Glitches as a Measure of Video Quality Degradation Caused by Packet Loss*

İsmail Dalgıç   Fouad A. Tobagi

Department of Electrical Engineering
Stanford University
Stanford, CA 94305

ABSTRACT: *The loss or excessive delay of video data in networks causes glitching in the display of video. In this paper, we consider some glitch statistics (i.e., glitch rate, average spatial extent, and duration) to evaluate network performance when carrying video traffic. We present some numerical results for ATM and 100Base-T Ethernet networks, obtained by simulating these networks using real video traces for various contents and encoding schemes.*

1 Introduction

In the context of networked video, *glitching* is the word used to refer to the effect seen by the viewer in the display of video due to the unavailability of video data at the decoder when needed (i.e., when there is a discontinuity in the video stream). The unavailability of video data may be due to two reasons: (i) the data may have been lost or not delivered in time; (ii) the data may have been delivered in time, but it may depend on other parts of data that are not available. Such unavailability of video data results in portions of multiple consecutive frames not being displayed (or replaced by information from older frames).

In this paper, we use some glitch statistics for evaluating the network performance when carrying video traffic. (This is in contrast with *packet loss rate* often used as a measure of network performance.) We present some numerical results for ATM and 100Base-T Ethernet networks, obtained by simulating these networks using real video traces, and deriving the glitch statistics.

2 Glitches as a Network Performance Measure

In order to characterize glitching in the display of video, and to give quantitative measures for its occurences, we need to accurately define what constitutes a glitch. A *glitch* is an event which is said to begin when a portion of a frame is not displayed due to the unavailability of data, while its preceding frame is fully displayed; the glitch is said to continue as long as each consecutive frame after the beginning of the glitch contains a portion that is not displayed; and it is said to have ended when a subsequent frame is fully displayed.

We can now define three quantities of interest which can be used to characterize the network performance in terms of glitches.

glitch duration: The glitch duration may be expressed in units of time, or equivalently, in number of frames for a given frame rate.

spatial extent: The *spatial extent* of a glitch within a given frame is defined as the percentage of the undisplayed portions in the frame, or equivalently, the number of undisplayed macroblocks in the frame. Note that the shape and extent of the undisplayed region may vary within a glitch from frame to frame; this may for example be due to some data not being delivered to the decoder somewhere in the middle of the glitch, thereby causing additional regions not to be displayed. In this paper, in order to keep the definitions simple, we do not consider the shape of the undisplayed regions in a glitch, and focus only on the spatial extent. The simulation results that we present in Section 3 of this paper also justify this approach. Given that the spatial extent may vary in time within a glitch, it is also interesting to consider the *average, standard deviation, minimum*, and *maximum* values of the spatial extent for a given glitch.

glitch rate: The glitch rate is defined simply as the number of glitches per unit time that a video stream experiences.

It is important to note that all glitches do not have the same quality degradation effect to the viewer. In fact, some glitches may not be perceived at all, particularly if their duration is not more than 2-3 frames. For a given glitch, the greater its duration and spatial extent, typically the more objectionable it would be to the viewer. Also note that, when there is a glitch, some decoders may perform error concealment actions by using information from previous frames and from the available portions of the current frame. Such error concealment actions would affect the perceived effect of a

*This work was supported in part by NSF under grant NCR-9016032, and by Pacific Bell.

glitch on the viewer. Given the particular error concealment actions of a decoder, the glitch measures may be mapped to quality impairment ratings by means of models of the human visual system. However, in this paper, we only present the raw glitch statistics; we will study the effect of glitches on the perceived quality in the future. Once such a study is done, the glitches may be divided into several classes according to their perceptual effect, and glitch rates may be computed separately for each class.

Clearly, the duration, spatial extent, and rate of the glitches will depend on five factors: (i) the network, (ii) the traffic scenario, (iii) the video encoding scheme, (iv) the video encoder control scheme, and (v) the video content. In the following, we discuss the effect of these factors on the glitch statistics.

The network type affects the statistics of packet loss and delay. Furthermore, the packetization process and the maximum packet size depend on the network type. Therefore, it can be expected that the network type has an important effect on all three glitch quantities of interest. In the next section, we consider 100Base-T and ATM networks as examples, and show that there is a significant difference between the two as far as the glitch statistics are concerned.

The traffic scenario also has an effect on the statistics of packet loss, since it determines the load on the network. Clearly, as the network load increases, the glitch rate also increases. In the next section, we examine this effect, as well as the effect of the traffic load on the glitch spatial extent and duration.

The video encoding scheme affects the glitch statistics because for different schemes, the dependencies among frames are different. To elaborate more on this point, consider as examples H.261 [1] and MPEG [2] video encoding standards. In H.261, frames are divided into a number of Group of Blocks (GOBs); (a frame consists of 3, 10, or 12 GOBs, depending on the frame format). Typically, one GOB in each frame is intracoded, and the intracoded GOB is rotated cyclically among all the GOBs from frame to frame. Also, the motion vectors are typically confined within the limits of a GOB when such a cyclic intracoding of GOBs is used. In this case, when a glitch which affects only one GOB occurs, it will be recovered in the frame where the affected GOB is intracoded. Therefore, assuming that no other loss occurs in the middle of the glitch, its duration will be anywhere from one frame to the number of GOBs in a frame. Similarly, when a glitch which affects multiple GOBs occurs, it will be recovered one GOB at a time as the affected GOBs are intracoded in successive frames. Taken to the extreme, if the glitch affects the entire frame, it will take at least as many frames to fully recover from it as the number of GOBs in a frame. Therefore, in this case, the duration of a glitch is dependent on its initial spatial extent. By contrast, in MPEG, when a glitch begins in an I or P frame, it affects all the B and P frames dependent on that frame. Therefore, the duration of the glitch is equal to the number of all those frames that are affected. When a glitch begins in a B frame, it does not propagate; thus the duration of such a glitch is always 1 frame. Therefore, in MPEG, the duration of a glitch is generally independent of the initial spatial extent. We compare the glitch statistics for H.261 and MPEG encoded sequences in the next section.

In order to achieve certain data rate and/or quality objectives, encoders must be appropriately controlled. The most typical control scheme is *Constant Bit Rate (CBR)*, where the rate of the encoded video is kept constant at all times by dynamically adjusting the encoder parameters (typically the quantizer scale). This is achieved by placing a buffer at the output of the encoder, and using the buffer occupancy level as feedback to dynamically adjust the quantization scale [3, 4]. Another scheme is simply to keep the quantizer scale at a constant value at all times. This scheme is referred to as *Open-Loop Variable Bit Rate (OL-VBR)*. A third possibility is to use a feedback control mechanism to adjust the encoder parameters so as to maintain a constant level of quality [5, 6, 7]. This scheme is referred to as *Constant-Quality VBR (CQ-VBR)*. The video encoder control scheme has an important effect on the resulting traffic characteristics [7]. For example, while the CQ-VBR scheme can maintain a target level of quality using fewer bits than the CBR scheme, it produces more bursty traffic. Therefore, the encoder control scheme is expected to affect the glitch rate for a given average network load. We show in the next section that this is indeed the case; we also show that the glitch duration and spatial extent are not very much dependent on the control scheme.

When the video encoder control scheme is not CBR, the resulting traffic characteristics also significantly depend on the video content [8, 9, 6, 7]. Therefore, the video content is expected to have a significant effect on the glitch rate. In the next section, we show that this is indeed the case. As for the glitch spatial extent and duration, we show that the effect of video content is not significant.

3 Case Studies for ATM and 100Base-T

Consider a scenario consisting of a number of video sources transmitting video data to the same number of respective receivers, whereby there is a delay constraint of D_{max} from when the video is captured until it is displayed. Any data that arrives at the receiver in excess of the delay constraint is considered lost. We consider that the encoder and decoder are streamlined to the full extent possible in order to minimize the end-to-end delay; thus, they operate on a macroblock-by-macroblock basis. Therefore, the encoder and decoder delays are small, and they can be neglected. The end-to-end delay then consists of packet formation delay, and the network delay. The packet formation delay $D_f(n)$ is defined as the delay experienced by the first bit in packet n from the time the bit is generated by the encoder, to the time the packet is formed. The network delay consists of the time spent in queueing, contention for the channel, transmission, and propagation.

Here, we use two video sequences for traffic sources, each of them one-minute long. The first sequence is taken from the movie *Star Trek VI*; it contains a large amount of motion, and 29 scene cuts. The second sequence is of videoconferencing type, where a person sits and speaks in front of a camera. The video is digitized at SIF resolution; hence, there are 330 macroblocks in a frame. The frame rate is 30 per second. We have encoded the sequences using H.261 and MPEG-1. For MPEG-1, we have employed two GOP structures: (i) "IBBPBBPBBPBBI...," which we refer to as GOP Structure 1 (GS1), and (ii) "IPPPPPPPPPPPI...," which we refer to as GOP Structure 2 (GS2). For both MPEG-1 and H.261, we have employed the CBR and the CQ-VBR encoder control schemes. For the CBR scheme, the bit rate is chosen as 1.536 Mb/s, and the rate control buffer size is chosen as 153.6 kbits. The CQ-VBR encoder is operated at a target quality value of 4.5 on a scale of 1 to 5 (degradations rated at 1 being very annoying, and rated at 5 being imperceptible). For H.261, this resulted in an average rate of 640 kb/s for the Star Trek sequence, and 340 kb/s for the videoconferencing sequence. For MPEG-1 with GS1, the resulting average rates were 627 kb/s for the Star Trek sequence, and 384 kb/s for the videoconferencing sequence.

For a given simulation run, one of the encoded sequences is used to generate the traffic in all the sources, using a random starting frame in each source in order to have a more realistic traffic scenario. Glitch statistics are derived from packet loss by taking into account the dependencies among the macroblocks within frames.

As far as the network types are concerned, we consider a 100Base-T Ethernet Segment and an ATM multiplexer, which are described in more detail below.

3.1 100Base-T Ethernet Segment

When video is sent over an Ethernet, the packetization process and the packet size play an important role in the resulting performance [10, 11]. Here we use the following packetization process. At any time, if there are 1500 bytes (i.e., Ethernet's maximum packet size) of data available, then that data is packetized and queued for transmission. If fewer than 1500 bytes have been generated by the encoder when a predetermined time T_f has elapsed after the generation of the last packet, then all those bytes generated are packetized and queued for transmission. It is important to note that there is a tradeoff between T_f and the network delay. When T_f is small, the packet formation time will be small, but there will be many small packets contending for the shared channel, and the network delay will be large. When T_f is large, there will be fewer packets, and the network delay will be smaller, but the packet formation time will be large. Therefore, T_f is an important design parameter, representing a tradeoff between packet formation time and the network delay.

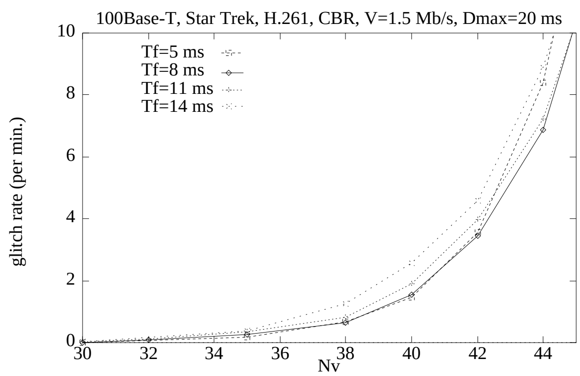
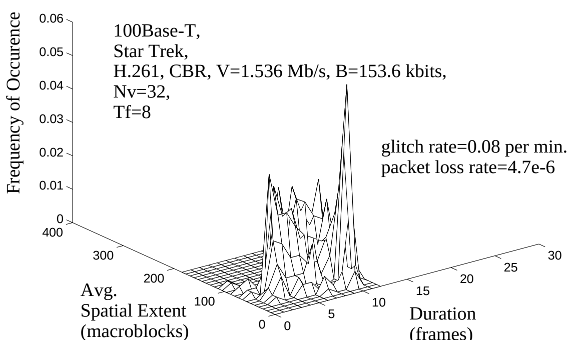


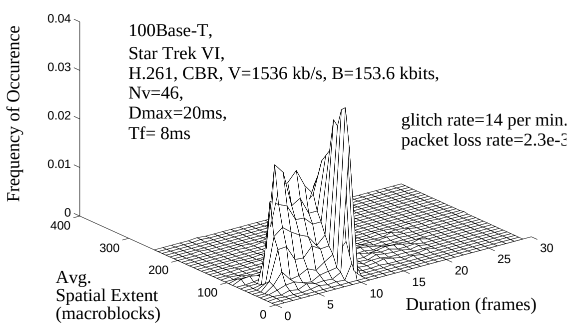
Figure 1: Glitch rate as a function of the number of video streams in a 100Base-T Segment, Star Trek sequence, H.261, CBR, D_{max}=20 ms.

In Figure 1, we show the glitch rate as a function of the number of video streams for various values of T_f from 5 ms to 14 ms for the Star Trek sequence, CBR encoded using H.261. It is clear that the best performance is achieved when T_f=5 ms and T_f=8 ms. In all the scenarios considered in this paper where D_{max}=20 ms, we have observed that T_f=8 ms gives the best performance for all the N_v values considered. Therefore, the remainder of the results are shown for T_f=8 ms when D_{max}=20 ms. The Figure 1 also indicates that the glitch rate versus N_v curve exhibit a knee around N_v=40. For reference, a glitch rate of about 1 per minute occurs when N_v=39, corresponding to a network utilization of about 60%.

In Figure 2, we plot the histogram of the average glitch extent and the glitch duration for N_v equal to 32 and 46 for the same scenario as above. It is interesting to note that the histograms for the two values of N_v look very similar, even though the resulting glitch rates are more than two orders of magnitude different between N_v=32 and N_v=46. Note that when N_v=46, the glitch rate is 14 per minute; this is clearly beyond the load level where the networks would be operated. Therefore, the glitch duration and spatial extent statistics do *not* depend on the number of video streams (or equivalently the network load) for all the practical values of the network load. This was observed in all the cases we have considered. Also observe in the figure that most of the glitches have a duration less than or equal to 10 frames. For example, when N_v=46, 58% of the glitches have a duration less than 10 frames, 31% of the glitches have a duration equal to 10 frames, and 1.8% of the glitches have a duration greater than 15 frames. Therefore, glitches with a long duration (which occur due to losing parts of data in different frames corresponding to different regions) are rare. This justifies our definition of a glitch, where the loss of data in consecutive frames constitute the same glitch, regardless of the shape of the affected regions in those frames.



(a) N_v=32



(b) N_v=46

Figure 2: Histogram of the glitch duration and spatial extent for a 100Base-T Segment, H.261, CBR, D_{max}=20 ms.

For large D_{max} (say, 100 ms or more), the optimum T_f values will be greater; therefore, the expected packet size will be greater. Consider for example D_{max}=500 ms. For that case, we have determined that the best packetization strategy is to choose a very large T_f (e.g., 100 ms), so that every packet is a full-size Ethernet packet. In Figure 3, we show the histogram of the average glitch extent and the glitch duration for the H.261 CBR en-

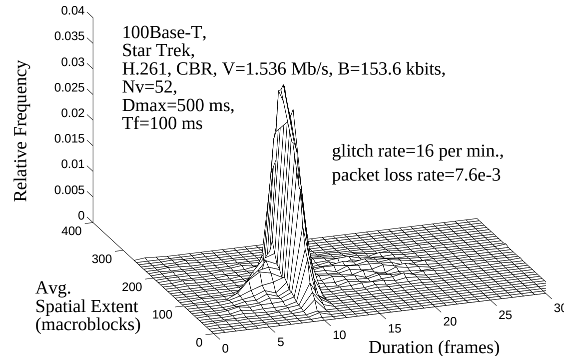


Figure 3: Histogram of the glitch duration and spatial extent for a 100Base-T Segment, H.261, CBR, D_{max}=500 ms.

coded Star Trek sequence when D_{max}=500 ms, and N_v=52. The figure indicates that both the glitch spatial extent and the glitch duration increased when D_{max}=500 ms compared to D_{max}=20 ms. The reason for the increase in the average glitch spatial extent is the larger packet size used. As discussed in the previous section, in H.261, when the glitch spatial extent is increased, the glitch duration also tends to increase. In this particular case, 15% of the glitches have a duration less than 10 frames, 46% of the glitches have a duration equal to 10 frames, and 6% of the glitches have a duration greater than 15 frames.

Now consider the effect of the encoder control scheme. As far as the glitch duration and spatial extent statistics are concerned, we have observed that in general, the CQ-VBR and CBR encoded sequences produce very similar results. As for the glitch rate, consider the H.261 CQ-VBR encoded Star Trek sequence at $\hat{s}_{target}$=4.5 (therefore resulting in an average rate of 640 kb/s), for D_{max}=20 ms. For this scenario, when the network load reaches about 52%, the glitch rate becomes around 1 per minute. Therefore, even though in this case 82 streams can be carried by the network compared to the 39 streams in the CBR case, the network utilization for the CQ-VBR is somewhat lower. This is explained by the more bursty nature of the CQ-VBR traffic.

We have also encoded and simulated the Videoconferencing sequence under the same CQ-VBR scenario. In this case, up to 164 streams can be supported by the network before a glitch rate of about 1 per minute occurs. This difference is due to the smaller average rate resulting from CQ-VBR encoding the videoconferencing sequence compared to the Star Trek sequence. As far as the glitch spatial extent and duration are concerned, in Figure 4 we plot their histogram when N_v=150. Here too, the histogram is not much different from the corresponding histograms for the Star Trek sequence, although there are somewhat more glitches with a duration smaller than 10 frames. This is mainly because of the smaller data rate of the Videoconferencing sequence, forcing smaller packets to be used for the same T_f. In general, we have seen that the dependence of glitch duration and spatial extent on the video content is weak.

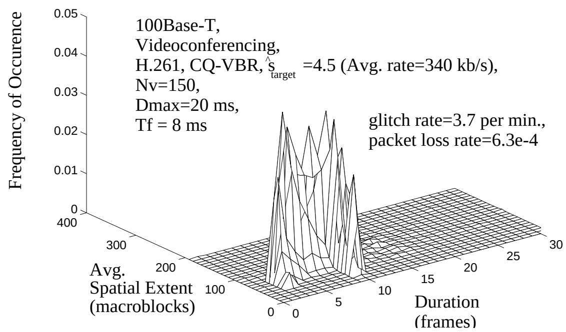


Figure 4: Histogram of the glitch duration and spatial extent for a 100Base-T Segment, Videoconferencing sequence, H.261, CBR, D_{max}=20 ms.

Now consider the effect of the video encoding scheme. In Figure 5, we show the glitch spatial extent and duration histogram for the Star Trek sequence, CBR encoded using MPEG-1 GS1 scheme, when D_{max}=120 ms. (Note that for the GS1 scheme, since the frames are transmitted out of order, there is a delay of 3 frame times (i.e., 100 ms) in the encoder and the decoder. Therefore, the network delay is equal to D_{max}−100 ms, in this case, 20 ms.) It is interesting to note that about 60% of all glitches have a duration of 1 or 2 frames. These are due to the packet losses which occur during the transmission of B frames. Most of the other glitches have durations 5,8,11, or 14 frames. These are caused by the packet losses which occur during the transmission of an I or P frame; such losses affect the previous 2 B frames, as well as all the subsequent frames until the next I frame. For a given glitch rate, this pattern of glitch durations is better than the one for H.261, since the glitches of duration 1–2 frames are likely to cause little or no perceived quality degradation.

In Figure 6 we plot the corresponding histogram for the MPEG-1 GS2 scheme. Here the histogram looks quite different from the GS1 case, since there are no B frames, and whenever there is a glitch, it persists until the next I frame. Therefore the glitch durations are mostly distributed between 1 and 12.

3.2 ATM Multiplexer

We consider that the encoded video data is placed into ATM cells as soon as it is generated, regardless of the macroblock boundaries. As soon as a cell is full, it is sent to the multiplexer. Note that, due to the small size of the cells, the cell formation time is on average about 0.6 millisecond for the video sequence that we consider, which is fairly small.

The ATM multiplexer is operated using the FIFO

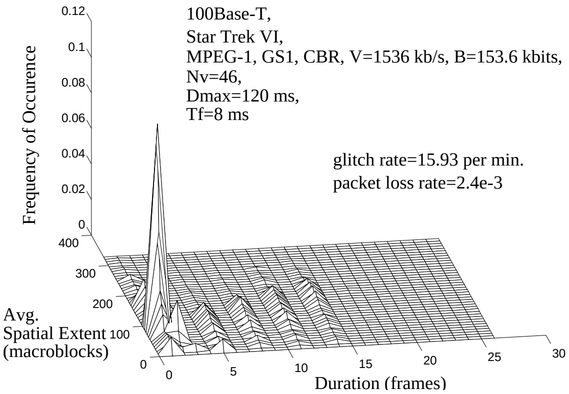


Figure 5: Histogram of the glitch duration and spatial extent for a 100Base-T Segment, Star Trek sequence, MPEG-1, GS1, CBR, D_{max}=120 ms.

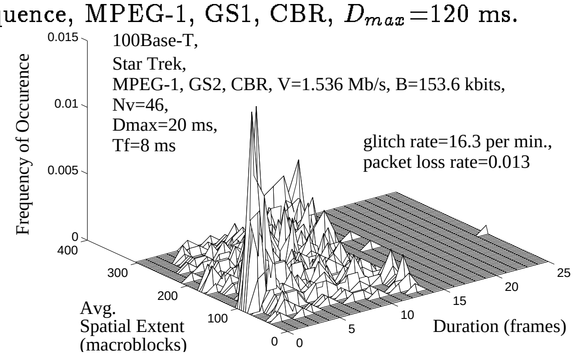


Figure 6: Histogram of the glitch duration and spatial extent for a 100Base-T Segment, Star Trek sequence, MPEG-1, GS2, CBR, D_{max}=20 ms.

scheduling discipline, and its output is fed into a channel with a capacity W of 100 Mb/s. Given that D_f is very small compared to D_{max}, the multiplexer buffer size M_{max} is chosen such that $M_{max}/W = D_{max}$. Cells that arrive when the multiplexer buffer is full are dropped.

In Figure 7, we show the glitch rate as a function of N_v for the Star Trek sequence, CBR encoded using H.261, and for D_{max}=20 ms. For this scenario, a glitch rate of 1 per minute occurs when N_v=50, as compared to N_v=39 for 100Base-T.

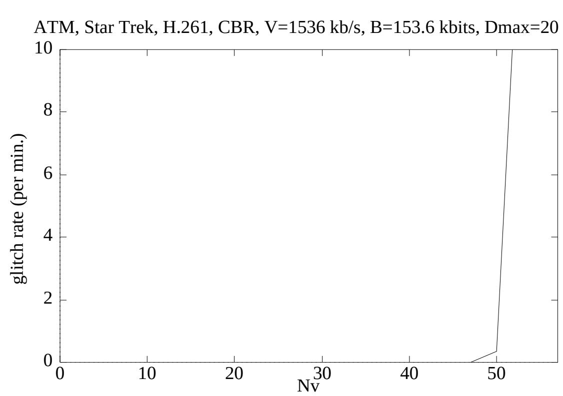


Figure 7: Glitch rate as a function of N_v for ATM, Star Trek VI, H.261, CBR, D_{max}=20 ms.

In the previous subsection, we have established that for 100Base-T, the main element that affects the glitch duration and spatial extent statistics is the video encoding scheme. Our ATM simulations show the same result as well. In Figure 8, we plot the glitch spatial extent and duration histogram for the ATM multiplexer, the CBR encoded Star Trek sequence using H.261, for $N_v=50$, and $D_{max}=20$ ms. The corresponding histograms for MPEG-1 GS1 is plotted in Figure 9. We have observed that for all of H.261, MPEG-1 GS1, and MPEG-1 GS2, the glitch spatial extent statistics are significantly smaller in ATM as compared to 100Base-T. The smaller spatial extent in ATM is clearly due to the small cell size. As for the glitch duration, for MPEG-1, the statistics are not very different in the two networks. For H.261, there are more glitches with small durations in ATM as compared to the Ethernet, as a result of the dependence of glitch duration on the spatial extent in H.261.

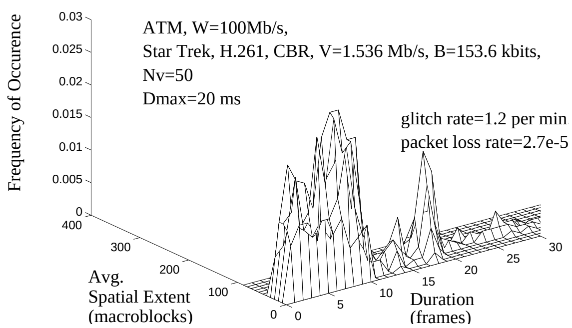


Figure 8: Histogram of the glitch duration and spatial extent for a an ATM multiplexer, Star Trek sequence, H.261, CBR, $D_{max}=20$ ms.

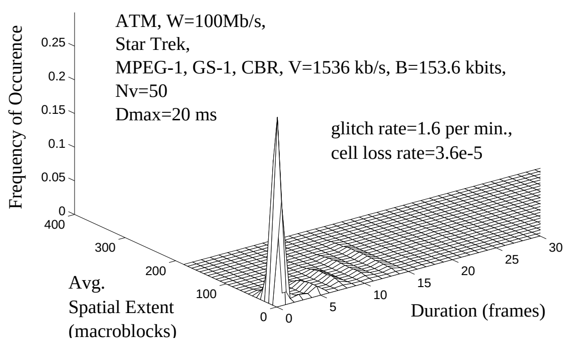


Figure 9: Histogram of the glitch duration and spatial extent for a an ATM multiplexer, Star Trek sequence, MPEG-1, GS1, CBR, $D_{max}=20$ ms.

4 Conclusions

We have used glitch statistics (duration, spatial extent, and rate) to evaluate the network performance when carrying video traffic. We have presented numerical results for ATM and 100Base-T Ethernets, obtained by simulating these networks using real video traces obtained by using MPEG-1 and H.261 video encoders, controlled for achieving either constant bit rate or constant quality. We have shown that the glitch rate is mainly affected by the network type, traffic load, encoder control scheme, video content, and the end-to-end delay constraint. The glitch spatial extent and duration are mainly affected by the network type and the video encoding scheme, and to a smaller degree, video content and the end-to-end delay constraint. We are currently working on a comprehensive evaluation and comparison of various networking technologies using glitch statistics as the network performance measure.

References

[1] "Video CODEC for Audiovisual Services at p×64 kbit/s," ITU-T Rec. H.261,Geneva,1990.

[2] "ISO/IEC JTC1/SC29, Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbits/s," ISO, 1991.

[3] "Description of the Reference Model 8," CCITT SG XV. Spec. Group on Coding for Visual Telephony, May 1989.

[4] "MPEG Video Simulation Model Three (SM3)," ISO-IEC/JTC1/SC2/WG8, 1990.

[5] İ. Dalgıç and F. A. Tobagi, "Constant Quality Video Encoding," in *IEEE ICC'95*, (Seattle, Washington, USA), pp. 1255–1261, June 1995.

[6] İ. Dalgıç and F. A. Tobagi, "A Constant Quality MPEG-1 Video Encoding Scheme and its Traffic Characterization," *PCS'96*, Melbourne, Australia, March 1996.

[7] İ. Dalgıç and F. A. Tobagi, "Characterization of Video Traffic and Video Quality for Various Video Encoding Schemes and Various Encoder Control Schemes," Computer Systems Laboratory Tech. Rep., Stanford U. (in preparation).

[8] D. Heyman and T. Lakshman, "Source models for VBR broadcast-video traffic," in *IEEE INFOCOM '94*, pp. 664–671, 1994.

[9] O. Rose, "Statistical Properties of MPEG Video Traffic and Their Impact on Traffic Modeling in ATM Systems," Research Report Series 101, University of Wurzburg, Inst. Comp. Sci., Feb. 1995.

[10] İ. Dalgıç, W. Chien, and F. A. Tobagi, "Evaluation of 10Base-T and 100Base-T Ethernets Carrying Video, Audio and Data Traffic," in *IEEE INFOCOM '94*, (Toronto, Canada), pp. 1094–1102, June 1994.

[11] F. A. Tobagi and İ. Dalgıç, "Evaluation of 10Base-T and 100Base-T Ethernets Carrying Constant Bit Rate Video Traffic," submitted to IEEE JSAC, special issue on Distributed Multimedia Systems and Technology.