Proceedings of the ACM International Conference on Multimedia,
ACM MM'2001, Ottawa, Canada, October 2001

Resource Adaptive Netcentric Systems: A case Study with SONET- a Self-Organizing Network Embedded Transcoder

JAVED I. KHAN, SEUNG S. YANG, QIONG GU, DARSAN PATEL,
PATRICK MAIL, OLEG KOMOGORTSEV, WANSIK OH, & ZHONG GUO

Media Communications and Networking Research Laboratory
Department of Math & Computer Science, Kent State University
233 MSB, Kent, OH 44242, javed@kent.edu

Abstract

In this paper we discuss architecture for network aware adaptive systems for next generation networks. We present in the context of a novel cognizant video transcoding system, which is capable of negotiating local network state based rate and let the video propagate over extreme network with highly asymmetric link and node capacities utilizing knowledge about the network, content protocol and the content itself.

Key Words: Adaptive Video Applications, Asymmetric Internet, Active Network.

1. Introduction

Adaptation is a fundamental phenomenon is natural systems. It seems that engineering of any large and complex system intrinsically requires inbuilt ability of its components to adapt. Internet has already grown into a meganet with global reach. Now with the emerging need of advanced applications it is poised to evolve into a complex system of systems. With its expansion the asymmetry of the Internet is also increasing. Historically the initial Internet architecture has been conceived to cope with the heterogeneity of network standards [15]. No sooner we thought the problem has been caged, it now appears a second era is evolving. It seems next generation of Internet will have to deal with more intrinsic (and perhaps harder to overcome) heterogeneity— the asymmetry of hard network resource such as bandwidth, or switching capacity [20]. This asymmetry can evolve from the fundamental physical limitations such as the power crunch in an intergalactic network element, or from something as close and insurmountable as socio-economical disparity- the digital divide.

In this research, we investigate a concept application that concentrates on creative adaptation[1]. We present an MPEG-2 rate transcoding mechanism, which addresses the issue of adaptation from two levels. It adapts with respect to two critical network resources—bandwidth and the processing resource at the junction nodes. While the link bandwidth adaptation has been addressed up to some extent in few of the recent research little attention has been paid to the node capacity adaptation.

The transcoder senses local asymmetry in link capacities at various junction points of a network. Based, on that it accordingly adapts the video stream rate. In the second level, the transcoder also senses the local computation power to execute its rate adaptation task. And thus, based on the network computational power it demonstrates self-organization behavior. In each of these adaptive behaviors it employs a number of techniques. For rate adaptation in the first stage it uses full re-quantization based transcoding. For extreme rate scalability it further employs a focal object based region discriminating encoding. To adjust with the processing power problem, it first can shift back to a low computation mode of transcoding using motion vector computation bypass. However, when a single node becomes insufficient dynamically migrates computations to neighboring nodes in search of increased processing power. While we are building the prototype of this concept system, in the process we are identifying the requirements, for a next generation network architecture for adaptive systems.

2. Pervasive Video Streaming

2.1 Asymmetry in bandwidth:

The bandwidth disparity between various segments of the network is increasingly dramatically. The advent of non-traditional devices already demonstrates the limits of current networked applications [19,20]. Technology is needed so that pervasive applications can be designed which will operate irrespective of the diversity and should not stop operating when encountered with network with 10-100 times lower bandwidth. Unfortunately, for most demanding applications such rate adaptation cannot be conveniently performed with the current network level operation. When congestion arises, generally routers drop packets. However, current routers do not consider the applications view and treats them as unrelated items. Because of complex inter packet data dependency dropping only 20-25% of the UDP packets can render an entire stream useless [12,2,18], though a network accounting can falsely continue to show high throughput. Effective rate adaptation requires reduction of the content, which inherently requires content awareness. Unfortunately, currently content awareness is only available at the network end-points. There are only three strategies possible from the end-points- none of which is satisfactory. Sending only the high-speed stream cuts off the low speed clients. Sending the

[1] There are quite a few works tending adaptive systems—particularly in the areas of scalable video communication, web caching, and very recently the wave sparred by mobile information systems. Many however, depend on resilience by means of redundancy of information or of onboard processing logic to survive variation in transport characteristics. Less they depend on network awareness, sort of accepting the statuesque. Mostly for space limitation we could not add review of these systems here.

lowest speed stream penalizes others by forcing the lowest quality to all. Sending multiple streams, burdens the network. Fig-1 illustrates the non-optimality for the *heterogeneous clientele* scenario. It seems effective rate adaptation requires the technique to come from deep content awareness, but the resulting action to be committed in close proximity of the impairment-- both in the senses of time and space. Consequently we choose a network embedded video transcoding. A **network embedded transcoder** can solve this riddle and provide optimality even at the level of individual links.

2.2 Asymmetry in Computational Power:

It seems that in near future with the rapid advancement of the VLSI technology some nodes may be able to garner enough processing power for real-time high fidelity video transcoding. However, on any given large network the fact of the matter is that there will be always inequality of processing capability just like the asymmetry in bandwidths. Consequently, it is also important to develop applications, which is capable of **self-organization.** Transcoding is inherently computation intensive. Logically transcoding operation included both decoding and encoding. A number of techniques have been investigated for accelerated transcoding by us and other researchers, such as fast DCT domain transcoding, parameter bussing [11,7,8,6]. A video transcoding is a three way-trade-off between the quality-loss, rate reduction, and the computation involved. The proposed system should be built with as little possible expectation about the computational power of the underlying network elements. It should be flexible enough to choose the right operating state in this trade-off based on the network processing resources available at the junction points.

2.3 Dynamic Adaptation, Self-Organization and Seamlessness:

The probability of dynamic change in the environment increases with the size of the system and the duration of the session lifetime. Therefore, in a large and complex system the operational modules have to be built in such a way that it can respond to dynamic changes. This calls for at-least two important issues. First of all network layer services are required which will can extract network local states. Also it is important to note that these services must be available to the adaptive elements not only to the parent layer to the application. Adaptive elements should be built in a way that while underlying dynamics changes the adaptation and reorganization is minimally disruptive to the service abstraction it itself provides. For example, if the processing power at the node changes the migration of the computation should occur with minimum disruption of the video carriage.

3. System Architecture

Recently we are building a transcoding system to gain increased understanding about adaptive complex network systems and in particular identify the requirements of the next generation network infrastructure, if it were to support **resource adaptation**. We have further carefully selected a subset of features to be incorporated into this experimental system to understand the principles of complex networked system's adaptation. For the rate adaptation we study the impact and mechanics of the infusion of content-awareness at two grades—

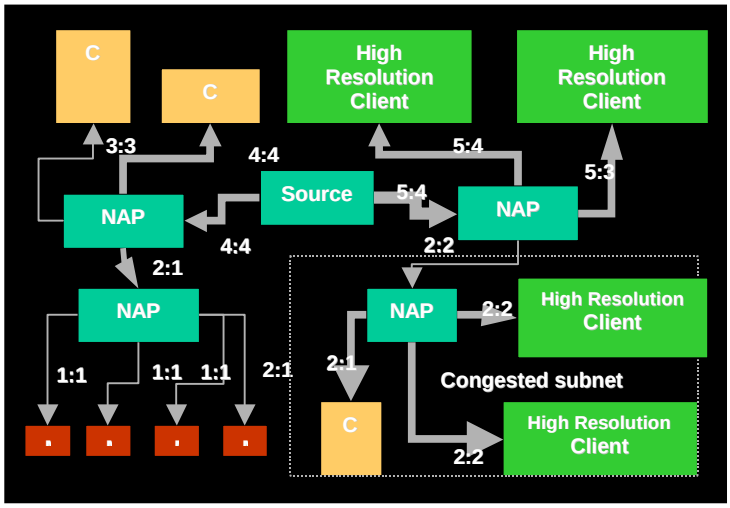

Fig-1 Video multicast in a large network with heterogeneous links and clients. Labels show the capacity and flow. Network embedded rate adaptation (NAP) minimizes the net traffic and maximally satisfies client requirements. This problem has no optimal solution in end-to-end rate adaptation paradigm.

awareness about the content protocol and awareness about the content itself. For the compute power adaptation we study- the impact and mechanics of **modular self-organization** and **computation diffusion**. The system's adaptive ability is built into two architectural layers. Below we first provide a close functional description of the transcoder, and then we discuss the systems engineering.

3.1 Active Networking

The scope of the research and the proposed concept system is such that where many of the required supports as expected are not available in the current network infrastructure. Consequently, we have selected conceptual framework of active network as the base system, which at the moment promises ability to embed programmed capsules into network [9,10,6]. Our implementation is based on a user space realization of it.

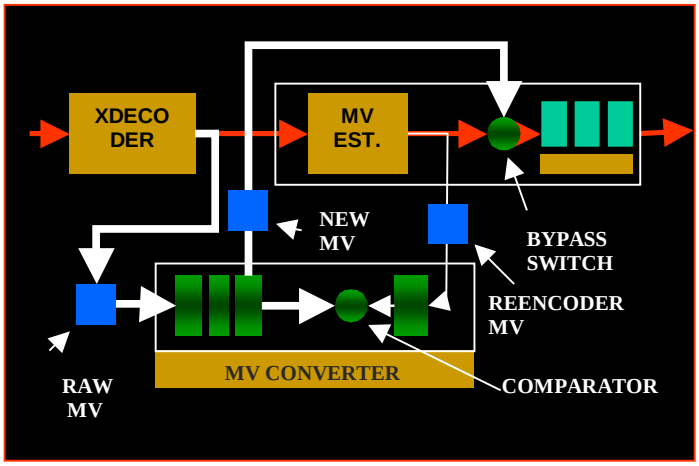

Fig-2 Functional diagram of the cognizant transcoder. It can accept plug-in processors for content aware video transformation.

3.2 Transcoder Architecture

Logically a transcoder is a cascaded decoder and encoder. However, the re-encoding process can be made different from conventional encoding. Fig-2 shows the schematics of the transcoder. The transcoder activates a number of sample processing techniques for rate reduction based on the extent of rate reduction required. The final unit uses a dynamic piece-wise CBR (constant bit rate) rate control mechanism. Like TM-5 [5], it is a feedback control mechanism, which adjusts the quatization steps on group of GOP basis. At the start of each GOP it reinitializes its target bit allocation, and the carryover is discontinued after every specified number of GOPs. Pure requantization based reduction [7,11] is inadequate for larger downscaling. For second stage compression we offer a sample tiling mechanism that achieves about 1/50 times rate scalability [6]. In this mechanism, the quatization process is further being integrated with a region based sample fusion technique—both in time and in spatial dimensions.

There is another experimental feature in this transcoder. For low bit-rate transcoding the object becomes increasingly important for perceptual quality [16]. For this case, rate control process assumes that object information (either by high level image analysis or even by direct feedback) is available about the perceptual significance of the various spatial areas of the video. Correspondingly, at low rate, bits are then taken out selectively based on the object significance. An area identified with higher significance is last to loose its bits. In current TM-5 the activity parameter designed to account for Human Visual System (HVS) is calculated as simple variance of the pixel values. In our rate controller we have enhanced the model. The system can accept perceptual foveation information in two ways. It can explicitly accept object definition from an interface. Also recently we have integrated a motion based object processing logic module, which extracts object information from macro-block motion and accordingly modulates individual *macroblock activity* parameters of the rate controller. Fig-3(a) and (b) respectively show the difference created by this scheme. For two sample frames[2]. Both the frames have same number of bits but the region-wise distributions are different (area SNR shown in Fig-4(a) and (b)).

The above mechanisms provide cognizant means for enhancing the rate adaptation ability of the system. The overall computational task is nevertheless daunting. The first level of adaptation for computation power is offered by the optional motion vector bypass mechanics. When switched to this mode, the transcoder can also extracts the motion vectors [17] from the incoming stream and instead of full search, uses them to estimate the new motion vectors. Based on the search space specification it reduces (25-40% computation). Because of the existence of the complex frame dependency, and existence of the intra-frames the extracted motion vectors cannot be used directly. We use several preprocessing. These preprocessing is done in the motion converter unit, which substitutes the motion estimation unit of full logic encoding. It accepts a series of motion processing converter plug-ins. It works in parallel with the macroblock-processing pipe of the main encoder section. The bypassed motion vectors extracted from the incoming streams are fed into them. As the original macro-block data passes through tiling, temporal sub-sampling, spatial tiling, and requantization, etc. stages, similarly the motion vectors in parallel passes though these converted to match the final macroblock forms. A second level of complexity adaptation is performed by another novel aspect of this system-- computation migration. If the processing power available/allocated in a current network point becomes inadequate if needed part of the transcoding operation can be further migrated into multiple network nodes. This is achieved at systems level to be described next.

[2] Not sure if the difference will show in the printout. On screen more details appears in the center. Compare the face, the white jaw line of the whale, and the tone of the water.

3.3 System Components

The overall system has been architect with a rather unique view so that the entire transformation (and the associated transcoding operation) on a video stream during its passages can diffusely take place on a (active) subnet rather than on a single network point. This provides important benefit in resource adaptation. The key to the design is the modular decomposition of the transcoding operation—where instead of a single monolithic implementation we took an experimental approach of building them as dynamic hyper-linkable capsules with easily separable data flow optimized concurrent modules. Within the main units we similarly allowed for dynamic hyper-linkable plug-ins. In our current implementation the main units are (a) GOP-Encoder (GOP-ENC), (b) decode-demultiplexer (DE-DEMUX) and (c) GOP multiplexer (GOP-MUX). One ENC-MUX and one GOP-MUX are activated at the designated sub-net entry and exit points. However, multiple instances of the GOP-ENC units are then kept in a dormant state in designated active nodes for seamlessness. Active modules are then connected by a feedback system. The GOP-MUX module senses the resultant frame rate. If the current aggregation is less than a given target frame rate, it then sends signals to DE-DEMUX unit, which then activates dormant GOP-ENC units in the active subnet to join in. One of the complex tasks here is the coordination. To keep it manageable in this experimental implementation we maintain a single Meta-Controller (MC), which is responsible for user interaction, system initiation and dismantling and setting up initial boundary state configurations. All network state information is collected locally, but is relayed to the MC. MC converts the incoming states into global operational state and generates appropriate control signals. The coordination between various modules is performed by two parameter sets. The local parameters are visible only to individual instances of the units, and the MC. The global parameters are visible to all units but only the MC can modify it.

3.4 Demonstration Plan

The implemented prototype is called *Self-Organizing Network Embedded Transcoder* (SONET). We have recently made a demonstration of the concept system in a DARPA meeting. In this demonstration we plan to present few adaptive aspects of the system with MPEG-2 video. Following is the tentatively scenario. We will setup a simple asymmetric network. It will contain both high performance subnets as well as well as a wireless subnet. An MPEG-2 server will be placed in the high-speed segment but for serving a client into the wireless network. We will first demonstrate how an incoming live video feed can be dynamically downscaled via this transcoder at a router machine splicing the two segments. Next we will artificially introduce bandwidth impairment in the wireless segment of the

network and demonstrate the dynamic rate adaptation mechanics. In the process we expect to highlight the resilience and superiority of the transcoded video stream (which is empowered with MPEG-2 encoding knowledge infusion) as opposed to simple TCP or UDP streams (which has no knowledge about the content and resists congestion only with network level information). We will then further constrain the outgoing link and demonstrate the perceptual transcoding kicking-in. This should show how extreme adaptation could result from high level network embedded logic (which is empowered with even deeper knowledge infusion about the MPEG-2 stream content), which is otherwise very difficult to achieve. In the next stage we will demonstrate resource adaptation by the self-organization of the transcoder itself. We will also introduce processor overload into the junction machine and show the final feature-- process migration[3] of the system.

Fig-3 (a) rate transcoded output with uniform perceptual encoding using conventional TM-5 quantization.

Fig-3 (b) rate transcoded output with object analysis based perceptual encoding. More bits are given at the moving areas.

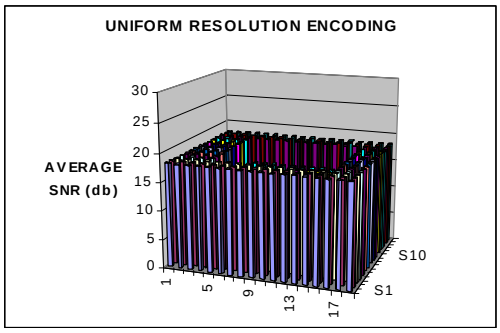

Fig-4 (a) corresponding macroblock wise SNR distribution trace after encoding.

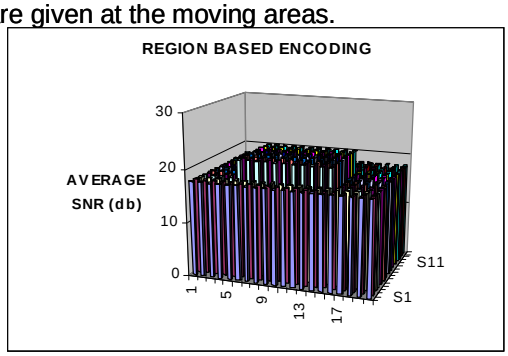

Fig-4 (b) Increased SNR in the area where perceptual significance is higher.

4. Remarks

Adaptation is a complex process. In this research we have focused only on resource adaptive network based systems to function in extreme environments. Success in adaptation comes from cognizance. In this, transcoder we have demonstrated how knowledge about the network, knowledge about the content protocol and the knowledge about the content itself can be infused for pervasive communication. There is little support for cognizant adaptation in today's Internet infrastructure or in the deployed protocols. A visit to any multimedia site can reveal the state of the affair. Today sites pre-encode and store contain multiple copies of the same media one for each popular rate class (LAN, DSL/Cable, 56K, 28.8K etc.) [13,1,3,4]. The idiosyncrasy of the whole arrangement is that we expect the end-user to provide the information about the bandwidth between the server and his/her end-point. Should it not the end-user be the last person to answer such a pure network parameter? Within the current infrastructure there is no easy means to extract this information, so we simply pass on this question to the poor end-user! Perhaps infrastructure for the cognizant adaptation has to be built from core if we expect to build applications, which are truly user friendly and easy to use. The work is currently being funded by the DARPA Research Grant F30602-99-1-0515 under it' s Active Network initiative.

[3] Demonstration of this part needs extra machines. Will depends on the logistics.

5. References:

[1] Amir, Elan, Steven McCanne, and Randy Katz, Receiver-driven Bandwidth Adaptation for Light-weight Sessions, Proceedings of ACM Multimedia ' 97, Seattle, WA, Nov 1997,

[2] Bhattacharjee, S., Kenneth L. Calvert and Ellen W. Zegura. An Architecture for Active Networking. High Performance Networking' 97 ,White Plains, NY, April 1997. [http://www.cc.gatech.edu/projects/canes/papers/anarch.ps.gz, October 98]

[3] Fluckiger, F, Understanding Networked Multimedia Applications and Technology, Prentice Hall, UK, 1995.

[4] Haskell B. G., Atul Puri and Arun Netravali, Digital Video: An Introduction to MPEG-2, Chapman and Hall, NY, 1997.

[5] Information Technology- Generic Coding of Moving Pictures and Associated Audio Information: Video, ISO/IEC International Standard 13818-2, June 1996.

[6] Javed I. Khan, S. S. Yang, Medianet Active Switch Architecture, Technical Report: 2000-01-02, Kent State University, [available at

URL http://medianet.kent. edu/ technicalreports.html, also mirrored at http:// bristi.facnet.mcs.kent.edu/medianet]

[6] Javed I. Khan, Q. Gu, Darsan Patel and Oleg Komogortsev, , Medianet Active Video Transcoder Performance, Technical Report: 2000-01-02, Kent State University, [available at URL http://medianet.kent. edu/ technicalreports.html, also mirrored at http:// bristi.facnet.mcs.kent.edu/medianet]

[7] Keesman, Gertjan; Hellinghuizen, Robert; Hoeksema, Fokke; Heideman, Geert, Transcoding of MPEG bitstreams Signal Processing: Image Communication, Volume: 8, Issue: 6, pp. 481-500, September 1996,

[8] Javed I. Khan, Motion Vector Prediction in Interactive 3D Video Stream, Proceedings of the World Congress on Advanced IT Tools, IFIP '96 IT, Canberra, Sept 96 , pp533-539.

[9] Tennenhouse, D. L., J. Smith, D. Sincoskie, D. Wetherall & G. Minden., "A Survey of Active Network Research", IEEE Communications Magazine, Vol. 35, No. 1, Jan 97, pp 80-86

[10] Wetherall, Guttag, Tennenhouse, "ANTS: A Tool kit for Building and Dynamically Deploying Network Protocols", IEEE OPENARCH'98, San Francisco, April 1998. Available at: http://www.tns.lcs.mit.edu/publications/openarch98.html

[11] Youn, J, M.T. Sun, and J. Xin, "Video Transcoder Architectures for Bit Rate Scaling of H.263 Bit Streams," will beappeared to 'ACM Multimedia 1999', Nov., 1999. pp243-250.

[12] Jill M. Boyce and Robert D. Gaglianello, Packet loss effects on MPEG video sent over the public Internet; Proceedings of the 6th ACM international conference on Multimedia , 1998, Pages 181 - 190

[13] White Paper, Delivering RealAudio® or RealVideo® from a Web Server, RealNetworks Technical Blueprint Series, 1998 [URL:http://service.real.com/help/ library/blueprints/ hosthtml/webhost.htm Last Download: Sept 20, 2000]

[14] Javed I. Khan, & S. S. Yang, Made-To-Order Custom Channels for Netcentric Applications over Active Network, Proc. Of the Internationaal Conf. On Internet and Multimedia Systems and Applications, IMSA 2000, Nov 2000, Las Vegas, pp22-26.

[15] Katie Hafner and M. Iyon, Where Wizards Stay up Late: the Origins of the Internet, Simon and Schuster, New York, 1996.

[16] IOverview of the MPEG-4 Standard, ISO/IEC JTC1/SC29/WG11 N3444, Geneva – May/June 2000, [Also available from http://www.cselt.it/mpeg/standards/mpeg-4/mpeg-4.htm]

[17] Javed I. Khan, Darsan Patel, Wansik Oh, Seung-su Yang, Oleg Komogortsev, and Qiong Gu, Architectural Overview of Motion Vector Reuse Mechanism in MPEG-2 Transcoding, Technical Report TR2001-01-01, Kent State University, [available at URL http://medianet.kent. edu/ technicalreports.html, also mirrored at http:// bristi.facnet.mcs.kent.edu/medianet] January, 2001]

[18] Jill M. Boyce and Robert D. Gaglianello, Packet loss effects on MPEG video sent over the public Internet; Proceedings of the 6th ACM international conference on Multimedia , 1998, Pages 181 - 190

[19] Gary Stix, The Triumph of the Light, Scientific American, January 2001, pp31-35 [HTTP:// http://www.sciam.com/2001/0101issue/0101stix.html]

[20] [Peter Forman and Robert W. Saint, Creating Convergence, Scientific American, November 2000, pp.10-15[HTTP://http://www.sciam.com/2001/0101issue/0101stix.html]