Coding/Compression Requirements from the Network Protocol's Viewpoint

Yee-Hsiang Chang
Communications Research
MCNC Center for Communications
Research Triangle Park, North Carolina 27709

Abstract

Several major efforts are under way in the Internet community to bring real-tim communications into reality. Most of the efforts are targeted to solve network-r issues and leave the coding/compression problems to be answered by industry market value becomes obvious. However, at this point, most of the codec vend focusing on the constant-bit-rate network, which is the current telephone netw infrastructure. In this paper, we will demonstrate the importance of developing bit-rate compression algorithms and standards. This effort is not only for user Internet community, but also for users in the next generation of the telecommu network (broadband ISDN), which uses ATM (Asynchronous Transfer Mode) te Due to a strong dependency between coding/compression and networking for v rate networks, we will attempt to specify the requirements for the compression designs and experiments from the networking point of view.

1. Introduction

The current Internet, based on packet-switched technology, possesses an async
nature (also called variable-bit rate services). It is very desirable to design com
algorithms to take advantage of this characteristic. However, none of the codin
currently available (e.g., H.261 [LIOU91; FOX91], MPEG [GALL91; FOX91], JPEG
[WALL91; FOX91]) are designed for this purpose.

In fact, all the compression schemes, in their natural form, generate asynchron
For synchronous networks, these data are buffered and sent out in constant rat
actually easier to design the compression algorithm for an asynchronous netwo
synchronous one. One example is the data generated by the ADPCM (Adaptive I
Pulse Code Modulation) scheme. ADPCM is used for voice and video
coding/compression, which transmits only the temporal difference between the
voice signal and the previous one. The signal difference varies according to th
conversation, and results in a variable data rate.

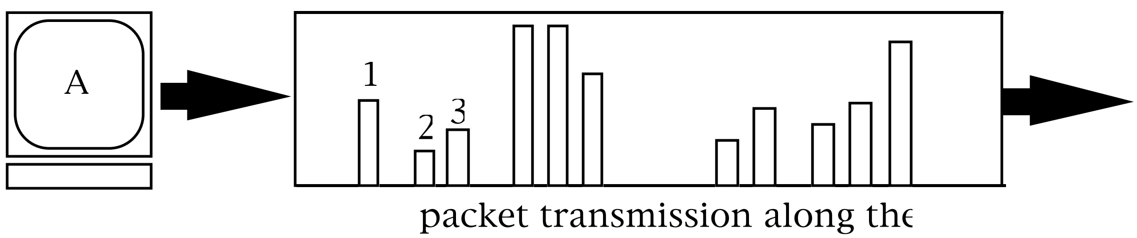
In next few years, other areas will also drive coding/compression development
asynchronous networks. One area is LAN-based networks, which are in the cat
asynchronous networks as well. According to most analysts, between 29 and 4
computers will be on LANs in the U.S. in 1995 [FORR91]. Another area is the fu
BISDN (Broadband Integrated Service Digital Networks), which will be based on
(Asynchronous Transfer Mode) and also will be asynchronous in nature. For th
we believe that variable-bit-rate compression is the trend for video compression
future.

In the Section 2 of this paper, we will show the basic characteristics of a variabl
network and demonstrate the dependencies between the compression algorithm
network traffic. In Section 3, we will list the requirements needed to design a c
asynchronous networks.

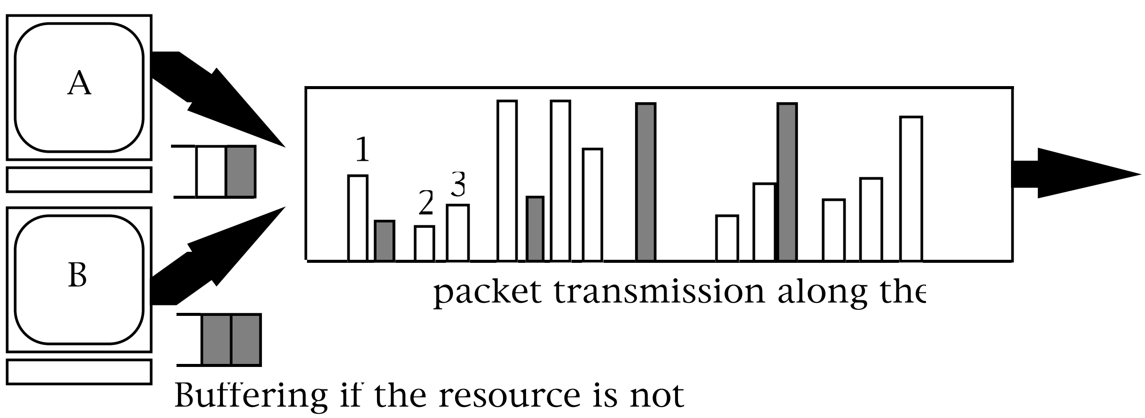
2. The Major Characteristics of the Variable-Bit-Rate Networks

The major characteristic of an asynchronous network is a traffic pattern with st
opposed to deterministic) interarrival characteristics (Figure 1a). If there is onl

sending the data through the network, the traffic pattern is the same as the traf there are more sources using the same link, their traffic will be multiplexed toge according to the time any individual message arrives. If there are multiple mess the same time, some of them will be buffered and delayed (Figure 1). When the so heavy that the buffer can not handle all the data, some messages will be drop further delayed through a congestion control mechanism. The delay can be dif each packet. For example, message 1 in Figure 1 is delayed due to multiplexing traffic, but messages 2 and 3 do not have any delay. This variation of the delay messages is called jitter. Jitter is not desirable for applications when it varies tc general, the asynchronous network environment provides better network utiliz introduces the potential for message delay and lost and out-of-sequence messag the basic idea of statistical multiplexing and resource sharing in packet-switche



(a) Traffic that has only on



(b) Traffic combined by two

Figure 1. Delay and Congestion Caused by Asynchronous Networks.

For this kind of environment, an optimal scheme should be designed to work w network having statistical traffic characteristics. This is different in synchronou environments, in which a constant-bit-rate channel is allocated for every applic synchronous network, once this type of channel is set up, the application does interact with the network. It only has to regulate its flow. In the case of asynch networks, the compression scheme has to be designed to interact with asynchro networks to deal with the extra delay generated by buffering as well as with me due to congestion.

To support real-time applications, we believe most future networks will have re management schemes to support timing requirements. These management sch information from the application to specify its traffic pattern, which normally ir average rate, peak rate, burst period, and jitter. There is a stronger relationshi application and the network for an asynchronous network compared with a syr network. To illustrate this point, consider the case of compression of compress schemes. For an synchronous network, the compression scheme needs only on - bit rate. For an asynchronous network, the compression scheme needs four p

3. Relationship Between Coding/Compression and Networking

The following is a list of items that show the relationship between coding/comp networking.

Timestamping information should be provided by the compression/coding function [CASN91].

The compressed data for every frame has to be marked with a timestamp. The purpose is for timing recovery of messages at the receiver side, considering the in the network and buffering at the receiver side. Furthermore, the receiver ca synchronize media from different sources without the timing information. This also can be used by the networking layers to know the time limit for transmittin forwarding the message across networks, which is a way to reduce the jitter.

For using timestamping, an accurate global clock for each host is required. Thi done in the Internet community with the network time protocol [MILL89], which

synchronize the network clock to millisecond accuracy. In the future ATM envi
we expect the network will provide this service.

Packetizing information should be given by the compression/coding function.

Indication for the same packet: The compressed data for each unit of the frame must b
marked to show that they belong to the same frame. For example, the subband
divides the image into several frequency bands. These bands then are packetiz
several packets. There should be an indication to show that these packets belor
same frame.

Match compression unit (block size) with the network segment size: Another example is
that some compression schemes divide the image into blocks such as DCT (Disc
Transform), then send the blocks over networks. Because the network will segr
data according to its characteristics (e.g., one ATM cell has 48 bytes of data), th
should be matched closely with the network segment size.

Indication for the location of the same image: The information that shows the location of
blocks in the same image must be given to the network protocol layers. In this
will be higher tolerance of out-of-order messages for blocks in the same image,
doesn't make any difference for the receiving end which blocks are within the s

Good error tolerance is needed in both compression/coding and network levels

There are two different kinds of network errors. One kind is bit errors, which
the same way for both synchronous and asynchronous network environments.
error pattern exists only in the asynchronous network environment. Due to the
nature of the environment, the asynchronous network is more likely than the s
network to have packets lost, out of order, and delayed. These phenomena are
desirable because once the compression ratio becomes higher, the system is less
errors. There are several places where this problem can be dealt with this. One
compression/coding layer. Some adequate redundant information can be adde
recovery. However, we don't want to add so much information that we nullify
from compression. The other place is the network layer. The application shoul
specify its reliability requirements to the network. This reliability index should
a fully reliable state to an unreliable state. For example, the ATM uses the *lost pr*

to show the degree of reliability. So, a strong relationship between compressio
control and network layer error control needs to be defined. For example, the
layer should be able to scale the error control functionality up or down to fit wi
network layer error control (Figure 2).

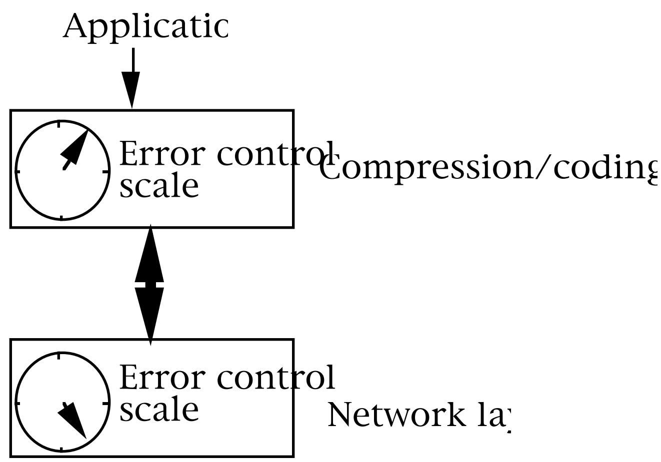


Figure 2. The Relationship Between the Error Control Adjustment in both La

Scalable performance is needed.

The image or video/audio quality should be scalable once the network bandwid
available (say from 19.2 Kb/s to 300 Mb/s). Because the future ATM network
bandwidth on demand, the compression schemes should be able to scale smoot
example, the standard of H.261 (also called Px64) is scalable to multiples of 64
other hand, the current MPEG 1 standard specifies the bit rate around 1.5 Mb/
MPEG 2 has the bit rate around 10 Mb/s. Although there is a range can be adju
MPEG 1 and 2, neither of which is flexible enough to scale in the whole range.

Interoperability among various performance levels is required.

If a host has a connection to a high-bandwidth network and is talking to two ot
the same time who are transmitting at different rates because of their network
this host should be able to decompress the low- and high-bandwidth traffic wit
problem. We call this "interoperability" among different bandwidth levels. If
as the example, the receiver who can send at 128-Kb/s rate should be able to re
Kb/s signal without any problem.

Adaptive coding schemes are needed.

One relationship with the network is adaptivity related to network traffic. If the temporarily congested, this information should be able to feed back to the code the traffic (and quality) and then come back to the original quality once the con over. Because the current codec already has a feedback mechanism to adjust its compression from the output buffer occupancy information, we need to extend feedback mechanism from internal buffer to networks. The adjustment of the c also needs an implicit or explicit handshaking between the sender and the recei changing compression parameters. It is worthwhile to note that the network co function must take the network round trip time delay and the duration of the c account. If the congestion is transient relatively to the round trip delay, an end adjustment will not help. It will help when the congestion is longer than the ne round-trip delay. When the congestion lasts about the same time as the round- there is a possibility that the control will cause oscillation.

Ways of coping with network resource management parameters are required.

Future networks will require the application to specify its traffic parameters. Fo there have been suggestions to use average rate, peak rate, burst period, and jit traffic parameters. These parameters are used by the network to figure out the of network resources. The compression algorithm should take advantage of the of the traffic parameters to design a lowest cost coding scheme.

Compression should specify the priority to be put on packet header.

The network should support priority levels to distinguish different real-time rec of the applications. The compression/coding scheme should take advantage of and specify the priority of the compressed data to the network layers. In this w network can map the priority requirement into packet headers. For example, th progressive image coding and transmission [DREI87; HUAN90] is first to send in features of the image followed by the less important ones. This scheme can use level supported by the network to make sure the important parts arrive on time important parts can have lower network priority. The layer coding scheme or h coding scheme will also benefit from the network priority.

Ways of dealing with multipoint (multicast) requirements are needed.

The current solution for multipoint conferencing does not depend on the netwc multicasting function. Its configuration is similar to the one in Figure 3. A cent point-to-point connections to each participant. The audio and video messages c each source will go through the process of audio mixing and video switching an redirected to all the participants.

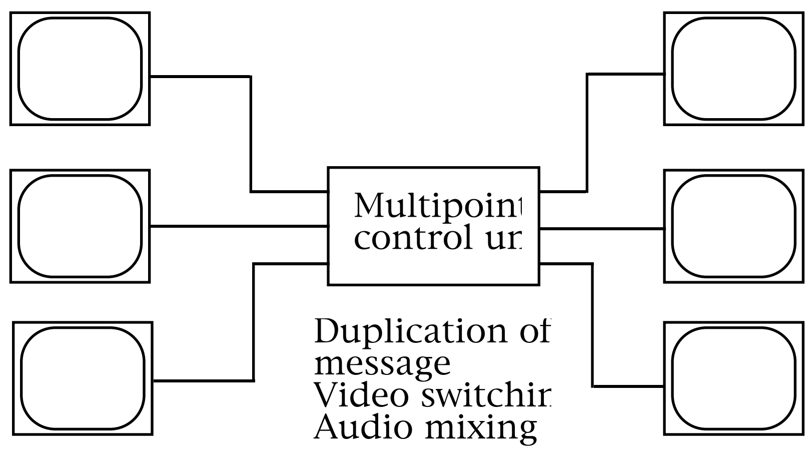


Figure 3. Current Multipoint Control.

For networks that have multicast capability (i.e., the network can deliver the me one source to multiple receivers), we need to add only video switching and aud ability. The current Internet has been implementing multicast capability [DEER8 several ATM switches are incorporating this feature as well [FORE92; TURN88]. where should the audio mixing and video switching functionality be in the Inter ATM network environments? Do we still need a centralized system? The answer because centralized solution is not flexible enough to support different numbers participants. We feel that these functions should be distributed to each end (Fi

One solution similar to this one is from IBM [KAND92]. Although there are mult signals coming from every participants to a receiver, these signals are combined single frame for update. In this case, only one codec is necessary at the every r This technique currently applies only on motion JPEG, which should be extende coding/compression algorithms.

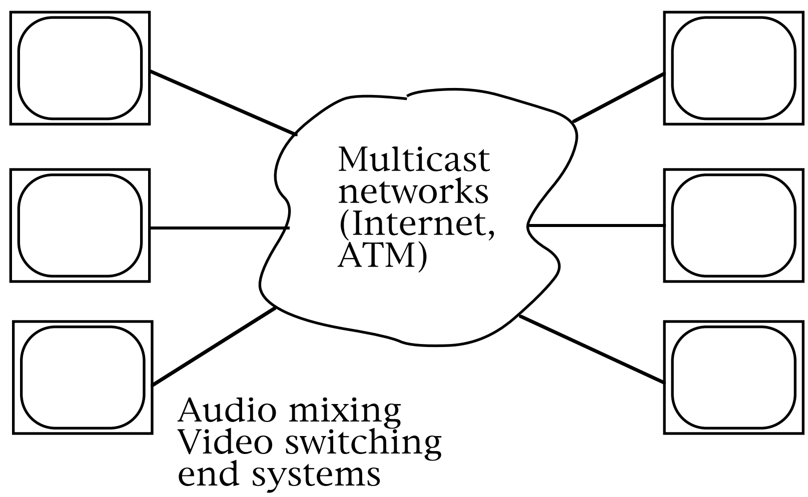


Figure 4. The Future Multipoint Control, which will Use the Network Multicast

Software implementation for experiments is needed.

Software implementation will not be required for future production use but is q
important at the early stages for experiments. We need a software codec implen
that can be used to test the integration of coding/compression function and the
based on the current workstation technologies. A possible next step solution af
software implementation is to build video-based DSP (digital signal processing)
processors.

Other Issues not Directly Related to Networking.

Support multiple coding format [WROC91]. The codec has to support several different coding formats.

Real-time echo cancellation. Real-time echo cancellation is important when the networ
delay is large, which is normally the case in wide-area communications. We nee
real-time echo cancellation to get rid of the echo. Ideally, this function should b
codec.

Audio Mixing and Video Switching and Integration with the Window System. If the future
multimedia station is on a personal computer or workstation, some functions ha
incorporated into the window system (e.g., X window). For a conferencing sess

multiple participants, there should be a way to show all the participants' images screen, and indicate the person who is currently speaking. For example, we can this person's talking-head window. In other words, we can add this function to for video switching and audio mixing.

4. Conclusion

In this paper, we have shown the motivation to push the development of compression/coding algorithms and standards for the asynchronous network. demonstrate basic characteristics of asynchronous networks, which require sign interaction with coding/compression schemes. The major contribution is a list requirements and possible opportunities for the coding/compression design for purpose. We believe there is a need to remove the barriers between coding/con researchers and networking researchers. In order to accelerate the developmen multimedia communications, a mutual understanding of each others' requireme important. This is especially true for the asynchronous network environment. is an effort to specify the coding/compression requirements from the network point of view.

5. Acknowledgements

I would like to thank the following persons for their technical comments on this which makes the codec requirements more complete. They are Dan Winkelstein Casner, and Dan Stevenson.

References

[CASN90] Casner, S., Seo, K., Edmond, W., and Topolcic, C., "N-Way Conferencing with Packet Video," *Proceedings of The Third International Workshop on Packet Video*, March 1990.

[CASN91] Casner, S., "Higher Layer Protocols," *Proceedings of Packet Video Videoconference Workshop*, MCNC, Research Triangle Park, NC, December 1991.

[DEER89] Deering, S.E., "Host Extensions for IP Multicasting," RFC 1112, Augu 1989.

[DEER90] Deering, S.E., "Multicast Routing," *ACM Computer Systems*, 1990.

[DREI87] Dreizen, H.M., "Content-driven Progressive Transmission of Grey Le Images," *IEEE Trans. Communications*, vol. COM-35, March 1987, pp. 289-296.

[FORE92] Fore Systems, "SPANS: Simple Protocol for ATM Signaling," Fore Systems, 1992.

[FORR91] Forrester Research's Network Strategy Reports, "The Network Strate Report: LANs for Free?" November 1991.

[FOX91] Fox, E., "Advances in Interactive Digital Multimedia Systems," *IEEE Computer Magazine*, October 1991, pp. 9-21.

[GALL91] Gall, D.L., "MPEG: A Video Compression Standard for Multimedia Applications," *Communications of the ACM*, vol. 34, no. 4, April 1991, pp. 47-58.

[HUAN90] Huang, Y., "Prioritized Source Coding and Adaptive Progressive Ima Transmission," Ph.D. Dissertation, Illinios Institute of Technology, Chicago, Illinois, August 1990.

[KAND92] Kandlur, D.D., "MMT System Architecture," Proceedings of the Seco
Packet Video Workshop, Research Triangle Park, N.C., December 19

[LIOU91] Liou, M., "Overview of the Px64 Kbit/s Video Coding Standard," *Communications of the ACM*, vol. 34, no. 4, April 1991, pp. 59-63.

[MILL89] Mills, D.L., "Internet Time Synchronization: The Network Time Proto
RFC 1129, Oct. 1989.

[TURN88] Turner, J.S., "Design of a Broadcast Packet Switching Network," *IEE*
Transactions on Communications, vol. 36, no. 6, June 1988.

[WALL91] Wallace, G.K., "The JPEG Still Picture Compression Standard," Communications of the ACM, vol. 34, no. 4, April 1991, pp. 30-44.

[WROC91] Wroclawski, J., "The Integrated Service Internet Project: Work in Pro
at MIT, Xerox PARC, LBL, Bellcore," *Proceedings of Packet Video Videoconference Workshop*, MCNC, Research Triangle Park, NC, December 1991.