

Effects of Voice Compression on the Operation of a N-ISDN/B-ISDN IWF

Nasser-Eddine Rikli
*Faculty of Engineering,
Universiti Putra Malaysia,
43400 UPM Serdang, Selangor*

ABSTRACT

The significant widespread of N-ISDN, which has recently gained momentum, will make it hard for the emerging B-ISDN whether to ignore its presence or to phase it out and replace it in the near foreseen future. Consequently, the ATM of the interoperability specification CES-IS V2.0 (af-vtoa-0078.000), which defines emulation standards for circuit characteristics of constant bit-rate (CBR) traffic within ATM. A critical attribute of a circuit emulation service (CES) is to achieve a performance comparable to that are still pending will be addressed. Special interest will be given to devising methods that will enable voice, which is a sizeable component of the current N-ISDN traffic, to be carried efficiently over ATM network. First, a multiplexing technique for voice sources will be presented. Then, assuming that speech silence detection is being used, a technique for dealing with the partially filled cells will be suggested, analyzed, and then simulated. The results will be then presented and analyzed, followed by conclusions and suggestions.

Keywords: N-ISDN, B-ISDN, ATM, voice, VOATM, speech silence detection, multiplexing, internetworking function, VBR

INTRODUCTION

It is becoming almost uncontested that ATM will be the premium choice for the efficient transport of multimedia traffic, such as voice, data, images, and video. Thus, we find an extensive coverage and study of ATM support for these new types of traffic. However, when it comes to the research and implementations being carried out for ATM using voice traffic, only a few will be found. This is due to the three main obstacles that may be summarized in:

- the high cost of early ATM interfaces,
- the diversity of features available on voice networks and not on the ATM's,
- the unfinished standardization process for voice over ATM.

Fortunately, it can be argued rightfully (Wright 1996) that the two first hurdles have been entirely overcome. Our goal here is to try to alleviate the third and hopefully last hurdle.

The need to consider the transport of traditional voice traffic (not as part of multimedia) over ATM network is steamed mainly by the objective of ATM to become a global technology for all networking technologies. Thus, as the ATM-based B-ISDN network is expected to integrate all types of communication service and applications, it is imperative to include support for the existing interfaces and service in general. However, the N-ISDN should be given high priority as its deployment started recently to gain momentum and significant worldwide spread (Petri and Schwetje 1996).

As voice traffic is a real-time but low bit-rate applications, a new challenge is facing ATM researchers, due to the fact that ATM was inherently designed for high data rate applications. Thus, an important function that has to be included in the corresponding ATM Adaptation Layer (AAL) is the efficient conversion of voice-band signal (as defined

in ITU-T Recommendation G.711) into ATM cells, and vice versa. It is worthwhile mentioning that the ITU has already specified that the ATM network has to transport voice plus its corresponding signaling information.

This being said, it is believed that any proposed ATM adaptation layer for voice should be evaluated taking into account three criteria: the structuring of data, the source clock recovery, and the detection of cell loss or misinsertion. The first criterion will give an indication on the network efficiency and the transmission overhead. The second criterion is necessary, as the considered application is real-time. Another issue that may be related to this criterion is the delay requirement for voice. This delay will include four components: the assembly of speech traffic into ATM cells, the switching delay, buffering delay, and transport delay. The third and last criterion becomes important in the case of voice, only if the cell loss surpasses a certain predetermined threshold.

Carrying voice signals on other network than the Public Switched Telephone Network (PSTN) is not a new idea, it has been considered for many years. Many standards have been added to the TCP/IP protocol suite so as to allow phone conversation, especially the long distance ones, to be carried cheaply on the Internet. The Real Time Streaming Protocol, or RTSP, (RFC 2326; White 1997; Brazilai et al. 1998) which is an application-level protocol for control over the delivery of data with real-time properties. This protocol is intended to control multiple data delivery sessions, provide a means for choosing delivery channels such as UDP, multicast UDP and TCP, and provide a means for choosing delivery mechanisms based upon RTP (RFC 1889 and RFC 1890). Also, voice traffic has been considered for transport over local area network (Dempsey et al.), such as FDDI (Mei and Everitt 1996). Ethernet (Qigang et al. 1996), and more recently wireless (Visser and Jarki 1995) and ATM (Okubo et al. 1997).

In this paper, we will consider voice sources generating variable bit-rate signals, thus of VBR type. Furthermore, we will assume that the Internetworking Function (IWF), or gateway, between the N-ISDN and B-ISDN is equipped with Silence Detection equipment that will allow the ATM network to discard silence periods. We will address the issue of unfilled cells and front-end clipping resulting from these proposed techniques. The performance measures that will be utilized are the channel utilization, the percent stuffed cells, and the percent lost cells.

The rest of the paper is organized as follows. In section two, we will present the definition, operation, and service of the IWF as devised by the ATM Forum. In section three, we will describe the voice source statistical model and the components of the system under consideration, and then present the technique proposed in dealing with the unfilled cells. In section four, a simplified mathematical analysis will be presented. In section five, we will present some simulation results and compare them to the analytical ones. Then, in the last section, we included our conclusions.

ATM INTERNETWORKING FUNCTION (IWF)

Proposed Reference Models

The ATM forum has currently addressed internetworking scenarios of users on a B-ISDN accessing N-ISDN service through internetworking functions (IWF). Three tyoes of attachments are described in (Voice and Telephony 1997), namely:

- Broadband terminal (B-TE) to a B-ISDN (private or public), despiceted in Figure 1;
- Private B-ISDN to Public B-ISDN, depicted in Figure 2;
- Private B-ISDN to Private N-ISDN, depicted in Figure 2.

Voice Compression on the Operation of a N-ISDN/B-ISDN IWF

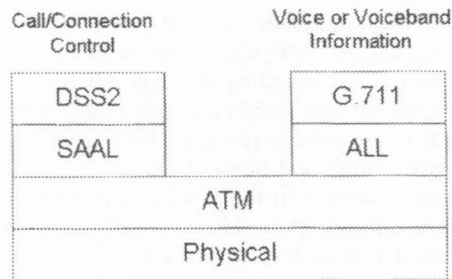


Fig. 1. Protocol reference model for B-Terminal to B-ISDN attachment

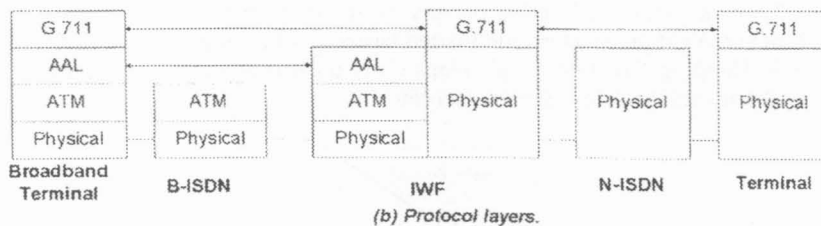


Fig. 2.

The role of the IWF, in Figures 2-(a) and 2-(b), is to convert between TDM encoded voice-band information and ATM cell-based voice-band information. It may be considered as a logical function that communicates with a peer IWF on the other side of the ATM network. It may be either a stand alone device with physical interfaces to the ATM and N-ISDN network, or it may be integrated into either the ATM and network equipment or the Narrowband network equipment (Voice and Telephony 1997).

The terminals considered so far were represented by a single 64 kbps A-law or μ -law PCM-encoded voice-band signal. Each ATM Virtual Channel Connection (VCC) is mapped to one N-ISDN channel dynamically on a per call basis. The ATM Forum did not include yet the techniques that use alternative coding for voice transport and silence removal. In this paper, we will try to address the implications of introducing some of these techniques on the system performance.

Basic Operation

The IWFs are connected on one the ATM network via physical interfaces defined in the ATM Forum UNI Specification, and on the other to standard constant bit-rate (CBR) circuits (e.g. DS1/DS3, J2 or E1/E3). The role of the two IWFs is to extend the CBR circuit to which they are connected across the ATM network. They are to do this in a manner that is transparent to the terminating equipment of the CBR circuit. This means

that the ATM portion of the connection should retain its bit integrity. The IWF is responsible of multiplexing and demultiplexing several VCCs, performing segmentation and reassembly on one VCC, and assuming the timeslot mapping function. It should maintain bit count integrity, i.e. the number of bits entering through the segmenting unit should be equal to the one leaving through the reassembling unit. The cells that are lost or misinserted within the ATM network should be detected by the reassembly function using the sequence numbers in the AAL1 headers. In the case of not too many cells lost, dummy cells consisting of 46 or 47 octets will be inserted to maintain the bit count integrity. As misinserted cells are expected to be rare, all misinserted cells are dropped.

The reassembly function will also require a buffer in which the reassembled cell stream is first stored before being transmitted on the output Service Interface. The buffer size should be large enough to accommodate expected Cell Delay Variation (CDV), while small enough to limit the delay in the emulated circuit. The desirable design criteria would be to minimize the buffer size while maintaining low probabilities of both overflow and underflow conditions. In case of a buffer overflow, a number of bits that is implementation-dependent should be dropped.

Example of a Circuit Emulation Service (CES)

In Figure 3, we have shown a layering perspective if a DS1 (or E1, J2) Unstructured CBR service. From an ATM perspective, the shaded layers may be considered as an "ALL User Entity". The Mapping Function simply maps every bit between the AAL1 layer and the 1,544 (or 2,048, 6,312) Mbps Service Interface.

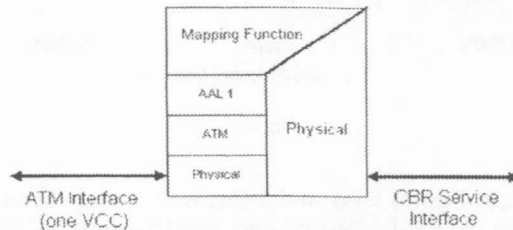


Fig. 3. Layering Perspective of a CES Interworking Function

The number N of timeslots available on the N-ISDN side are carried across the ATM network and reproduced at the output edge. This timeslots assigned to a VCC are not required to be contiguous at the input. However, at the output the IWF must delivery octets in the same order they were received at the input from the CBR source. Also, the $N \times 64$ service must maintain 125- μ sec frame integrity across a virtual channel.

In this case where Unstructured Data Transfer is used, the bits received from the service interface are packed into cell without regard to framing (CESIS 1997; PISN 1998). In other words, no particular alignment between octets from the frames on the Narrowband side (DS1, E1, or J2) and octets in an ATM cell on the Broadband side. However, correct bit ordering must be maintained. For example, the 376 contiguous bits (8×47 bytes/ cell) that will be packed into one SDU will constitute of:

- the MSB of the first octet will come from the first bit received on the Narrowband side;
- the bit placement proceeds in order (i.e. first bytes filled first, and MSB bits filled first and LSB last);
- the last bit received will be placed in the LSB of the 47th octet of the SDU.

In the case of $N \times 64$ service carrying N DSO timeslots organized into blocks, with each block being an integral number of octets in size, then a pointer mechanism is used to delineate these repetitive blocks. If the block size is one octet, then this is simply achieved by aligning each AAL user octet with an ATM cell payload octet. However, if the block size is greater than one octet, then a pointer indicating the start of the sequence is inserted in a single cell of a set of eight cell payloads with AAL sequence count cycle (i.e., 0,1,2,...,8). The selected cell should have an even sequence count value (i.e., 0,2,4, or 6). A similar multiframe structure, with extra signaling information carried, applies to the case where we have DS1/E1 circuits instead of DSO (CESIS 1997; PISN 1998).

Channel Utilization

The delay incurred at the IWF on the sender side is mainly due to a fixed delay and a variable delay. The fixed delay results from the cell payload assembly (the time it takes to collect enough data to fill a cell), and the variable delay results only if unfilled cells are allowed to be carried on the ATM network portion (CESIS 1997; PISN 1998). However, if cell padding is used through dummy octets, this delay portion will be reduced at an expense of a waste in bandwidth. It is worthwhile mentioning that in accordance with ANSI T1.630, the IWF shall fill the entire 47-octet cell payload with DS1/E1/J2 data.

Service Requirements

The service requirements specified by the ATM Forum (Voice and Telephony 1997), may be summarized as follows:

1. Call-by-Call routing: this allows more effective use of the transmission and capabilities of the ATM network.
2. Bandwidth on demand: this allocation is based either on predetermined traffic patterns or on the requirements of a new call.
3. Bandwidth sharing: the freed bandwidth may be assigned to other services.
4. Support of DS1/E1 with signaling: this is a required service.

Within our framework, we will consider only the second and third services. Routing and signaling are beyond our scope.

SYSTEM DESCRIPTION

3.1 Source Model

The voice sources to be considered here are allowed to be of the VBR type. They will be using VOCoders with silence detection, which is also known as voice activity detection (VAD). The ATM Forum has not yet addressed the transport of compressed voice with the option of VAD (Voice and Telephony 1997). In this paper, we will try to tackle some aspects of these issues. A source is allowed to alternate between talkspurt and silent periods according to exponential durations with statistical averages of 352 msec and 650 msec respectively (see Figure 4).

Packetization

A voice burst, arriving at the IWF for packetization into fixed size cell, may be categorized, based on its size into two types; either the voice burst is less than a cell

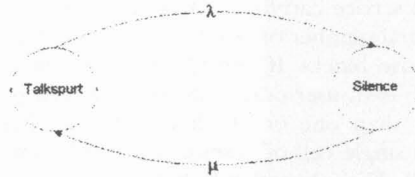


Fig. 4. Voice source model

payload, as know in Figure 4(a), or it is larger than a cell payload as shown in Figure 4(b). In both cases, we can notice from the diagram the relative portion of the unfilled cell. However, as can noticed from the diagrams, the relative portion of the unfilled cell to the total burst will be higher in (a) than in (b). Also, as the burst lenght become longer, this ratio become smaller. As we will see in our results, this is very closely related to the source-encoding rate (compression factor), in addition to the statistical variability in the signal duration.

Other effects may also result from allowing the source to be of VBR type. Indeed, the more compression is applied to speech, the longer the effect of cell loss. Furthermore, if encryption is used, recovering from cell loss or misinsertion can result in additional errors. The listener may therefore notice the effect of cell loss when a single highly compressed and/or encrypted voice call is carried over an ATM connection. However, only the first effect (loss) is to be considered in here.

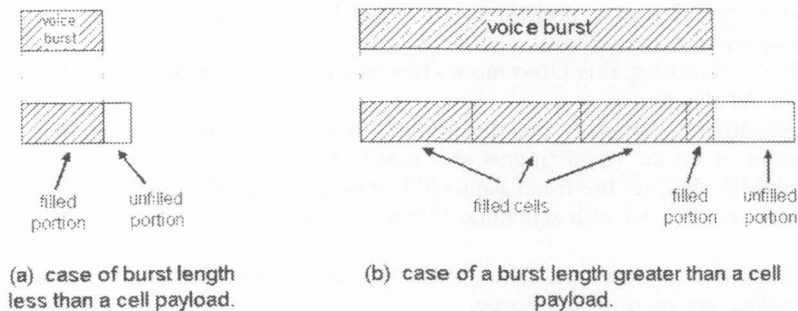


Fig. 5. Packetization of a voice burst into ATM cells

Switching

The system to be considered consists of a number of voice sources multiplexed through an IWF on to a common fixed bandwidth ATM channel. Using Speech Silence Detection systems, a source accepted on the network will transmit only while in talkspurt mode. During its silence period, the unused bandwidth is used by the other sources already in talkspurt mode.

Of course, there will be times where a source returns from a silence period to a talkspurt period and does not find any available bandwidth. Such situations will induce what is known as front-end clipping (see Figure 6). Limiting the number of source accepted into the network will control the loss due to this phenomenon. Thus, depending on the voice quality required, a limit on the voice sources accepted into the network ought to be fixed. Of course, since we are using statistical multiplexing (an inherent feature in ATM), the aggregate source rate should be larger than the total bandwidth allocated for voice traffic.

3.4 Unfilled Cells

The first we want to deal with is the unfilled portion of a cell. The importance of this issue depends on the conversion rate (compression ratio) of the voice encoder. For instance, if the generation rate is 64 kbps, then a voice burst of 500 msec will generate 32,000 bits, which will be converted into 83 cells plus once cell with 1/3 filled and 2/3 unfilled. The same voice burst with an 8 kbps encoder will produce 10 cells and 3/7 of a cell. Table 1 illustrates the relative wasted bandwidth.

TABLE 1
Comparison of wasted bandwidth as function of encoder rate

Encoder rate (kbps)	Voice burst (msecs)	Number of cells	Unfilled fraction	Relative waste
64	500	84	2/3	0.8
8	500	11	3/7	3.8

It is very clear that there is a higher effect of the unfilled cell with lower encoder rate than with higher rates. As the current trend is in lowering the encoder rate, the effect of partially filled cells seems to be the limiting factor that has to be seriously addressed.

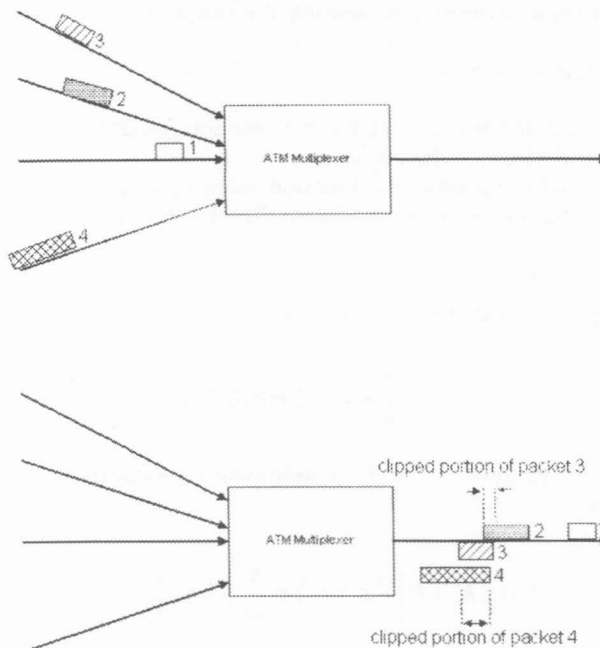


Fig. 6. Proposed switching technique

In what follows, we will propose the following technique in addressing this difficulty: send any unfilled cell as a whole cell by using cell padding (i.e. bit stuffing). The proposed technique is based on the assumption that speech silence detection is used. That is the bandwidth assigned to a terminal TE is used only while in talkspurt mode. As soon as it become silent, it releases the bandwidth, which is then assigned to another

source (if there is any). When the source become active again, it requests the required bandwidth, and gets it again if available, otherwise the front-end talkspurt gets clipped. Incidentally, a major criterion for accepting a new source call is the average percent clipping incurred by all sources.

Our proposed technique is, in some aspects, similar to the one already adopted by the ATM Forum. We assume that when a voice burst (talkspurt) is generated, it is packetized into fixed length cells, and the last partially filled cell is filled with dummy octets. It is expected that as this technique will improve the delay performance of the system, it will also deteriorate its efficiency in utilizing the available bandwidth.

ANALYSIS

Problems Formulation

Denote the voice talkspurt duration by random variable X which has an exponential distribution $f_X(x)$ with parameter λ . The filled portion of the unfilled cell will be denoted by the random variable Y , defined by:

$$Y = X \text{ mod } D \tag{1}$$

Where D is fixed value number representing the duration of cell (for given encoder rate).

The goals of this analysis are:

- (i) Find the distribution function of the new random variable Y .
- (ii) Derive an expression for the average
- (iii) Compare the resulting plot as a function of the product d with the simulation results (this part will be presented in the Results section).

Distribution Function $f_Y(y)$

Equation (1) may be rewritten equivalently as:

$$Y = \frac{X}{D} - k, \quad k = 0, 1, 2, \dots \tag{2}$$

Let us first find an expression for the conditional cumulative distribution function $G_Y(y, k)$ defined by:

$$G_Y(y, k) = P \left\{ Y \leq y / k \leq \frac{X}{D} \leq k + 1 \right\} \tag{3}$$

Knowing that $f_X(x)$ is given by:

$$f_X(x) = \lambda e^{-\lambda x}, \quad \text{for } x \geq 0 \tag{4}$$

Using Eqs. (3) and (4) and the conditional probability rule, we finally find:

$$G_Y(y, k) = \begin{cases} \frac{e^{\lambda Dk} - e^{-\lambda D y}}{1 - e^{-\lambda D}}; & 0 \leq y \leq 1 \\ 0 & \text{; elsewhere} \end{cases} \quad (5)$$

Consequently, the corresponding probability density function $g_Y(y, k)$ will be given by:

$$g_Y(y, k) = \begin{cases} \frac{\lambda D}{1 - e^{-\lambda D}} e^{-\lambda D y}; & 0 \leq y \leq 1 \\ 0 & \text{; elsewhere} \end{cases} \quad (6)$$

It can be easily seen that $g_Y(y, k)$ is independent of k . Thus, we have:

$$f_Y(y) \equiv g_Y(y, k) \quad (7)$$

Mean of Filled Portion Y

Using the definition of the mean and Eqs. (6) and (7), an expression for the mean μ_Y of the random variable Y may be obtained as:

$$\mu_Y = E\{Y\} = \int_{-\infty}^{\infty} f_Y(y) dy = \int_0^1 \frac{\lambda D}{1 - e^{-\lambda D}} e^{-\lambda D y} dy \quad (8)$$

After evaluating the integral, we find:

$$\mu_Y = \frac{1}{\lambda D} - \frac{1}{e^{\lambda D} - 1} \quad (9)$$

RESULTS

System Description

The simulated system consists of N independent and identically distributed (i.i.d.) voice sources following the model shown in Figure 4. These sources are assumed to be carried over an N-ISDN network. Then, through an IWF consists of detecting talkspurt periods, then checking for the availability of bandwidth on the broadband side, and then starting the conversion of talkspurt packets into ATM cells. If the last cell resulting from this conversion is unfilled with voice data, then stuffing bits are added to fill it. Also, if a talkspurt cannot be accommodated, we assume that data will be lost (front-end clipping). Once it senses that the source state is changing to silence mode, it releases the bandwidth allocated to the pertinent source.

Comparison with Analysis

In Figure 7 are depicted the average percentage of stuffed cells for various vocoder rates using both the analytical in Eq. (9) and simulation results.

The agreement between the curves obtained using the two techniques is very clear. It is, however, noticeable that the simulation results always under-estimate the theoretical ones. This is mainly due to the front-end clipping which is ignored in our analysis. As

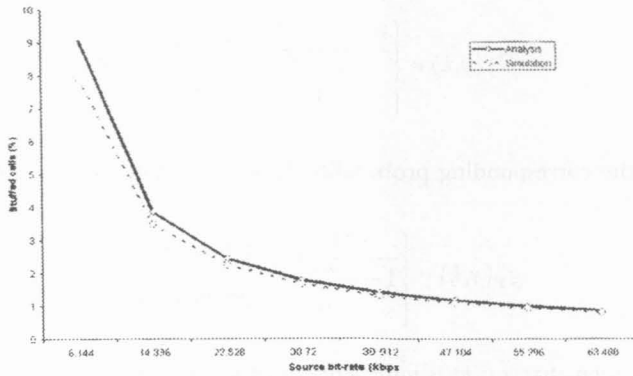


Fig. 7. Comparison of simulation and analysis results.

the front-end clipping effect gets stronger at low encoder rates (as will be seen later on), the difference between the two curves gets larger.

Effect on Utilization

First, we tried to change the average bit-rate generated by each source, and studied the resulting change in the relationship between the number of source, and studied the resulting change in the relationship between the number of sources and the utilization of the channel, which is used as a performance measure. As shown in Figure 8, from a 64 kbps bit-rate up to 14 kbps there is a negligible improves significantly. Also, we notice that this improvement stay constant until the number of multiplexed sources become 60, and then starts diminishing until it becomes zero at about 90 sources, where the utilization is practically 100%

The lastly mentioned results are redisplayed in a clearer way in Figure 9, where we can easily see the advantage of the multiplexing process. As the number of sources increases, the utilization of the channel increases and approaches almost 100.

Effect on Bandwidth Wasted in Bit-Stuffing

Our second look is at the effect of the number of multiplexed sources on the bandwidth wasted due to the use of cell padding by introducing bit-stuffing. As can be easily noticed

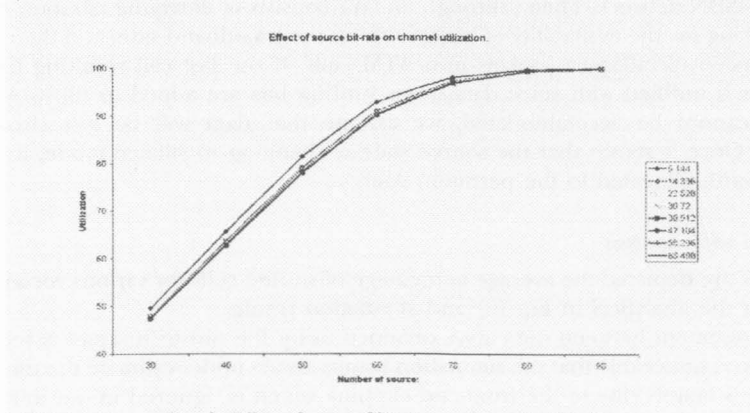


Fig. 8. Effect of source bit-rate on channel utilization

Voice Compression on the Operation of a N-ISDN/B-ISDN IWF

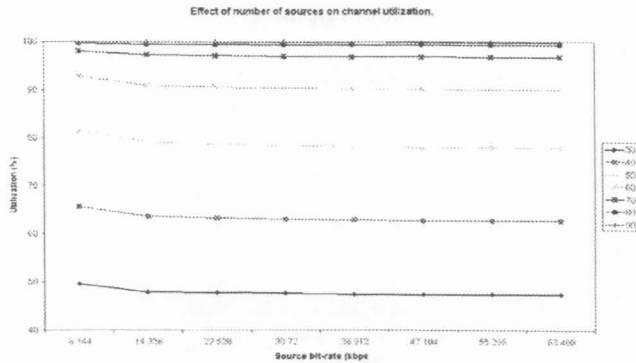


Fig. 9. Effect of the number of sources on the channel utilization.

in Figure 10, the performance of the system, vis-vis the percentage of stuffed cells, is practically independent of the number of multiplexed sources.. However, it is highly affected by the bit-rate generated by the sources. As the bit-rate decreases, the percentage of stuffed cells increases drastically causing a sharp decay in the system performance.

The effect of the source coding bit-rate is clearly depicted in Figure 11. In the range of 64 kbps coding rate down to a rate of 22 kbps, there is only a minor decrease in the performance of the system. Then, a sharp decrease is incurred at the rates of 14 kbps and 6 kbps, and evidently any rate below that.

Effect on Lost Cells

Due to the fact that we are using speech silence detection, after a source is accepted into the network, there are times where when it becomes active it does not find any available bandwidth. In such cases, the front-end of the talkspurt gets clipped (as we are not allowing any queueing). In Figure 12, we are showing the effect of the sources-coding rate on the percentage of lost cells. Here also, the effect of the sources-encoding rate become noticeable only when the rate is decreased below 22 Kbps, above that the effect is very negligible.

From both Figures 12 and 13, we can see that as the number of sources is increased above 50, the percentage of lost cells increases sharply and almost linearly.

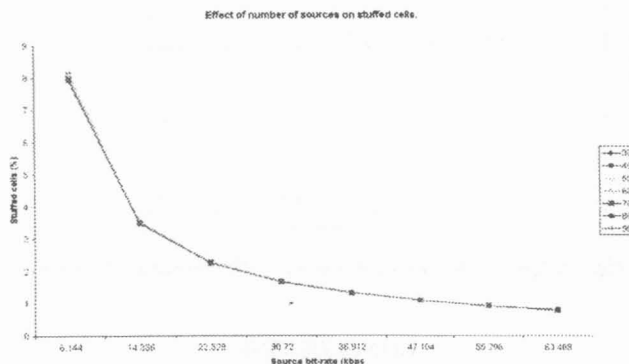


Fig. 10. Effect of the number of sources on the percentage of stuffed cells.

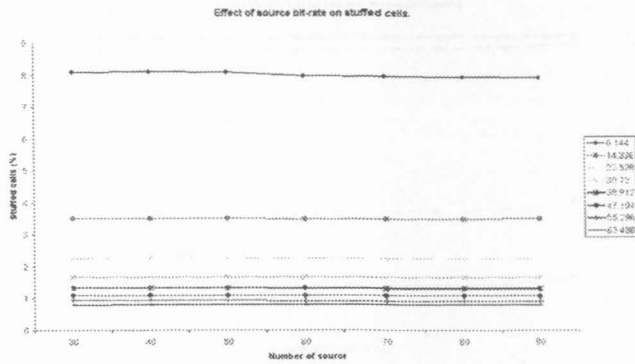


Fig. 11. Effect of the source bit-rate on the percentage of stuffed cells.

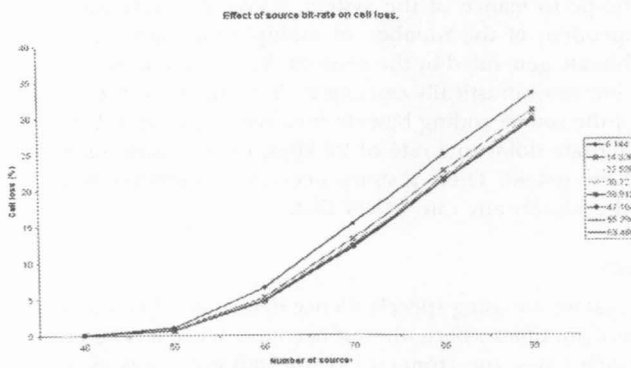


Fig. 12. Effect of the source-coding rate on the percentage of lost cells.

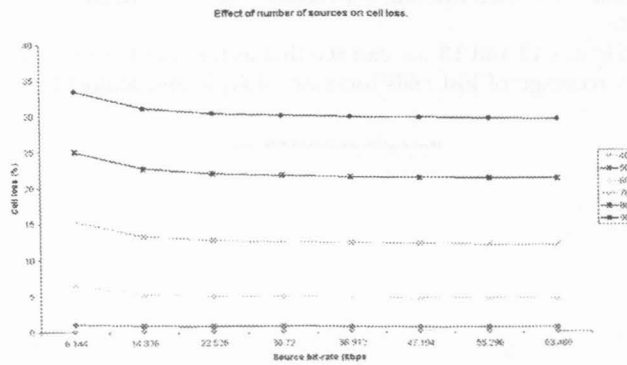


Fig. 13. Effect of the number of sources on the percentage of lost cells.

CONCLUSIONS

We have presented and evaluated a new technique in dealing with voice sources originating on an N-ISDN network and being carried on a B-ISDN network. As the

sources considered are of the VBR types that use both Speech Silence Detection and adaptive compression system. It is very hard to operate transparently over the B-ISDN. We tried to assess the degree of effect introduced by these networks.

The effects of decreasing the source-encoding rate on the system performance was on one hand worsened for front-end clipping and bandwidth waste due to stuffed cells. On the other hand, it was improved slightly for the system utilization. The effects of increasing the number of voice sources on the system utilization, and stayed practically unchanged for the percentage of stuffed cells. In our future work, we will try to study the system under other techniques that deal with the unfilled cells. Also, we need to consider other important issues as the delay incurred by the voice signals within the IWFs on the two sides of the B-ISDN network.

REFERENCES

- BERNHARD PETRI and DIETER SCHWETJE. Narrowband ISDN and Broadband ISDN Service and Network Internetworking. *IEEE Communications Society Magazine*, June 1996.
- BRAZILAI T. P. et-al. DESIGN and Implementation of an RSVP-Based Quality of Service Architecture for an Integrated Service Internet, *IEEE Journal on Selected Areas in Communications*, pp. 397-413, April 1998.
- Circuit Emulation Service Interoperability Specification (CESIS). 1997 v.2.0, af-vtoa-0078.000, January .
- DAVID J. WRIGHT. Voice over ATM: An Evaluation of implementation alternatives, *IEEE Communications Society Magazine*, May '96.
- DEMPSEY B.J., J. LIEBEHERR and A. C. WEAVER. A New Error Control Scheme for Packetized Voice over High-Speed Local Area Network, at <ftp://ftp.cs.virginia.edu/pub/techreports/CS-93-23.ps>.Z
- MEI M. and D. EVERITT. 1996. Voice traffic in timed token protocol local area network, ICC'96, pp. 1078-1082 Vol.2.
- PRIVATE INTEGRATED SERVICE NETWORK (PISN). 1998. Circuit Emulation Specification- Emulation of Basic Access by ATM Network, ECMA-227, June 1998.
- QIGANG Z. , Z. XIKONG and Z. SHUJING. 1996. , Performance analysis of an integrated data/voice protocol on bus LANs. ICCT'96, pp. 634-637 Vol.2.
- RFC 1889. RTP: a transport protocol for real-time applications, January 1996.
- RFC 1890. RTP profile for audio and video conferences with minimal control, January 1996.
- RFC 2326, Real Time Streaming Protocol (RTSP). April 1998.
- SAKKE OKUBO, et al. 1997. ITU-T Standardization of Audiovisual Communication System in ATM and LAN Environments, *IEEE Journal on Selected Areas in Communications*, pp. 965-982, August 1997.
- VISSER M. A. and M. EL ZARKI. 1995. Voice and data transmission over an 802.11 wireless network, Personal, Indoor and Mobile Radio Communications, 1995 (PIMRC'95), pp. 648-652 Vol.2.
- Voice and Telephony 1997 over ATM to the Desktop, af-vtoa-0083.000, May 1997.
- Voice and Telephony 1997 over ATM-ATM Trunking using AAL1 for Narrowband service, af-vtoa-0089., July 1997.
- WHITE P. P. RSVP and Integrated Services in the Internet: A Tutorial, *IEEE Communications Society Magazine*, pp. 100-106, May 1997.