upon which type of frame is lost. In the absence of results determining the effects of VBR frame rate losses on quality, we assumed that all frames are of equal importance in determining video quality. Future work might determine the effects on video quality for different types of frames.

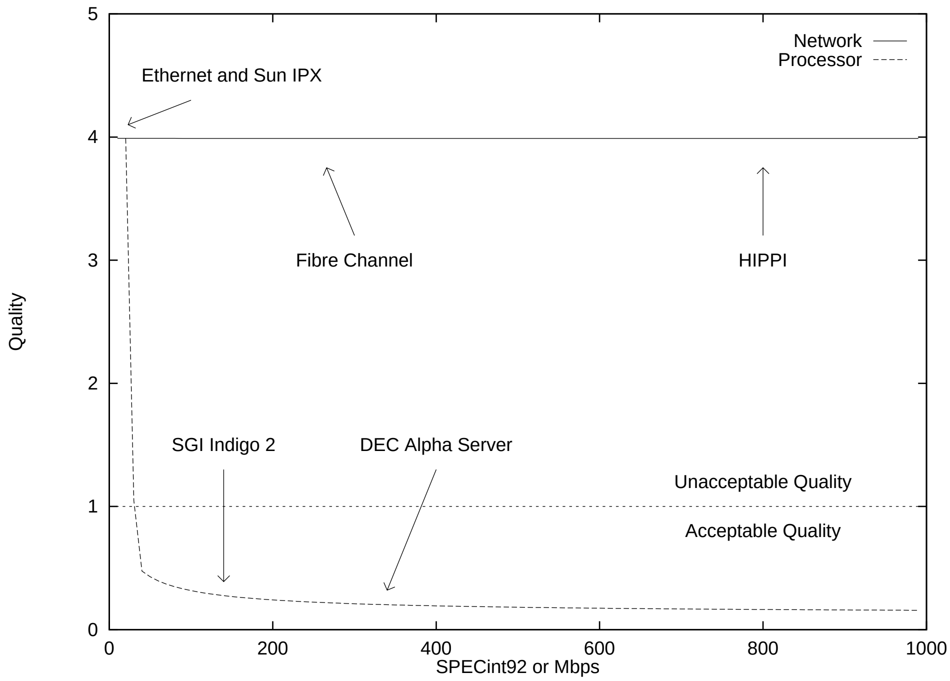

Figure 7: Videoconference Quality versus Processor or Network Increase. The horizontal axis is the SPECint92 power of the workstation or the network Mbps. The vertical axis is the predicted quality. There are two scenarios depicted. In the first, the processor power is constant, equivalent to a Sun IPX (SPECint92 = 22), while the network bandwidth increases. This is depicted by the solid curve. In the second scenario, the network bandwidth is constant, equivalent to an Ethernet (10 Mbps), while the processor power increases. This is depicted by the dashed curve. The horizontal line marks the limit between acceptable and unacceptable videoconference quality.

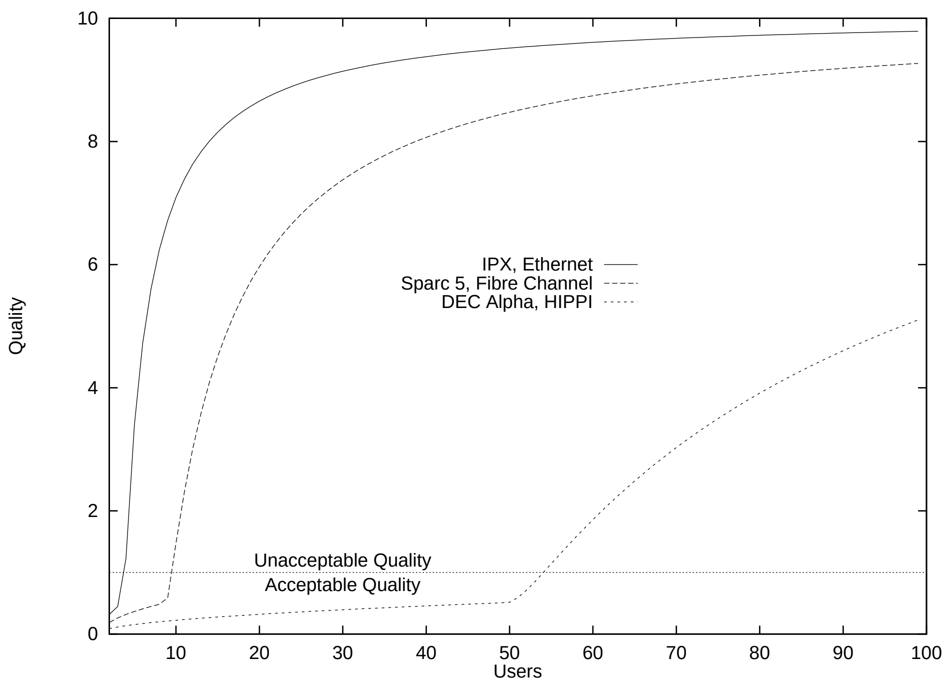

Figure 8: Videoconference Quality versus Users. The horizontal axis is the number of users. The vertical axis is the predicted quality. There are three scenarios depicted. In the first, the processors is a Sun IPXs connected by an Ethernet. In the second, the processors is a Sun Sparc 5s connected by a Fibre Channel. In the third, the processors are DEC Alphas connected by a HIPPI. The horizontal line marks the limit between acceptable and unacceptable videoconference quality.

References

[AFKN95] Ronnie T. Apteker, James A. Fisher, Valentin S. Kisimov, and Hanoch Neishlos. Video acceptability and frame rate. *IEEE Multimedia*, pages 32 – 40, Fall 1995.

[BP80] Barberis and Pazzaglia. Analysis and optimal design of a packet voice receiver. *IEEE Transactions on Communication*, February 1980.

[Com94] The DIS Steering Committee. The DIS vision - a map to the future of distributed interactive simulation. Technical report, Institute for Simulation and Training, May 1994.

[CR98] Mark Claypool and John Riedl. End-to-end quality in multimedia applications. In *Chapter 40 in Handbook on Multimedia Computing*, CRC Press, Boca Raton, FL, 1998.

[CRG$^+$94] J. Carlis, J. Riedl, A. Georgopoulos, G. Wilcox, R. Elde, J. H. Pardo, K. Ugurbil, E. Retzel, J. Maguire, B. Miller, M. Claypool, T. Brelje, and C. Honda. A zoomable DBMS for brain structure, function and behavior. In *International Conference on Applications of Databases*, June 1994.

[DCJ93] Spiros Dimolitsas, Franklin L. Corcoran, and John G. Phipps Jr. Impact of transmission delay on ISDN videotelephony. In *Proceedings of Globecom '93 – IEEE Telecommunications Conference*, pages 376 – 379, Houston, TX, November 1993.

[Fer92] Domenico Ferrari. Delay jitter control scheme for packet-switching internetworks. *Computer Communications*, 15(6):367–373, July 1992.

[GKL$^+$98] Steven Gringeri, Ghumip Khasnabish, Arianne Lewis, Khaled Shuaib, Roman Egorov, and Bert Basch. Transmission of MPEG-2 vidoe streams over ATM. *IEEE Multimedia*, 5(1):58 – 71, Jan-Mar 1998.

[HSK98] V. Hardman, M. A. Sasse, and I Kouvelas. Successful multi-party audio communication over the internet. *Communications of the ACM*, 41(5):74 – 80, 1998.

[IKK93] Satoru Iai, Takaaki Kurita, and Nobuhiko Kitawaki. Quality requirements for multimedia communcation services and terminals – interaction of speech and video delays. In *Proceedings of Globecom '93 – IEEE Telecommunications Conference*, pages 394 – 398, Houston, TX, November 1993.

[Jon95] Rick Jones. *Netperf*. Hewlett-Packard, 1995. The netperf home page can be found at http://www.cup.hp.com/netperf/NetperfPage.html.

[JSS92] K. Jeffay, D. Stone, and F. Smith. Kernel support for live digital audio and video. *Computer Communications*, 15(6), 1992.

[JVS91] Saimin Jin, Dhadesugoor R. Vaman, and Divyendu Sina. A performance mangement framework to provide bounded packet delay and variance in packet switched networks. *Computer networks and ISDN Sytems*, pages 249 – 264, September 1991.

[KN82] Kleinrock and Naylor. Stream traffic communication in packet switched networks: Destination buffering considerations. *IEEE Transactions on Communications*, COM-30(12):2527 – 2534, December 1982.

[MS94] Michael J. Massimino and Thomas B. Sheridan. Teleoperator performance with varying force and visual feedback. In *Human Factors*, pages 145 – 157, March 1994.

[PHH98] Colin Perkins, Orion Hodson, and Vicky Hardman. A survey of packet-loss recovery techniques for streaming audio. *IEEE Network Magazine*, September/October 1998.

[PSR93] Ketan Patel, Brian C. Smith, and Lawrence A. Rowe. Performance of a software mpeg video decoder. In *Proceedings of ACM Multimedia*, Anaheim, 1993.

[Roy94] Radhika R. Roy. Networking contraints in multimedia conferencing and the role of ATM networks. *AT&T Technical Journal*, July/August 1994.

[RS94] Ramachandran Ramjee, Jim Kurose, Don Towsley and Henning Schulzrinne. Adaptive playout mechanisms for packetized audio applications in wide-area networks. In *Proceedings of the 13th Annual Joint Conference of the IEEE Computer and Communications Societies on Networking for Global Communciation. Volume 2*, pages 680–688, Los Alamitos, CA, USA, June 1994. IEEE Computer Society Press.

[SJ95] D. Stone and K. Jeffay. An empirical study of delay jitter management policies. *ACM Multimedia Systems*, 2(6):267 – 279, January 1995.

[SW93] Merryanna Swartz and Daniel Wallace. Effects of frame rate and resolution reduction on human performance. In *Proceedings of IS&T's 46th Annual Conference*, Munich, Germany, 1993.

[WS98] A. Watson and M. Sasse. Measuring perceived quality of speech and video in multimedia conferencing applications. In *Proceedings of ACM Multimedia*, pages 55 – 60, Bristol, September 1998.