Traffic Shaping in an ATM Environment

Prabhakar T V[1], Burkhard Stiller, Thomas Walter
Institut für Technische Informatik und Kommunikationsnetze TIK
ETH Zürich, Gloriastrasse 35, CH – 8092 Zürich, Schweiz
Phone: +41 1 632 7016, FAX: +41 1 632 1035
E-Mail: [stiller | walter] @ tik.ee.ethz.ch

Abstract

Although Traffic Shaping is an optional implementation for users and network operators, it is important when submitting traffic to public ATM networks. Also, shaping may assist the operators in dimensioning the network more cost-effectively. However, today's ATM end device sources like video codecs, do very little to address this feature and as a result the transmitted ATM cells are either in bursts or well spread out. We take a look at how such end device sources are suitable for traffic shaping at ATM switches. In particular, we look at issues concerning traffic shaping of audio and video streams on a single physical ATM switch port. A detailed investigation on a FORE Systems™ ATM switch is carried out to conclude why such a scheme may not be possible in our current ATM environment.

1. Introduction

At present, for an Asynchronous Transfer Mode (ATM) based teleteaching or distance education scenarios, the networking and multimedia equipment comprises of systems such as the ATM switch, Video and Audio Codecs, cameras, TV systems, microphones, audio and video mixers etc.. The traffic handled by the ATM switch in such a setup comprises of two or more digital video and audio streams. As an example, we could have the "lecturer" and "classroom" video and audio streams of the local site switched to the remote site and the "audience" video and audio stream of the remote site switched to the local site as, for instance, in the Telepoly scenario /WSHP97/. Since a network is required to transport in real time the various above mentioned multimedia streams, we make use of public ATM network service providers. As a result of this, the traffic crosses one or more local administrative domains (local ATM LAN), and at least one external administrative domain (ATM WAN). Most often, the external administrative domains are government agencies and are referred to as PTT's. However, private telecom companies may also provide end users such facilities. Therefore, service providers are the key players in the transport of real time multimedia traffic.

Since the service providers charge customers for the services offered in carrying traffic, a customer is required to declare a priori the type of traffic likely to be generated. In other words, one should not only know the number of video and audio streams that are to be handled, but also the traffic pattern or burstiness of the traffic generated by the different sources. Based on this, the service providers fix the various parameters such as peak cell rate, cell delay variation tolerance, sustainable cell rate, and burst tolerance of the traffic on their switch. Such a situation implies that input traffic should comply to the negotiated parameters. If input sources do not comply thisresults in most cases in the loss of precious multimedia data. As an alternative, the network may agree to carry the exceeded traffic, because more traffic means more income. With an associated risk, it may drop this traffic, if a congestion occurs and, therefore, such a system does not guarantee loss of data only, but may also attract high penalties for violation. However, violation of negotiated data rates may be unintentional, since the cause could be an

1. Author's current address: Centre for Electronics and Design Technology, Indian Institute of Science, Bangalore 560 012, India, E-mail: tvprabs@cedt.iisc.ernet.in

inherent design property of the source or a sudden misbehaviour of the source due to faulty components or malfunctioning of the local switch. This brings us to traffic shaping of input streams to ensure a disciplined input to public ATM networks.

1.1 Problem Statement

Specific to the hardware in use (FORE Systems™ ATM ASX-1000 and K-NET's CellStack™ video codecs), present video conference transmissions shape video (one or more streams) on one physical port on the ATM switch and audio on another port (cf. Figure 1). Since such a

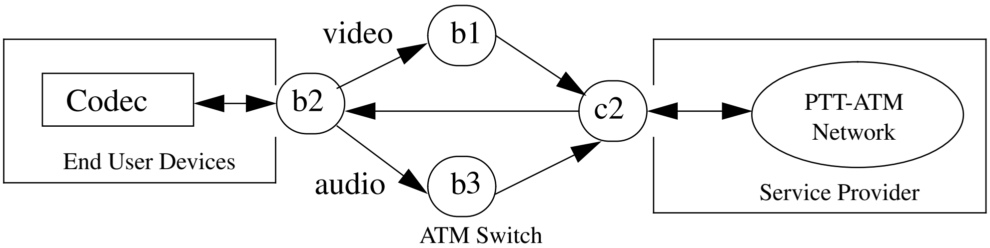

FIGURE 1. Present Scenario for Traffic Shaping Video and Audio Streams

setup wastes resources in terms of physical ports of our local ATM network, we explored the possibility of shaping two outgoing virtual paths (VP) using a single physical ATM port to shape both video and audio streams. Traffic shaping audio and video streams using a single physical port, results in a scrambled output. The video picture is lossy and the audio is jarred. Figure 1 indicates the existing scenario for shaping the video and audio streams. It uses two ATM switch ports (b1, b3) for shaping video and audio streams. Figure 2 indicates the scenario where an attempt is made to shape both audio and video using a single switch port (b1). Port c2 in both cases connects to the service provider network. No shaping is required for input streams (we mean streams coming into the local network) and therefore are directly switched to the concerned port (b2).

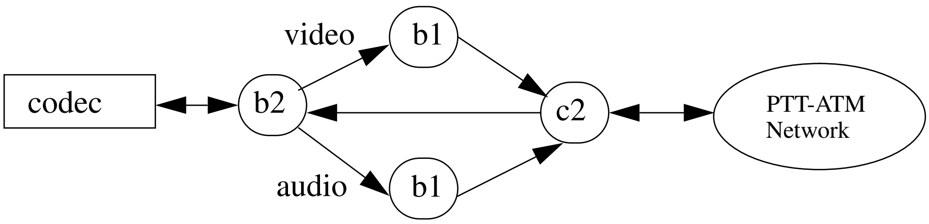

FIGURE 2. Problem statement Scenario for Traffic Shaping Video and Audio Streams

Since compressed digital video is a variable bit rate (VBR) source, and our setup handles video using Motion JPEG (Joint Photographic Expert Group) compression scheme, we take a look at some important aspects affecting its bandwidth requirements. However, the ATM service provider supports only CBR like traffic. As for the digital audio, since it is handled in an uncompressed form, it nataurally offers a constant bit rate signal. Traffic shaping is applied on outgoing VPs to achieve a CBR stream.

1.2 Outline of the Report

This report on "Traffic Shaping in an ATM Environment", is organized as follows. Initially, Section 2 presents the technical networking environment which contains an overview of an ATM network from traffic shaping view point. The section also includes the definition of traffic shaping and its consequences, a brief overview of digital video and MJPEG compression scheme. Section 3 contains the experimental setup applied. Within Section 4, all investigated scenarios are presented. Section 5 contains the summary of results. Section 6 has the discussions of the various investigated scenarios. Section 7 comprises of conclusions for shaping multimedia streams on an ATM port. Finally, Section 8 has the list of references.

2. Technical Networking Environment

The followin briefly introduces the essential features of ATM, traffic shaping, and audio and vidoe communication. For further details the reader is referred to /HHSc94/ and /Pryk95/ on ATM, /YSSa92/ on traffic shaping, and /StNa95/ on multimedia communication.

2.1 ATM network operation

In ATM, the data transfer unit, called cell, is fixed at 53 Byte. The payload may contain a maximum of 48 Byte and 5 Byte is used as header. A host connects to an ATM switch using an interface referred to as User Network Interface (UNI). The term User in UNI, is normally an end-system host and the term Network refers to the ATM switch. Since the UNI interface standards have evolved over the years, ITU-T Q.2931, UNI 3.0, UNI 3.1, and presently UNI 4.0 is developed. We also have proprietary standards from switch manufacturers. One such standard is Simple Protocol for ATM Network Signalling or SPANS™, from FORE Systems™.

ATM connects hosts in a point-to-point fashion. If a host A (calling host) requires to connect to host B (called host), it first attempts to setup a call to B, using the UNI interface signalling protocol. Apart from situations where the network and host B are busy, the success of a call depends on various parameters declared by host A. These parameters need to be negotiated and agreed upon by all parties involved, namely, host A, ATM switch and host B. The parameters include the maximum instantaneous cell rate (also called the peak cell rate or PCR), its allowed margin of variation - Cell Delay Variation Tolerance (CDVT), the Sustainable Cell Rate (SCR) and its allowed margin of variation - Mean Burst Size (MBS). When connections are set up using the above mentioned parameters, the network is now required to guarantee these connections the negotiated bandwidth. Also, in addition to guaranteeing network bandwidth, the ATM network may also require to deliver data to destinations within a specified delay, or, the cells need to arrive with strict timing constraints as required by the application. In other words, the network may be required to deliver data in real time for any usefulness. Therefore, cell delay variation (CDV) and its bounds also needs to be addressed. What is now expected of the network is a defined Quality-of-Service (QoS), addressing these issues on bandwidth and delay.

After the call succeeds, a switched virtual circuit (SVC) connection is established and the routing information is included in the header of the cell in terms of Virtual Path Identifier (VPI) and Virtual Channel Indentifier (VCI). The VPI identifies the Virtual Path Connection (VPC) which contains one or more virtual connections (VCs). Each Virtual Channel Connection (VCC) in the path is identified by a VCI. Every cell originating from host A and terminating at host B always traverses a fixed path or route.

The next issue handled by the ATM switch concerns policing the above discussed parameters on bandwidth and delay. Such a policing mechanism is referred to as Usage Parameter Control (UPC), which works at the UNI interface. A number of UPC algorithms have been proposed, each having distinct advantage over the other. One of the effective and popular mechanisms presently, seems to be the Leaky Bucket UPC policing algorithm. The input parameters for policing include the PCR, CDVT, SCR, and MBS. Cells conforming to the negotiated traffic are accepted and routed. Non-conforming cells are either marked in the header using the Cell Loss Priority (CLP) bit, or dropped by the network depending on what was negotiated during the call setup phase. Table 1 shows the UPC parameters which are required for setting CBR and VBR traffic. Based on the set parameters, the algorithm may police data using a single (for CBR) or a dual leaky bucket (for VBR).

Finally, after the completion of data transfer from host A to host B, the connection is closed.

As opposed to the above mentioned scheme of call setup and connection by signalling, one could also establish a connection between two hosts by manually configuring the VPC and VCC on the switch. Such a connection is referred as permanent virtual connection or PVC.

TABLE 1. Required UPC Parameters for CBR and VBR Traffic

Traffic	Example	Typical Characteristics	Bandwidth Required	PCR	CDVT	MBR	BT
CBR	Audio	Strict end-to-end timing required	2 Mbit/s	Yes	Yes	No	No
VBR	Video	Strict end-to-end timing required	8 to 16 Mbit/s	Yes	Yes	Yes	Yes

2.2 Definition of Traffic Shaping

Traffic shaping may be defined as a method of spacing in time all cells emitted into a VCC or VPC to attain the negotiated traffic characteristics of a connection. Traffic shaping may also be regarded as actively altering the traffic characteristics in order to comply with the negotiated traffic characteristics. This is done by smoothing out peaks in cell streams at the expense of adding more delay and limiting the burst length /CuSa93/. Traffic shaping may be implemented at either the ATM end device source or at the ATM switch. When traffic flows are shaped at the source, cells emitted into a VCC are spaced and if done at the switch, cells originating from a VCC are shaped /Pryk95/. Traffic shaping shares a "cause and effect" relationship. While the cause to introduce traffic shapers in a network is to ensure that faulty and misbehaving sources need to be checked for their cell emission and, therefore, meet the negotiated peak cell rate, the effect is, it partially compensates for the effects of CDV on the peak cell rate.

What we mean by the above sentence is illustrated by the following example. Consider the cells arriving at an ATM switch, experience random delays due to physical layer functions in the terminal equipment, medium access controls in the customer equipment, and ATM multiplexing in the network infrastructure /BGSC92/. Some cells experience a larger delay compared to previously arrived cells, resulting in larger interarrival time for these cells. Such a situation may be termed as "Cell Dispersion". Another situation may also exist wherein cells experience a shorter delay compared to previous cells and therefore the interarrival time for such cells may be less. Such a situation may be referred to as "Cell Clumping" /BGSC92/. By shaping (spacing) traffic flows one can easily compensate the Cell Delay Variation (CDV) which includes cell dispersion and clumping. It is important for traffic shaping to maintain a cell sequence integrity of the connection. Figure 3 shows the functional block diagram of traffic shaping. The traffic shaper smoothes the input cell traffic to reduce its burstiness. Shaping is done by inserting empty cells at the physical layer which in effect delays the user payload carrying cells. The delay may therefore be controlled by number of empty cells inserted into the traffic stream. At the receiving side, the empty cells are discarded at the physical layer.

FIGURE 3. Functional Block Diagram of Traffic Shaping

Figure 4 shows the algorithm applied for traffic shaping. Having known the rate at which the cells need to be emitted (T_o), the shaper has knowledge of the departure time of the previous cell $td_{(i-1)}$ and having acquired the time of arrival of the cell to be emitted, the shaper can easily decide either to delay (insert empty cells) or emit the cell into the concerned VCC /YSSa92/.

It is important to note that in the traffic shaping algorithm, the peak cell rate is used only to fix the interarrival time of cells. Also, the mean cell rate or the average rate of cells is left undisturbed. Therefore, cells are unlikely to be dropped and this provides an efficient method to control congestion in the network, particularly when the traffic has very short spikes. This is slightly different from GCRA (Generic Cell Rate Algorithm) and in particular the leaky bucket algorithm where PCR, CDVT, SCR and MBS may be applied. Further, the function of the UPC at a UNI is to check for conformance of the traffic and dealing with non conforming cells. Therefore, the function of a UPC is to "traffic police" the data and not "traffic shape" the data.

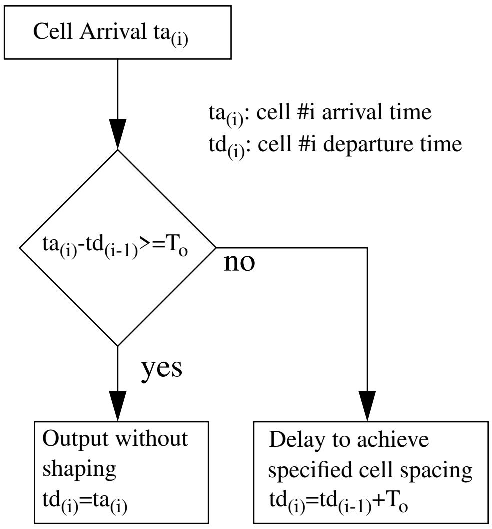

FIGURE 4. Traffic Shaping Flow Chart

2.3 Digital Video and Compression - An Overview

Presently, video signals captured from familiar equipment like video cameras are analog by nature and therefore need to be translated into digital form. Such a translation process is known as Digitization. Digitization involves two processes. The first is referred as sampling and the other is quantisation. As an example, sampling of audio refers to one dimensional discrete equal time intervals, and in the case of video, sampling refers to two dimensional object space being cut into smaller rectangular regions. As regards to quantisation, it refers to assignment of integers based on the amplitude for the audio intervals and smaller regions of video, respectively. A digitized image lies on an imaginary regular, rectangular grid and is represented by a matrix of the above mentioned quantised values. One important convention adapted in the computing world is the use of a sampling grid having equal horizontal and vertical sample pitch, called square pixels /Poyn96/.

Colour vision is achieved through proportional intensities of three signals, Red, Green and Blue (R,G and B). Therefore, the value associated with each pixel is a function of information pertaining to intensity or brightness and colour of the image surrounding the corresponding point of the image. In otherwords, there is a luminance (or luma) and chrominance (or chroma) component for every pixel, which is worked out based on empirical measurements on brightness of each colour. While a luminance signal is represented by convention using the letter Y, given by the relation: $Y= 0.30\, R + 0.59\, G + 0.11\, B$, chrominance is represented by two difference signals U (B-Y) given by the relation: $U = (B-Y) * 0.493$ and V (R-Y) is given by the relation: $V = (R-Y) * 0.877$. Another convention is to represent U as Cb and V as Cr. Summing up, any captured image has a spatial resolution (pixels*pixels) and colour encoding (bits/pixel), which encompasses the luminance and chrominance video signals. It should be noted that the above formula is specific for YUV signals /StNa95/.

It is well known that the eye is more sensitive to luminance or brightness changes than to chrominance changes. Therefore, in colour encoding of the images, one could take advantage of reducing the data capacity for the chroma and providing more data capacity or full bandwidth for luma. However, in practice there are both methods, some which do not consider such advantages and some which do. For example, in ‘Composite Coding’ the entire video signal is converted into one digital representation of 8 bits for all the three colours. This is also called pseudo colour. Since there is no conversion of R,G and B to Y, Cb and Cr it is easier than representing the signal as separate signal components. Such a method suffers from interference of luminance with chrominance signals. In the case of the ‘Component coding’ method, we have ‘RGB’ and ‘YUV’ submethods. While RGB codes the R,G and B as separate signals, (8 bits for each colour, totalling 24 bits- also called true colour), it does not consider luma.

The YUV (or YCbCr) encoding by far offers the best solution since the luma and chroma signals are encoded separately. The digital data are transformed using a multiplexing technique (4:2:2), defined in the recommendations of CCIR 601 for digital video. Presently this standard needs to be referred as ITU-R 601 standard for digital video. It should be noted here that CCIR 601 is only one of the standards for digital video. There are other methods which are referred as sampling by ‘square pixels’ and sampling at ‘4 times the colour subcarrier’. Other subsampling methods include (4:4:4) where no subsampling is done, (4:1:1) system where chroma components are subsampled by a factor of 4 horizontally, and (4:2:0) system where chroma are each subsampled in both horizontal and vertical directions.

TABLE 2. Chroma Subsampling in the Ratio 4:2:2

Y0	Y1
Y2	Y3
Cb	
Cb	
Cr	
Cr	

Referring to table 2, in an 8 bit system, if we consider a 2x2 pixel area and apply the (4:2:2) technique, we have Y represented as Y0, Y1, Y2 and Y3. The chroma may be represented as Cb, Cb and Cr, Cr. This is because we have subsampled (or removed by filtering) chroma by a factor of 2 along the horizontal axis. We have the Y component represented by 4 Byte and Cb, Cr represented by 2 Byte each, summing up we get 8 Byte. This results in 16 bits per pixel (total number of bytes equals 8 (or 64 bit) and number of pixels 4 (2x2). Hence 64/4=16). Therefore, by effectively subsampling the chroma we achieve full luma bandwidth and gain 2 Byte each for Cb and Cr, over a 2x2 pixel area.

It is needless to say the gain we obtain over larger pixel areas. The sampling rate in the above discussed (4:2:2) technique is 13.5 MHz for luma and 6.75 MHz for chroma. In the case of digital video PAL, if we assume a uniform 8 bit coding for each of the samples, we get (13.75+6.75+6.75)*8 bit = 216 Mbits/s as the bandwidth required to transmit such a system. It is needless to say that 216 Mbit/s bandwidth for one video stream is not directly usable for any present day network /StNa95/. Therefore, we need to compress the signal and significantly reduce the bandwidth requirements to acceptable levels.

Although bandwidth reduction is one of the main benefits of compression, the gains achieved are many more. The advantages of video coding are discussed at length in /KMHY89/. The benefits include consistent picture quality, multimedia integration and transmission efficiency. Some of the compression techniques used today are JPEG and its minor variant MJPEG, used for still images and H.261 for video conference. Other techniques include MPEG1, MPEG2, MPEG4, together used for both audio and video compression and DVI, which has the flexibility to be used for both still and continuous images.

2.3.1 Motion JPEG (MJPEG) Coding and Compression

Coding and compression of any video signal is a series of processes and therefore it is important to understand the compression scheme from a holistic view point. Some of the major steps are shown in Figure 5.

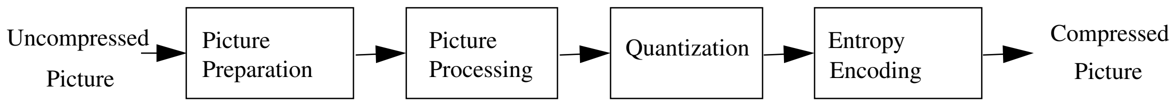

FIGURE 5. Block Diagram Representation of Coding and Compression of Digital Video

In Figure 5, picture preparation includes sampling and quantisation as discussed earlier. One more important function of this block is to divide the image into 8x8 pixel blocks.

Picture processing is the first step towards image compression. It is here that a conversion from the spatial time domain to spatial frequency domain occurs. In the case of JPEG, picture processing does an intraframe coding and compresses the previously prepared 8x8 pixel blocks by taking the redudancies of luma and chroma within the blocks (and not on the image). To put it in simple terms, while doing the coding for the blocks - if the pixels within the block are similar the system achieves more compression compared to blocks where there is large variation. In typical images we do have large parts which are similar in chroma. Examples of such images include trees, sky and wall-background. However, the luma component within these individual images may be different. For example some portion of the tree may be exposed to sunlight and some part of it may be in shade. Therefore, the compression efficiency which in turn affects the variation in the bit rate of the output video signal, is not only dependent on the kind of images one is trying to compress, but also on the luma or brightness of images. In otherwords, for the present, if one neglects the luma component, the difference between the output bit rate of an image with uniform colour will not be substantial compared to an image with varying colours. Summing up, the lesser the variation of luma and chroma within a block, the higher the compression efficiency for the block. One could extend this theory between images as well. For example, let us assume that the camera in a video conference setup presently is focused on the audience. If the camera is panned and zooms on an individual in the audience, there is a variation in the output bit rate since the luma and chroma may increase or decrease as compared to the bit rate of the previous image of complete audience. More examples could include slide presentation in a video conference and sporting event like soccer where the camera is continuously panning. These situations could have high fluctuations in bit rate of video due to varying luma and chroma.

The quantisation block in figure 5 refers to the allocation of bits of the processed block. Here, lower intensity and uniform colour (low luma and chroma) blocks will tend to occupy less bits compared to high intensity blocks. Finally, entropy encoding refers to 2 additional stages of compression. The first one is the well known RLE (Run Length Encoding) encoding. In this scheme, the compression is performed taking the redundancy in data. In otherwords, by replacing repeated byte sequences with the number of occurrences, a substantial reduction in data can be achieved. An example may be ABCCCCCCCCDEFGGG, which is now reduced to ABC!8DEFGGG /StNa95/. In this example, 8 Byte are reduced to 3 Byte. The “!’ is declared as a special flag.

The second and last stage of compression is Huffman Coding. In this scheme, continuing with the above example, if ABC!8DEFGGG is most frequently occurring and then follows say LMN!OPQRSTU etc., codes are alloted as shown in table 3. A variable length coding is applied based on the frequency of occurrences and the shortest code is assigned to numbers that occur frequently.

Concluding, compression introduces variable bit rate signals. The variation in the bit rate of a signal is dependent on the type of compression scheme and the handling of luma and chroma signals of the image. Therefore, for a compressed signal like video, one should bear in mind the above discussion, while fixing the shaping bandwidth based on peak cell rate.

TABLE 3. Huffman Coding: A Simple Example

Bits	Number
1	ABC!8DEFGGG
01	LMN!OPQRSTU

3. Experimental Setup

Bandwidth measurements were conducted using FORE Systems™ network management tool, ForeView (Version 4.2) /FView96/, running in stand alone mode on a SUN workstation. The results are mostly presented in graphical form. The measurement environment for video and audio along with the technical specifications of the equipment is as shown below.

3.1 Video Environment

One of the camcorders were placed in a typical office room, facing a window which opens into a footway with a background of few plants subjected to on and off sunlight. Another camcorder was placed to only face the office room.

3.2 Audio Environment

Audio was tested by speaking into a cordless microphone with the base unit connected to the codec. It has to be mentioned that the microphone generates empty cells at other times.

3.3 ATM Switch

Table 4 shows the specification for the ATM switch. The ATM switch - ASX-1000, presently houses two switching fabrics and series C network modules with four ports each and line speed is 155 MBit/s. C network modules support traffic shaping. The source and input port module used is 'B' with b2 connected to the source video codec and b1 being the shaping port. b1 is placed in diagnostic loopback mode for the purpose of experimentation. Since the module is likely to handle more traffic and queuing due to shaping, the traffic model of the module is set to model 5 as per recommendations of the manufacturers for VP shaping. Also, a PVC is applied between the source and the switch /FRunner96/.

TABLE 4. Hardware in Use - ATM Switch ASX-1000

Equipment	Make and Type	Modules	Signalling Applied	Software Version	AMI Version
ATM switch	Fore, ASX 1000	NM-4/155 MMSCC	PVC	FT410.74	4.1

3.4 Video Codec

Table 5 shows the specification for the audio/video codec. The codecs used are CellStacks™ from K-NET Ltd. They apply MJPEG compression technique and they support a resolution of 720*288 pixels, PAL (25 frames/s) bidirectional. In addition, they also support stereo audio, which is processed independently of the video configuration. The audio is sampled at 44.1 KHz CD quality and no compression is applied. The codecs generate video and audio AAL5 frames in accordance to UNI 3.1 interface implementation. The video produces cell streams which should be classified as conforming to ITU-T class B which is defined for VBR sources where a time relation exists between source and destination /Pryk95/. The physical interface supported by the codecs include 155 Mbit/s SDH/SONET /Knet/.

TABLE 5. Hardware in Use - Video/Audio Codecs

Equipment	Make and Type	Inputs	Outputs	Resolution	Audio Sampling	Com-pression Applied	Cabling
Codec	K-NET, Cell-Stack™	1 Video and 1 Audio	1 Video and 1 Audio	720x576 or 720x288 pixels	44.1 KHz CD, Uncomp-ressed	M-JPEG	155 MBit/s SDH/Sonet

3.5 VP Configuration

While table 6 shows the VPI and VCI numbers applied between codec and ATM switch, table 7 shows the VPI and VCI outgoing numbers used for shaping the streams.

Using the numbers from table 6 and table 7, a permanent virtual connection (PVC) is applied between the codec and ATM switch, for all experimented scenarios.

TABLE 6. Input VPI and VCI for Streams

Stream	VPI	VCI
Audio1	0	200
Audio2	0	210
Video1	0	201
Video2	0	211

TABLE 7. Output VPI and VCI for Streams

Stream	VPI	VCI
Audio1	6	200
Audio2	7	210
Video1	16	201
Video2	17	211

4. Scenarios

The following scenarios are based on certain assumption that are given below. In addition, some general background is explained beforehand.

4.1 Assumptions

The white papers /FArch/ and /FBand/ explain that the ASX-1000 switch implements distributed shared memory and uses *smart buffers* to dynamically allocate memory for virtual channels in support of per VC queuing at the output. In otherwords, series C network modules used in our investigated scenarios, use TDM for distribution to the shared memories. Each network module is responsible for switching and queuing cells destined for its local ports. The network modules queue each VC individually, named per VC queuing, in any one of the five queue sets (one queue set per network module and, in addition, one queue destined for backplane). Further, each queue set consists of four individual priorities. Such priorities are required to isolate high priority traffic from low priority traffic.

Smart Buffers are output buffers used for temporary storage for data travelling to an output port from the switch fabric. One place where Such buffers are essential is, when data travels from a LAN port of 155 Mbps to a WAN port of 45 Mbps. Another important reason to have them is to provide temporary storage for bursty UBR and ABR traffic. For UBR and ABR traffic, smart buffers located in the Series C network modules, are managed and allocated *dynamically* and, therefore, in case of a burst on a port, the entire buffer space could be allocated to a single port. This results in efficient buffer space management. However, for CBR and VBR traffic streams, there is an *automatic* allocation of buffer space.

It is concluded that traffic flows for each VCC is in a separate and distinguishable queue inside the switch. Such queues may be serviced based on queue service methods such as weighted fair queuing or weighted round-robin queuing. As a result of this conclusion, shaping two outgoing VPs on a single port is fairly simple.

4.2 Background

Before we delve into investigations, it is important to understand the difference between cell rejects and cell loss in an ATM environment.

4.2.1 Cell Rejects

Cell rejects normally occur at the UNI interface when they attempt to flow into a VCC connection which is being policed by a UPC contract algorithm. Such a situation may arise when the connection cannot handle any more input from a source due to the policing parameters set. In our scenarios, since we do not investigate with UPC, cell rejects do not occour and therefore are not considered.

4.2.2 Cell Loss

Cell losses occur mostly due to congestion in the network. For instance, a cell may be dropped by the policing algorithm when the CLP bit in the header is turned on. A cell may also be dropped by the switch because of output buffer overflow condition, or when two cells contend for a queue at the same time. Also, cell loss may occur due to queue overflow condition. Another reason for cell loss may be due to noisy or dirty lines, which are mostly recorded as hardware errors.

In our investigated scenarios, cell loss due to set CLP bit does not arise. This is because, as previously stated, we do not investigate with UPC. Also, cell loss due to hardware errors usually does not occour in a LAN environment.

4.3 Shaping One Video and One Audio Stream (One Input VP)

To begin with, a single video and audio stream were attempted for traffic shaping using a single port. Figure 6 shows the setup for the single stream mode. Port b1 is the shaping port for the video and audio stream. Further, b1 is placed in diagnostic loopback mode to return the signals back to the codec.

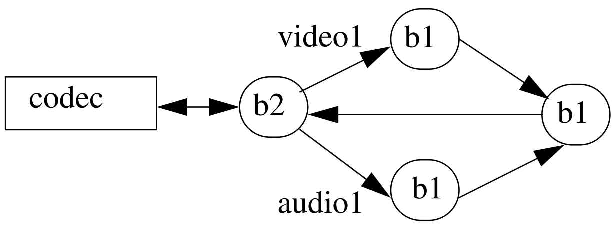

FIGURE 6. Scenario for Traffic Shaping a Single Stream.

Table 8 shows the input, shaping, and output bandwidth measured using the ForeView™ management tool. It should be noted that the video shaping bandwidth is significantly higher to take effects as discussed in Section 2.3. The audio input is uncompressed and, therefore, is a nearly constant bit rate source (cf. Section 2.2).

TABLE 8. Input, Shaping, and Output Bandwidth (Mbit/s)

Input Audio	Input Video	Audio Shaping	Video Shaping	Output Audio	Output Video
1.75	10	2	13	1.25	5

Figure 7 shows uncompressed audio input which almost has a fixed bandwidth of 1.75 Mbit/s. However, there are minor variations around this value, as observed by the dip in the curve.

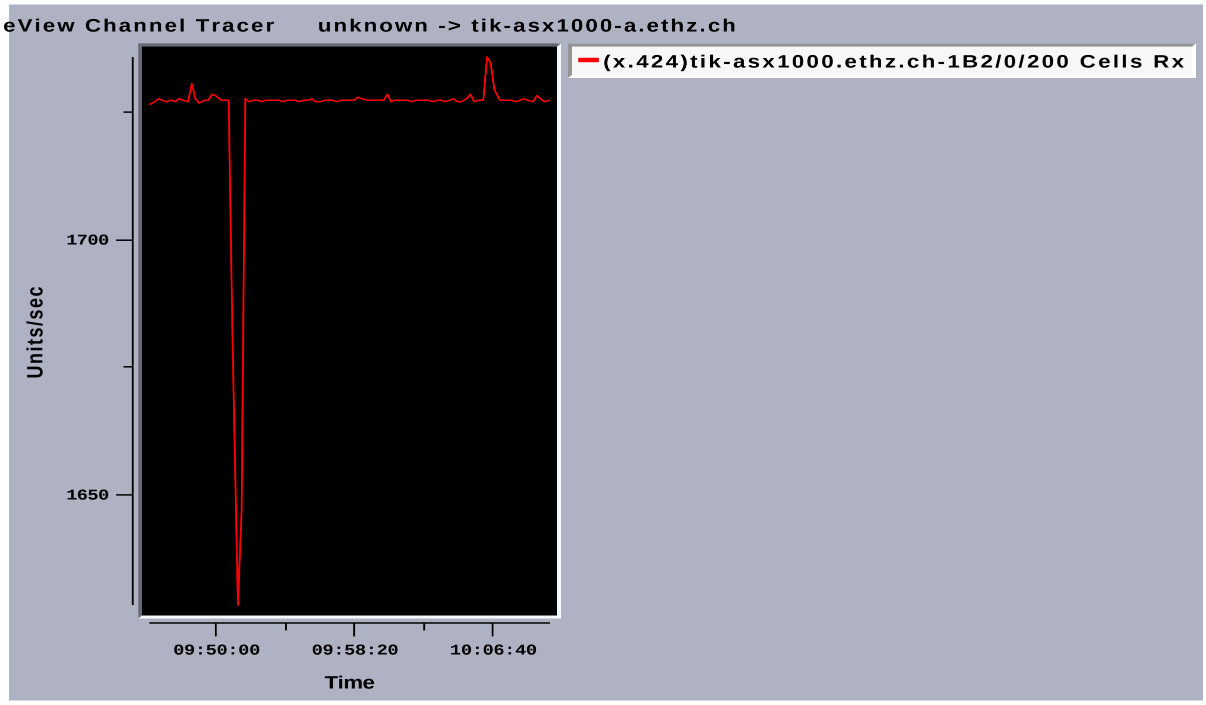

FIGURE 7. Input Audio - Bandwidth (1.75 Mbit/s)

Figure 8 displays the output audio, indicating around 1.25 Mbit/s. This clearly shows the effect of jarred audio. The occasional variation above this value is insignificant since the output required is far higher than achieved.

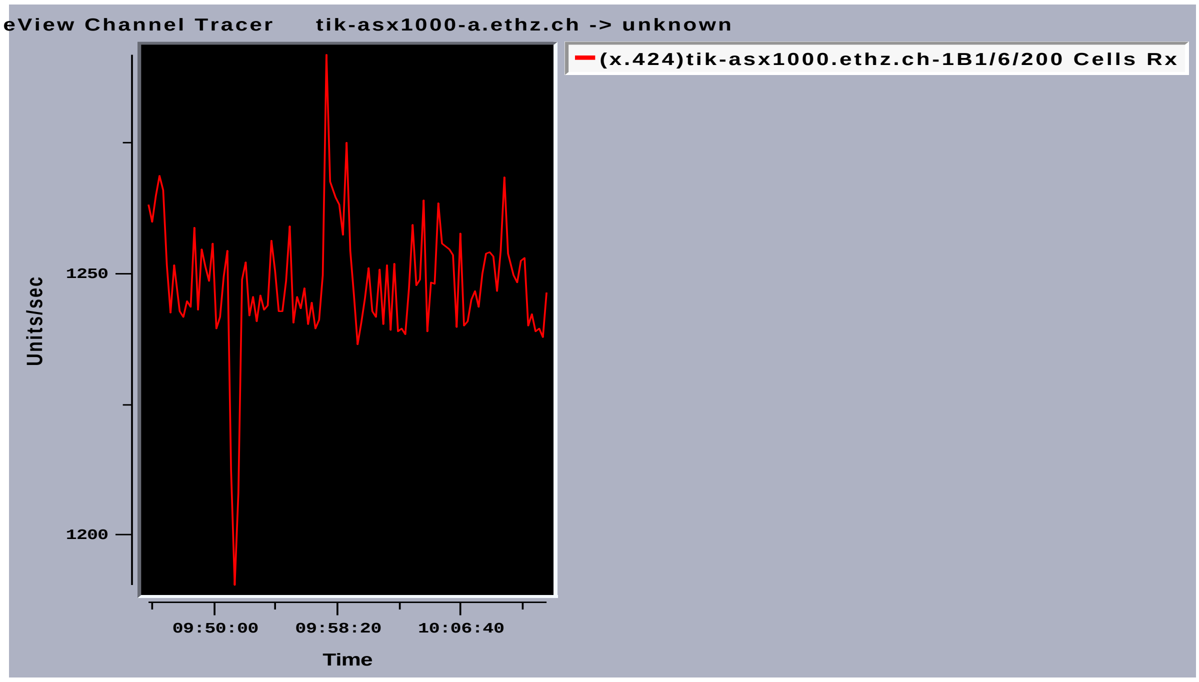

FIGURE 8. Output Audio - Bandwidth (1.25 Mbit/s)

Figure 9 shows a typical VBR input video source subjected to bandwidth variations due to possible effects of luma and chroma. The peak at measurement time was 10 Mbit/s.

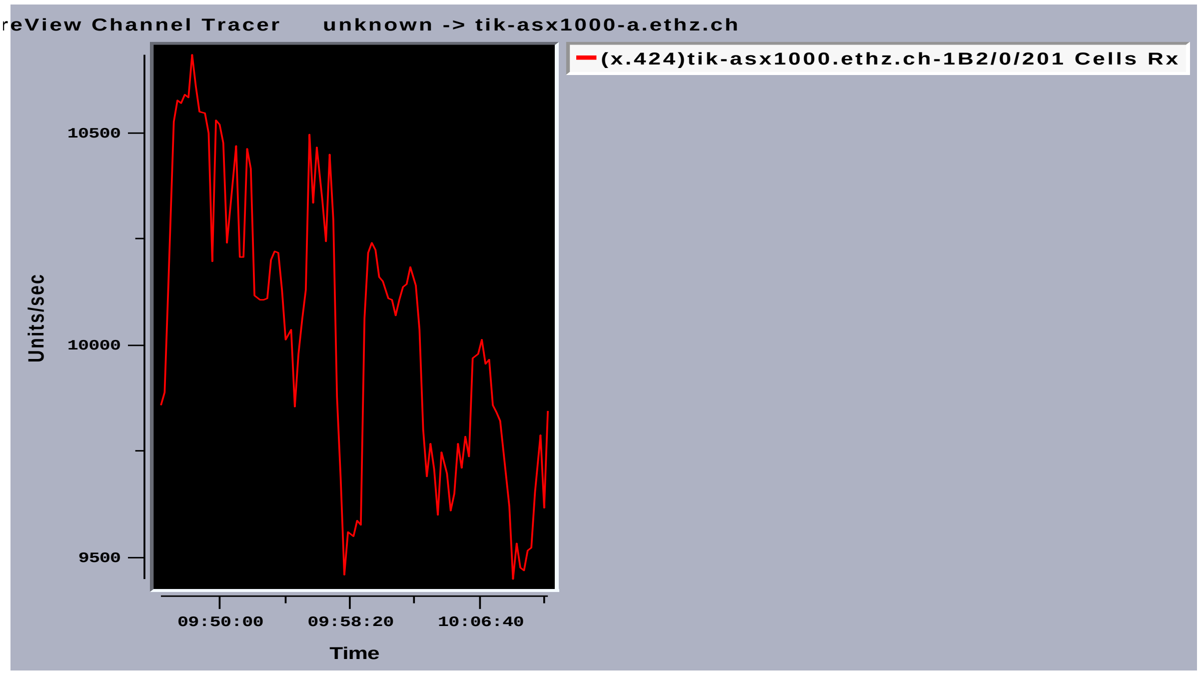

FIGURE 9. Input Video - Bandwidth (10 Mbit/s)

Figure 10 shows the output video. The output variation is around 0.3 Mbit/s. The value at measurement time 5 Mbit/s (the last point in the curve).

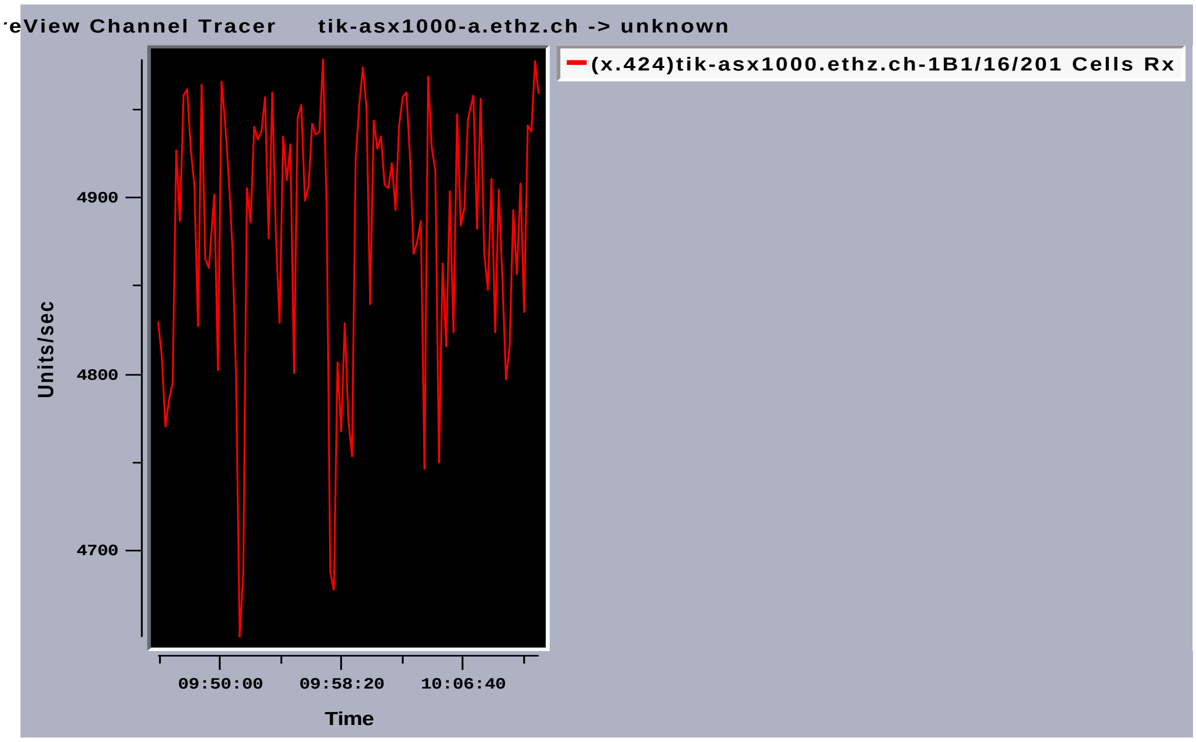

FIGURE 10. Output Video - Bandwidth (5 Mbit)

Figure 11 shows the cell loss for the above discussed single video and audio scenario. These

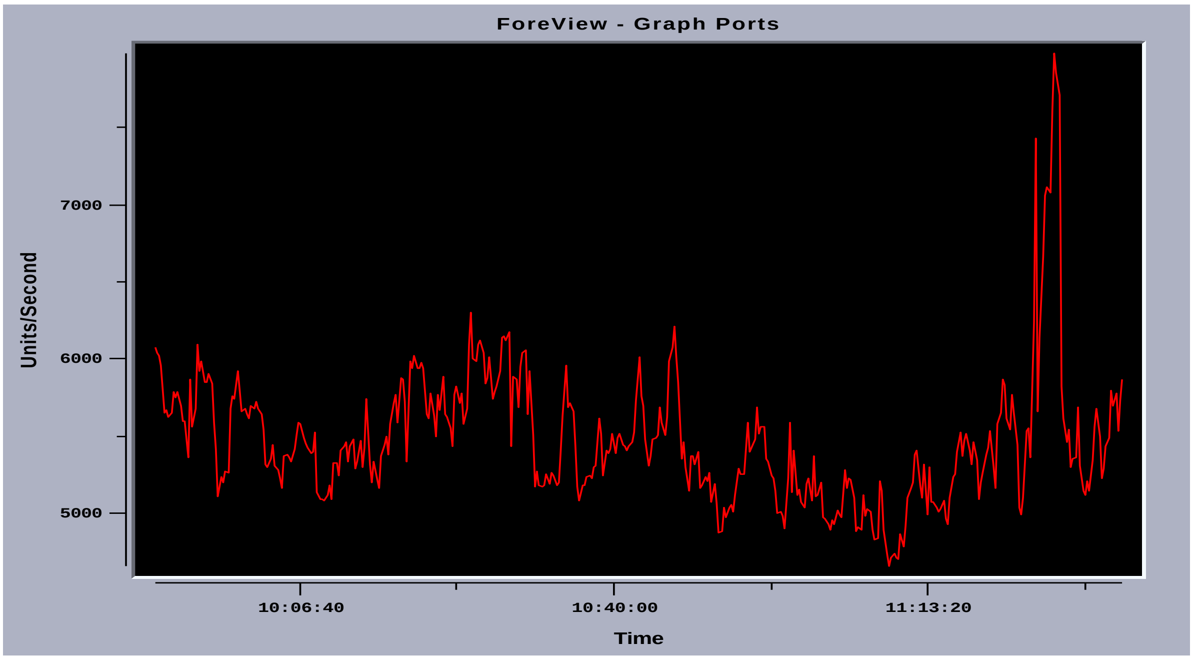

FIGURE 11. Cell Loss for Single Audio and Video Scenario - Bandwidth Variation (5 to 6 Mbit/s)

cell loss statistics were obtained for the port shared memory statistics under ABR-UBR connection, since this is the default policing. Due to practical reasons, the legend in Figure 11 is not included. The cell loss variation is between 5 Mbit/s and 6 Mbit/s. These values are dependent on the input and output characteristics.

Having observed input and output for audio and video streams, we now tabulate the shaping port (b1) statistics for single video and audio scenario. Since a comparison to existing methods is required, initially, video alone and audio alone are shaped and subsequently the problem scenario is applied.

Table 9 shows the port statistics when video alone is shaped. The figures represented are shown in percentages and also the variation within the range, measured by polling every 10 seconds for five minute duration.

TABLE 9. Shaping Port Statistics for Video (in Percentages)

Input Utilization	Output Utilization	Output Queue Loading	Queue Length (in Cells)	Overflows
65 - 75	65 - 75	0 - 8	0 - 450	0

Input utilization is the port's input capacity currently being used, which is maximum 75% of 155 Mbit/s. Output utilization shows the ports output capacity currently being used, which currently is the same as the Input Utilization. The output queue loading shows the port output buffer that is currently full. We observe that the buffer is completely free to a maximum of 8% occupancy. Queue length expressed in cells, is mostly null to a maximum of 450 cells, since the loading on the port is negligible. Overflows expressed in cells, indicate null since the queue is almost empty.

Table 10 Shows the shaping port statistics when audio alone is shaped. While the output queue loading is null, there are fewer cells in the queue compared to video.

TABLE 10. Shaping Port Statistics for Audio (in Percentages)

Input Utilization	Output Utilization	Output Queue Loading	Queue Length (in Cells)	Overflows
0 - 12	80 - 100	0	0 - 9	0

Table 11 shows the problem scenario. Here, the output queue loading is 100 percent all the time. The queue length is significant compared to table 9 and 10, and also maintains a constant number of cells.

TABLE 11. Shaping Port Statistics for One Video and Audio (in Percentages)

Input Utilization	Output Utilization	Output Queue Loading	Queue Length (in Cells)	Overflows
95 - 100	95 - 100	100	9050 - 9154	0

4.3.1 Video and Audio Shaping Bandwidth Combinations

Since the audio and video outputs in Figures 8 and 10 indicate that the total output is almost equal to the average of the two shaping bandwidths, we investigate this trend with a few combinations of shaping bandwidths. The combinations include twice and thrice (and also four times in the case of audio) of base shaping video and audio bandwidth. Table 12 indicates the results of such of investigations. We find that, an increase in audio shaping bandwidth, the output improvement is significant, when compared to large increase in video shaping bandwidth.

TABLE 12. Video and Audio Shaping Bandwidths (in MBit/s)

Video Shaping	Audio Shaping	Input Video	Input Audio	Output Video	Output Audio
13	2	10	1.75	5	1.25
26	2	10	1.75	7	1.5
39	2	10	1.75	6	1.65
52	2	10	1.75	8.5	1.75
13	4	10	1.75	8	1.5
13	6	10	1.75	9.8	1.7
13	8	10	1.75	10	1.75

4.4 Shaping Two Audio Streams

This section investigates the effect of shaping two identical audio streams. Such a scenario is shown in Figure 12.The output values at measurement time are indicated in Table 13.

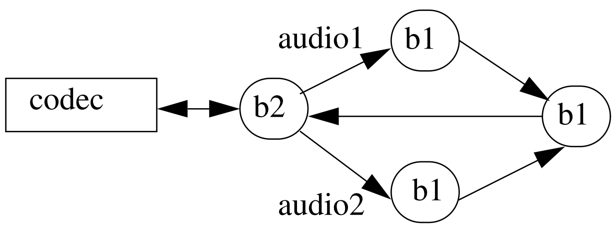

FIGURE 12. Scenario for Traffic Shaping Two Audio Streams

TABLE 13. Input, Shaping, and Output Bandwidth (in Mbit/s)

Input Audio 1	Input Audio 2	Shaping Audio 1	Shaping Audio 2	Output Audio 1	Output Audio 2
1.75	1.75	2	2	0.5	1.5

Although Figure 13 and 14 indicate a near constant audio input of 1.75 Mbit/s (CBR input), their outputs in Figure 15 and 16 indicate 0.5 Mbit/s and 1.5 Mbit/s, respectively. As discussed above, the output bandwidth is the average of shaping bandwidth. It is interesting to note the cyclic behaviour of the two outputs. While one output is at its peak, the other is at its lowest, and vice versa.

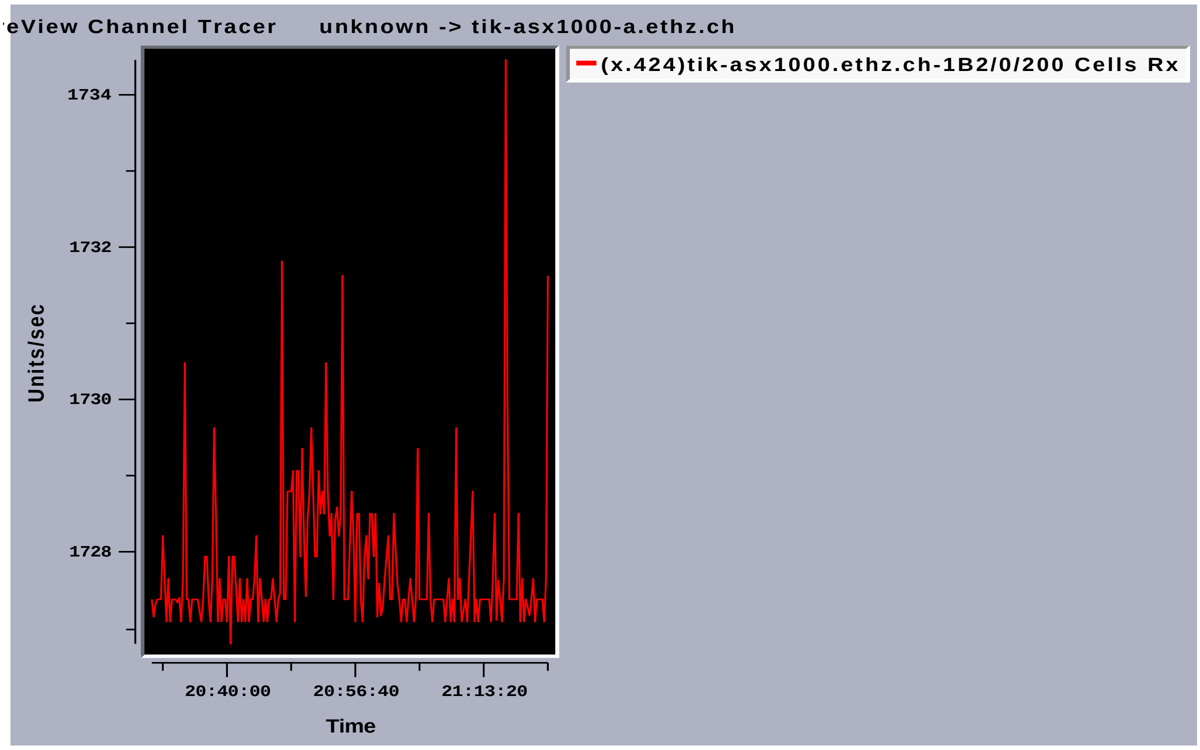

FIGURE 13. Input Audio1 - Bandwidth (1.73 Mbit/s)

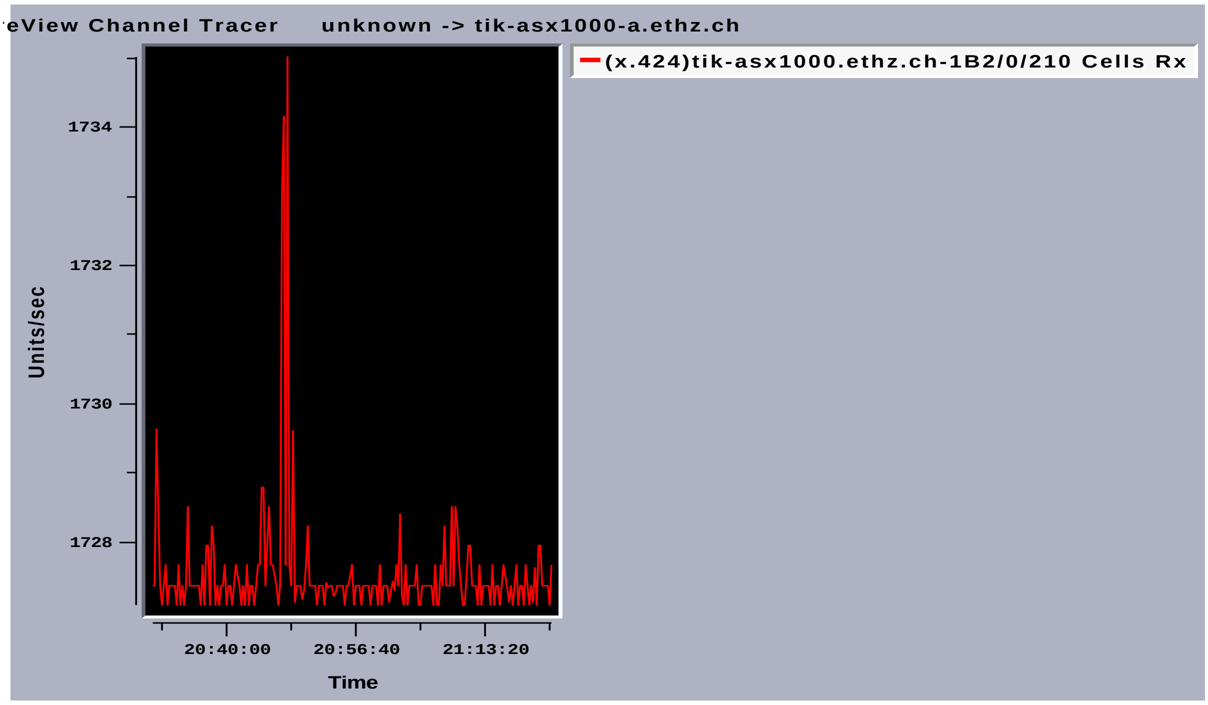

FIGURE 14. Input Audio2 - Bandwidth (1.72 Mbit/s)

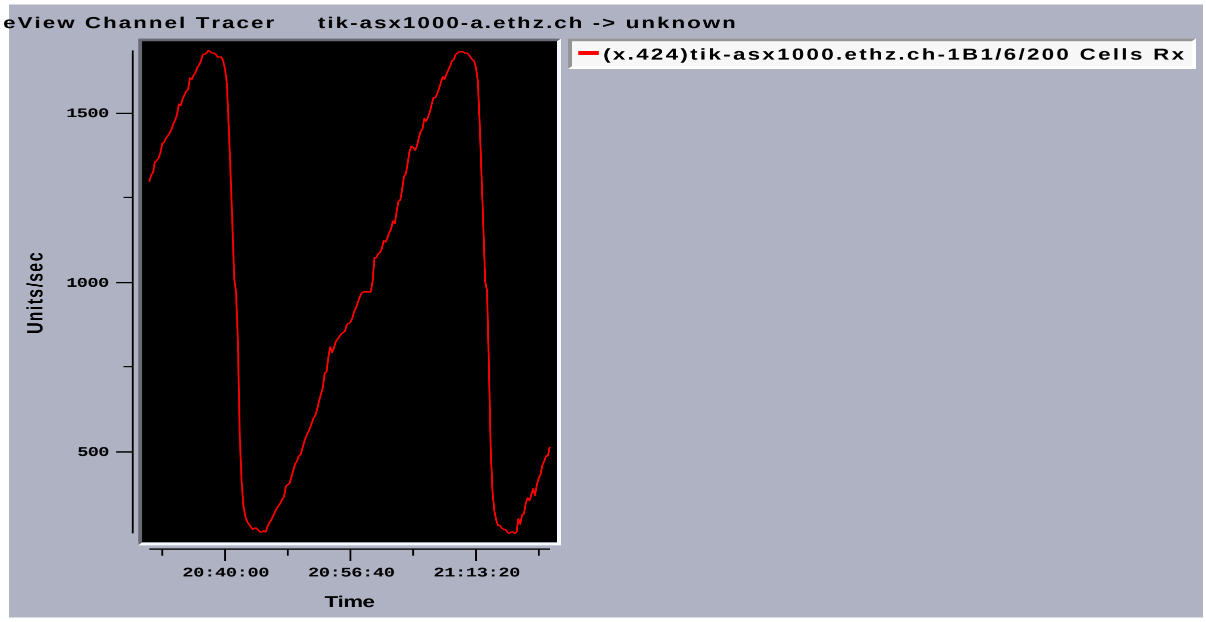

FIGURE 15. Output Audio1 - Bandwidth (1.5 Mbit/s - 0.5Mbit/s)

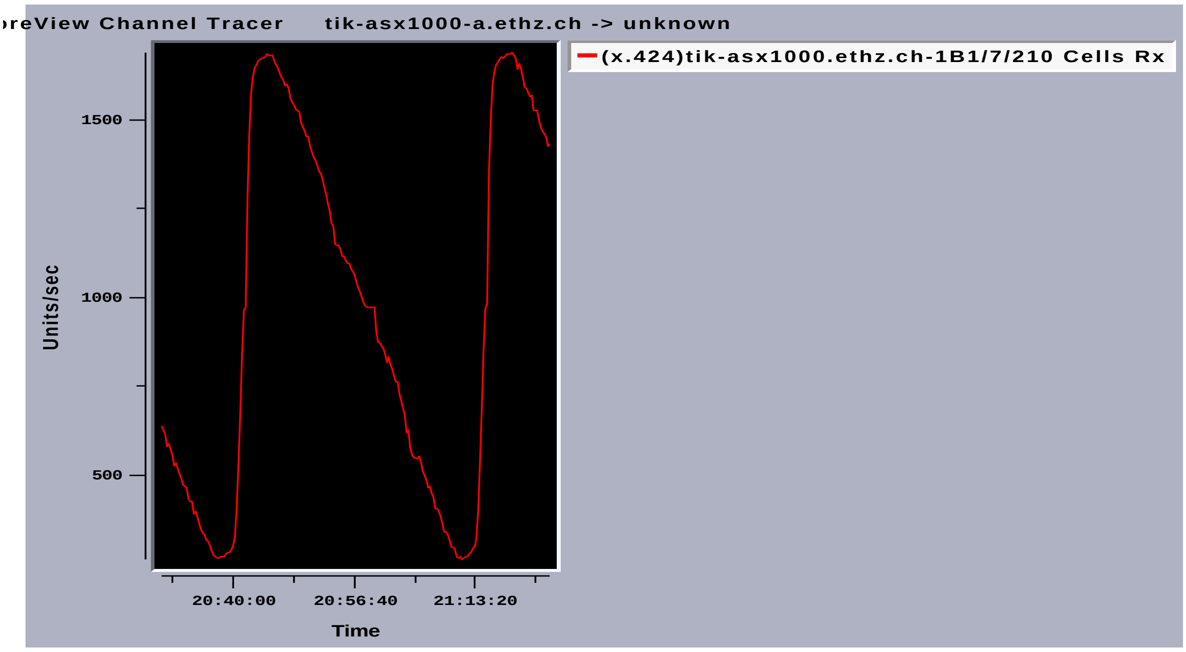

FIGURE 16. Output Audio2 - Bandwidth (0.5 Mbit/s - 1.5 Mbit/s)

Figure 17 is the observed cell loss for the two audio stream scenario. Although the cell loss is near constant, there are peaks at regular intervals. However, such peaks are insignificant since the momentary rise in cell loss is over a small range.

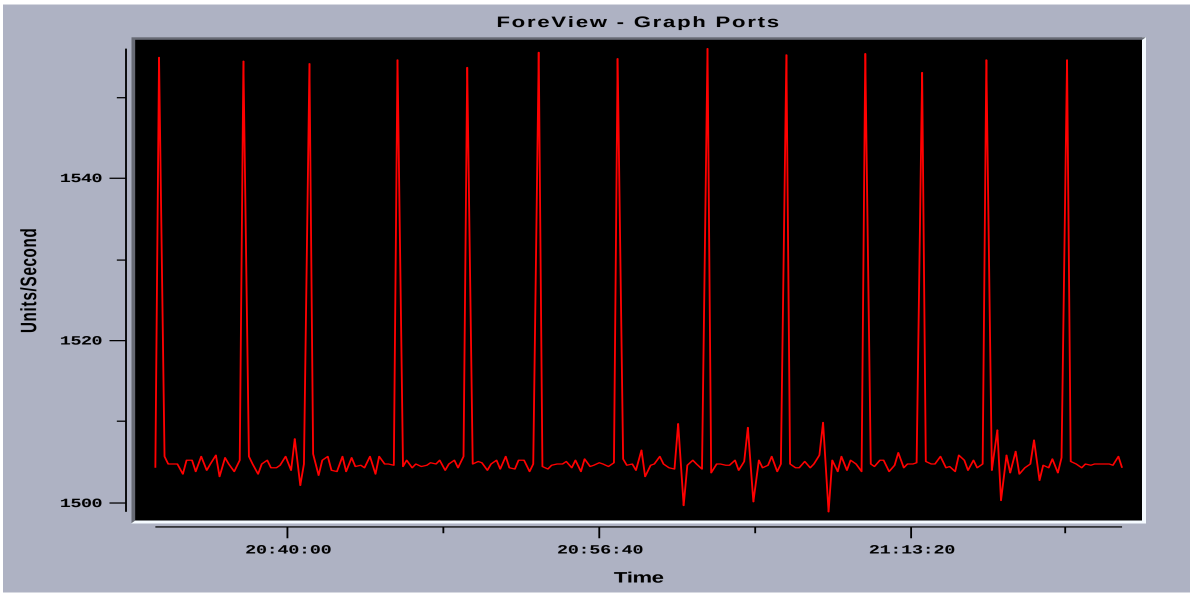

FIGURE 17. Cell Loss for Shaping Two Audio Streams - Bandwidth Variation (1.5 Mbit/s - 1.55 Mbit/s)

Table 14 shows the port statistics for the two audio scenario. While the output queue loading is 100 percent, the cells waiting in the queue are 9100-9200 cells.

TABLE 14. Shaping Port Statistics for Two Audio Streams (in %)

Input Utilization	Output Utilization	Output Queue Loading	Queue Length (in Cells)	Overflows
91 - 100	93 - 100	100	9100 - 9200	0

4.5 Shaping Two Video Streams

This section investigates the effect of shaping of two (VBR) video streams on a single ATM switch port. Such a scenario is shown in Figure 18.

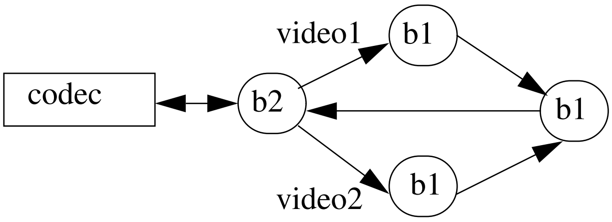

FIGURE 18. Scenario for Traffic Shaping Two Video Streams

The input, shaping and output bandwidths are as indicated in the Table 15.

TABLE 15. Input, Shaping and Output Bandwidths (Mbit/s)

Input Video1	Input Video2	Shaping Video1	Shaping Video2	Output Video1	Output Video2
9	10	13	13	8.8	4.9

The Figures 19 and 20 indicate the graphical input for video 1 and video 2. They indicate the typical characteristics of VBR sources with varying bandwidths due to possible effects of luma and chroma. During measurement, Figure 19 indicates an averaged and higher bandwidth of 9 Mbit/s, and figure 20 indicates 10 Mbit/s.

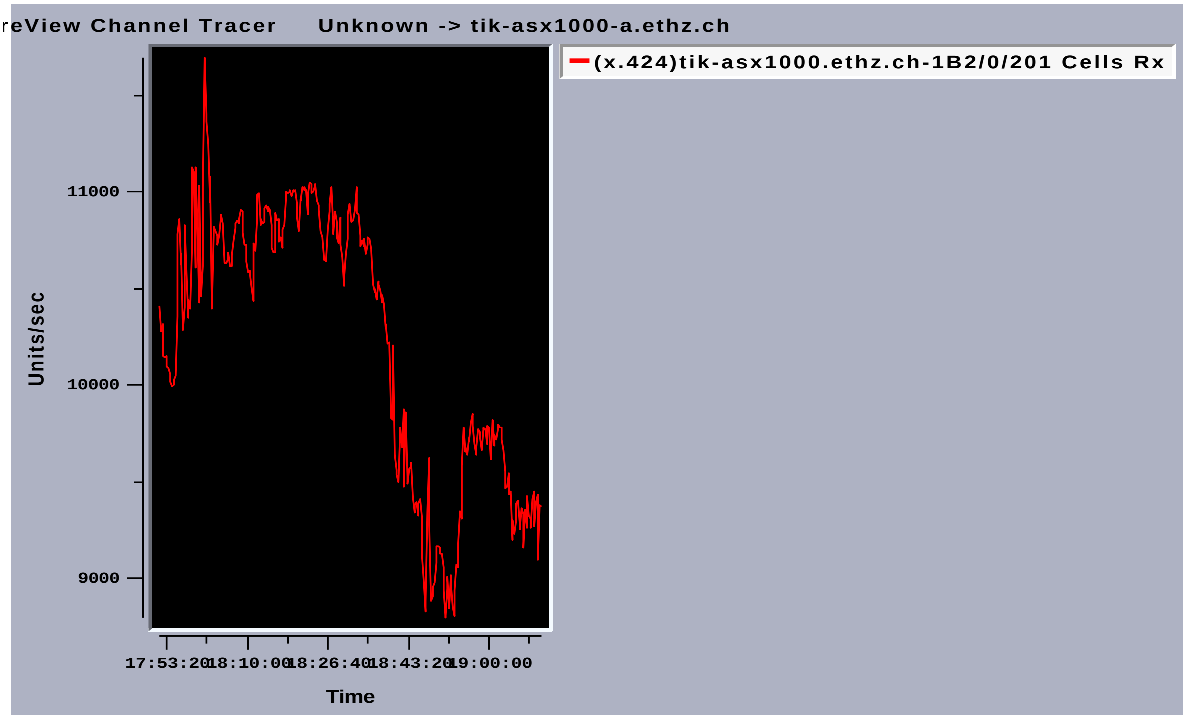

FIGURE 19. Input Video1 - Input Bandwidth (app. 9 Mbit/s)

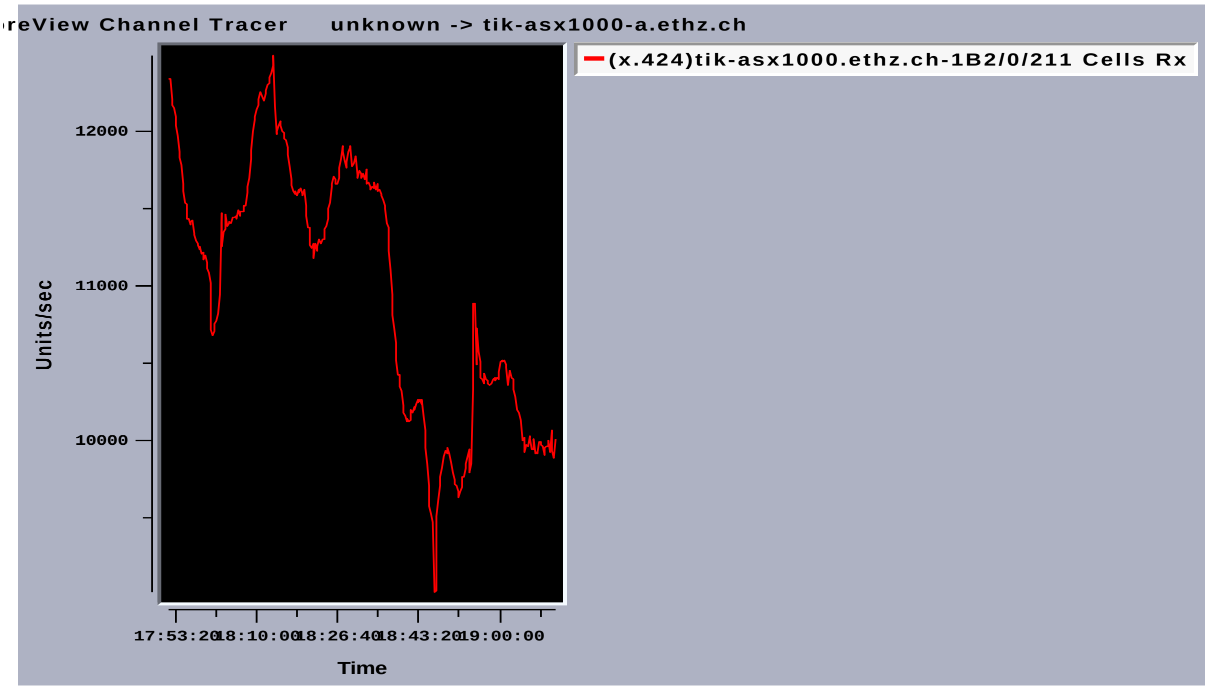

FIGURE 20. Input Video2 - Bandwidth (app. 10 Mbit/s)

Although the variation in the bit rate is reflected at the output, indicated in Figure 21 and Figure 22, the amplitudes are much lower than the inputs, which clearly explains the observed effect of lossy picture. The total output bandwidth obtained is the average of the shaping bandwidths. Table 16 shows the port statistics for the two video stream scenario. While the output queue length is between 8500 to 9500 cells, the overflows continue to be null since the queue can still accommodate cells.

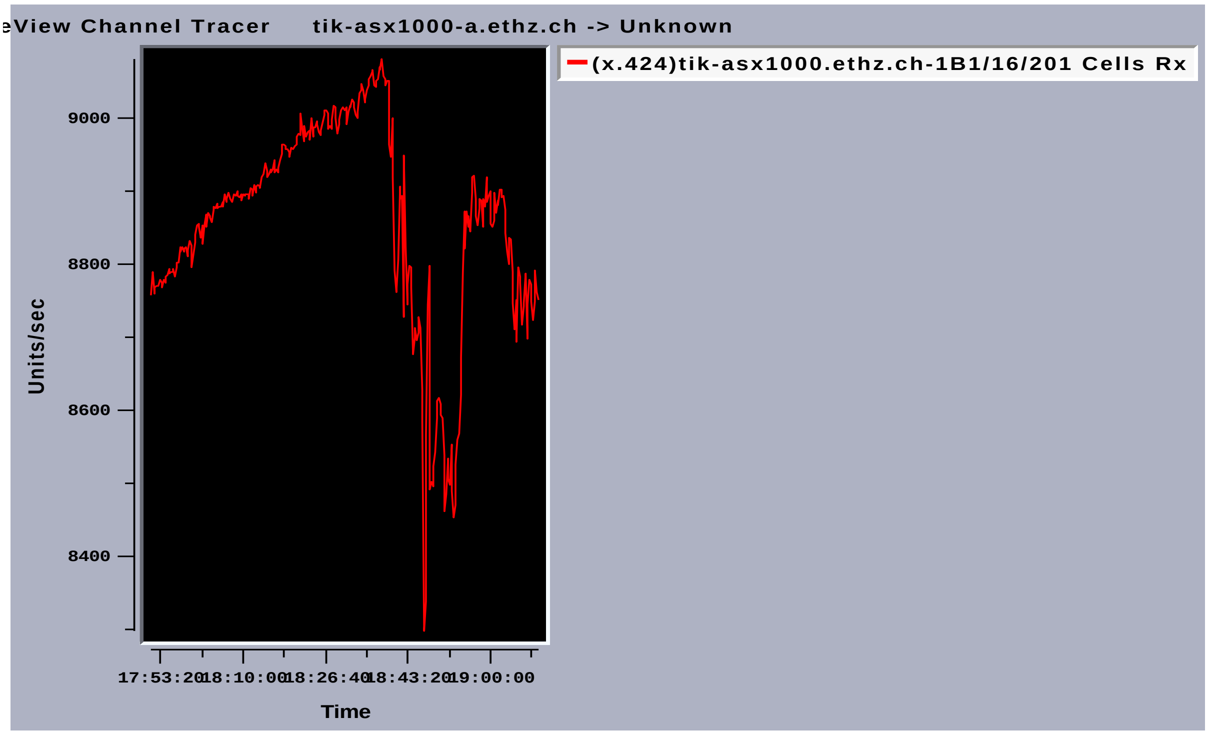

FIGURE 21. Output Video2 - Output Bandwidth (8.8Mbit/s)

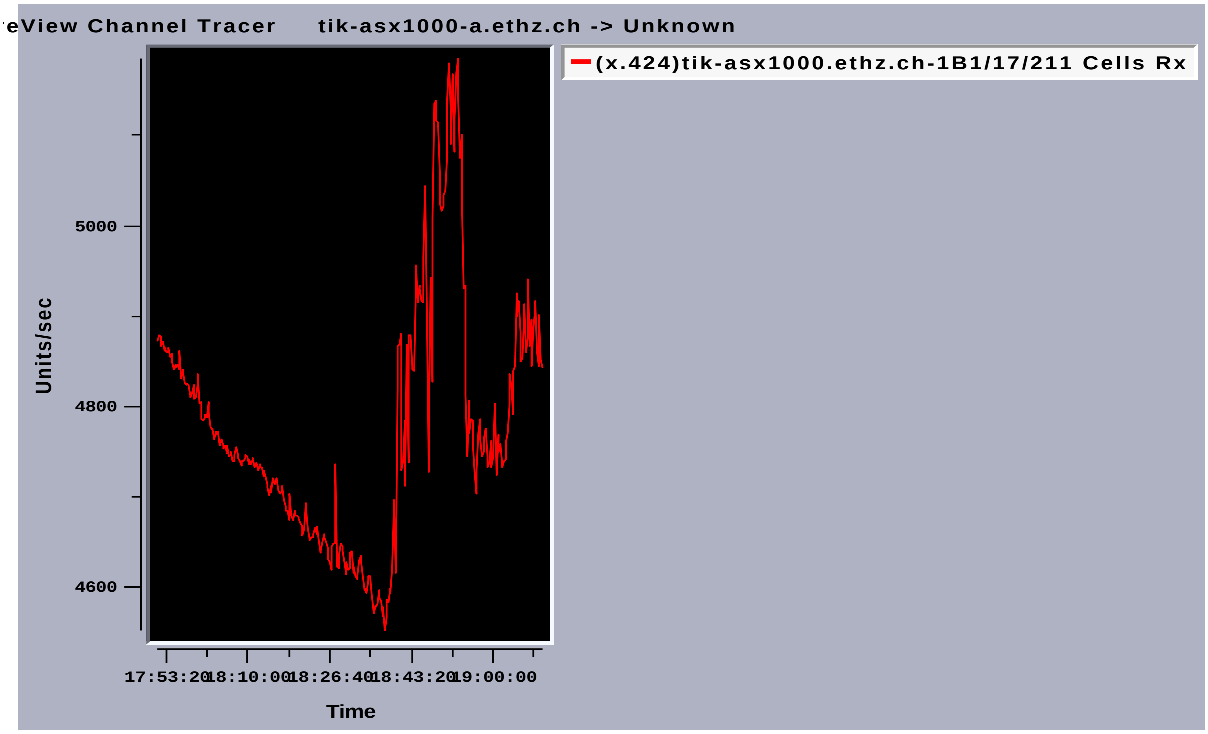

FIGURE 22. Output Video2 - Bandwidth (4.9 Mbit/s)

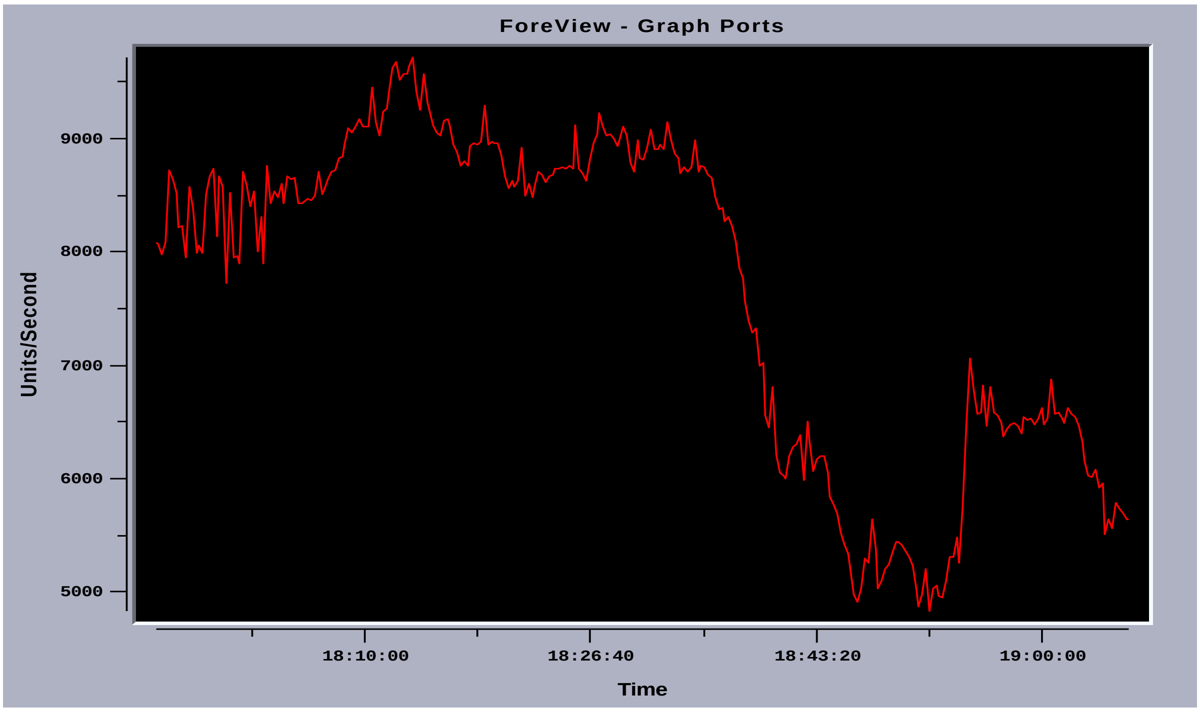

FIGURE 23. Cell loss for Shaping Two Video Streams - Bandwidth Variation (5 MBit/s - 9 MBit/s)

TABLE 16. Shaping Port Statistics for Two Video Streams (in Percentages)

Input Utilization	Output Utilization	Output Queue Loading	Queue Length (in Cells)	Overflows
95 - 100	95 - 100	100	8500 - 9500	0

4.6 Shaping One Audio and One Video (Two Input VPs)

In this scenario we assume that shaping two outgoing VPs on a single physical port may be possible if there are two different input VPs. While all the previous investigated scenarios use VPs indicated in Table 6 of Section 3.5, here we use VPs indicated in Table 11. Such a scenario is as shown in Figure 24. The VP of one of video inputs is changed by routing the signal using another port (b3). Subsequent to this, the two signals are shaped at port b1.

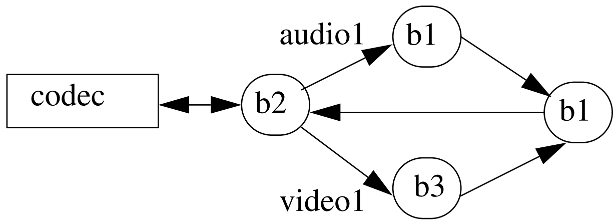

FIGURE 24. Scenario for Traffic Shaping a Single Audio and Video, Using Different Input VPs for Audio and Video

Table 17 shows the applied VPI and VCI for scenario 4.6. Observe that the video VP is now 5 instead of the regular 0. Such a situation is in accordance to our assumption. The results of these experiments to be the same as the results of section 4.3.

TABLE 17. Input and Output VPI and VCI for Scenario 4.6

Stream	Input VPI	Input VCI	Output VPI	Output VCI
Audio1	0	200	6	200
Video1	5	201	16	201

4.7 Other Single Video and Audio Investigations

Other single video and single audio investigations include combinations of video and audio with and without shaping. The general setup is shown in figure 25. Table 18 indicates the results of the investigations along with the combinations. Although other combinations such as two video inputs and one audio input, two audio inputs and one video input are possible, we only choose the ones indicated in table 12, since we wish to be in line with our previous measurements comprising of only two input signals. Here the term ‘no shaping’ means we don’t apply any shaping bandwidth limit on the outgoing streams. All the combinations indicate a good audio and video reception.

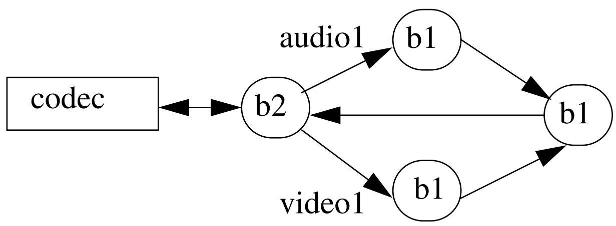

FIGURE 25. Other Scenarios for One Video and One Audio

TABLE 18. Scenarios for Single Video and Audio- Input, Shaping, and Output Bandwidths (Mbit/s)

Input Audio	Input Video	Shaping Audio	Shaping Video	Output Audio	Output Video
1.75	10	no shaping	no shaping	1.75	10

TABLE 18. Scenarios for Single Video and Audio- Input, Shaping, and Output Bandwidths (Mbit/s)

Input Audio	Input Video	Shaping Audio	Shaping Video	Output Audio	Output Video
1.75	10	2	no shaping	1.75	10
1.75	10	13	13	1.75	10

5. Summary

In Section 4.3, referring to Table 9 on port statistics for video, we observe that cells waiting in the queue is slightly higher compared to audio, possibly due to reason that video bandwidth is higher than audio.

Referring to Table 11, the problem scenario, we observe queue length varies over a small range. In otherwords, the number of cells waiting to be serviced is fixed. Cell loss in Figure 11, clearly explains the reason for lossy picture and jarred audio and therefore accounts for the loss of output bandwidth.

In Section 4.3.1, referring to Table 12, we observe that a small increase in audio shaping bandwidth achieves a better output compared to significant increases in video shaping bandwidth. The shaper does better if the shaping bandwidths of the two streams are converging to higher of two values (cf. Table 18). For example, by setting the audio and video shaping bandwidths equal to 13 Mbit/s, one can successfully shape two outgoing VPs.

In section 4.4, we observe the behaviour of the traffic shaper when the two inputs are near CBR audio sources. Table 13 once again shows that output bandwidth to be average of the two shaping bandwidths.

Referring to Figures 15 and 16, initially, audio 1 (VCI 200) does better for a while and later audio 2 (VCI 210) and vice versa. There is a sudden rise and fall in bandwidth, alternating with gradual rise and fall. This cyclic behaviour is observed over a long time interval.

In Section 4.5, we observe the behaviour of the traffic shaper when the two inputs are VBR type. Here again, the output bandwidth is the average of the shaping bandwidths. Also, video1 with VCI 201 does better than video2, which has a VCI of 211.

Table 16 shows the port statistics for the two video streams scenario. It is interesting to note that irrespective of the input streams combinations, the queue length is almost identical in all the investigated scenarios. One may confirm this by comparing Table 11, 14, and 16.

In Section 4.6, while most of the discussed scenarios use a single VP (VP=0), we try to observe the behaviour of the traffic shaper when the inputs are from two different VPs. However, the output is quite similar to Section 4.3. This proves that there exists only one shaper per port.

In Section 4.7, while all the previous scenarios considered shaping for both input streams, here we study the behaviour of traffic shaper when one of the inputs is shaped.

Referring to Table 18, the second and third rows clearly show that we can only shape one outgoing VP. All VCs running inside this VP are shaped. However, row one shows that we can still shape two outgoing VPs provided the shaping bandwidth of the outgoing VPs are made equal to the higher of the two values.

6. Discussion

By default, the ASX-1000 ATM switch polices all input traffic at the UNI. However, if no policing is set up, all traffic is regarded as ABR-UBR traffic. Since we do not apply any UPC on the VCs, all cell losses recorded in the investigated scenarios fall under default ABR-UBR connections. Further, in our assumptions it was stated that for all ABR-UBR traffic, there is a dynamic allocation of buffer space. However, since cell loss occours in our investigations, it is clear that such a dynamic allocation of buffer space and per VC queuing does not apply to shapable traffic streams.

This is because all shapable traffic streams, most likely enter another buffer space, the shaping buffer. The proposed shaping buffer's memory space is allocated and fixed depending on the outgoing shaping bandwidth set manually. In other words, because of shaping, there is a static allocation of buffers in order not to loose cells during cell spacing. Therefore, the concept of distributed and shared memory, discussed in white papers is not applicable. This also explains the reason for buffer overflow condition.

The reason for queuing of cells is because there is a contention between cells of the two VCs for entering the shaping buffer. Therefore, in order to resolve this contention, the cells are placed in a queue. Also, since the queue service time is fixed, the number of cells waiting in the queue is also fixed and therefore independent of the input stream bandwidth. This explains the reason for near constant queue length in all investigated scenarios.

Since the output bandwidth is the average of the input shaping bandwidth, the ATM switch randomly drops 50% of the information cells from either VCs. The remaining 50% surviving cells are spaced out.

The reason for cyclic behaviour in Section 4.4 is probably because for 50 percent of the time, the cells from one VCI 200 are dropped and the remaining 50 percent of the time the cells from VCI 210 are dropped. Thus still maintaining the observed average output.

Finally, the shaping buffer suffers from a drawback which is related to its size. While one could say that the shaping buffer has sufficient capacity to offer an average output of the input shaping bandwidths, such a scheme does not hold for the scenario discussed in Section 4.3.1, which is shown in Table 12.

7. Conclusions

From the ATM experimental setup and discussed scenarios, the following is concluded:

There exists only one traffic shaper per port. Such a traffic shaper can shape only one outgoing VP. Also, there exists an additional shaping buffer, where all shapable traffic flows into this buffer. Also, the method adapted in handling shaping buffers seems to be quite different from the output buffer handling.

Considering the high cost of ATM hardware and given the scenario, where a service provider expect shaped traffic only, if multiple VPs cannot be shaped on a single port, it is quite a serious shortcoming. However, if manufacturers of ATM end devices recognize the importance of traffic shaping by introducing shapers within their equipment, a considerable burden on ATM switches can be removed. Also, congestion due to short spikes may be handled effectively

8. References

/StNa95/ R. Steinmetz, K. Nahrstedt: *Computing, Communications and Applications;* Innovative Technology Series, Prentice Hall, Englewood Cliffs, New Jersey, U.S.A., 1*995.*

/HHSc94/ R. Händel, M. N. Huber, S. Schröder: *ATM Networks Concepts, Protocols, Applications;* Addison Wesley, Reading, Massachusetts, U.S.A., 1994.

/CuSa93/ L. G. Cuthbert and J.-C. Sapanel: *ATM - The Broadband Telecommunications Solution;* IEE Telecommunication Series 29, 1993.

/KMHY89/ F. Kishino, K. Manabe, Y. Hayashi, H. Yasuda: *Variable Bit Rate Coding of Video Signals for ATM Networks;* IEEE Journal on Selected Areas in Communications, Vol. 7, No. 5, June 1989.

/BGSC92/ P. Boyer, F. M. Guillemin, M. J. Servel, J.-P. Coudreuse: *Spacing Cells Protects and enhances Utilization of ATM Network Links;* IEEE Network, September 1992.

/Pryk95/ M. De Pryker: *Asynchronous Transfer Mode;* Prentice Hall, Englewood Cliffs, New Jersey, U.S.A., Third Edition, 1995.

/YSSa92/ N. Yamanaka, Y. Sato, K. Sato: *Traffic Shaping for VBR Traffic in ATM Networks;* IEICE Transactions on Communications, Vol. E75-B, No. 10, October 1992.

/Poyn96/ C. A. Poynton: *A Technical Introduction to Digital Video;* John Wiley and Sons, 1996.

/Knet/ K-Net Ltd: *CellStack User Guide;* K-Net Ltd, Hatchwood Place Farnham Road, Odiham, Hampshire RG29 1AB, United Kingdom.

/FView96/ Fore: *Foreview 4.0, Network Management User's Manual;* FORE Systems, Software Version 4.0, March 1996.

/FRunner96/ Fore: *ForeRunner ATM Switch Configuration Manual;* FORE Systems, Software Version 4.0, March 1996.

/FArch/ Fore: *ForeRunner ATM Switch Architecture;* FORE Systems, Version 1.0, , http://www.fore.com/atm-edu/pdf_files/frarch.pdf.

/FBand/ Fore: *ForeThought Bandwidth Management;* FORE Systems, Version 1.0, http://www.fore.com/atm-edu/whitep/ftbmwp.html.

/WSHP97/ T. Walter, B. Stiller, H. Hänni, B. Plattner: *Distributed Learning Scenarios supported by ATM - Experiences in Telepoly;* Fachtagung Mediengestützte wissenschaftliche Weiterbildung, Braunschweig, March 20-21, 1997, pp 69-83.