Efficient Multimedia Data Transmission Using Active Networks

Akbar[1], Buhari[1], Sahalu[2], Saleem[1]
[1]Lecturer, ICS Dept., King Fahd University of Petroleum and Minerals.
[2]Asst. Professor, ICS Dept., King Fahd University of Petroleum and Minerals,
31261 Dhahran, Saudi Arabia.
Phone: 966 3 860 4374
Fax: 966 3 860 2174
Email: mibuhari@ccse.kfupm.edu.sa

Published in CISST' 2001

Abstract

The transmission of continuous data like multimedia data requires guarantees of the restriction of the maximal end-to-end delay and minimal jitter. Since the most widely used network, Ethernet, cannot guarantee these, alternative solutions are developed to enable multimedia data transmission on Ethernet. These include using a dedicated Ethernet, bounding the maximum utilization of the network or using a start topology with a dedicated Ethernet at each station.

In this paper we propose the use active networks to support efficient multimedia data transmission in a distant learning environment where lecture slides need to be frequently accessed remotely by students. In active networks the routers take active part in processing the data being transmitted rather than acting as mere packets forwarders as in the traditional non-active networks. In particular, the routers in active network can cache frequently requested data from a server for faster onward transmission to clients. OPNET-based simulation of our active network model exhibits about 30% performance improvement over the same network model with passive routers.

Introduction

The current day of information technology is very fast and it deals with various fields of interest. Updating all the areas of information technology by a single person has become an uphill task. To handle this issue, we need lot of experts in this area to have the different fields of interest like Computer networks, computer graphics, digital image processing and so on. The development of the internet is faster and there are lot of information on the Internet. But, a person has to focus and try to improve his knowledge in a peculiar field of interest.

At the same time, to teach the bachelor degree courses, we need to teach all the courses in an introductory or general way. So, we actually people who are in various fields of information technology. The number of industries in this field keeps on increasing and so people are attracted towards the industry's too. Lack of experts is the major problem, especially, for the countries in the east, which are not able to pay the same as the west in terms of salaries.

To overcome this problem, we have come up with an idea of co-operation among various universities in one single region to get together to organize their sessions and be handle things towards for two campus at a specific time by a single lecturer. The process of using the Internet may cause the data to be not synchronized and so the video and audio might not match together. To avoid this problem, we propose an active networks system which handles this.

1. Active Networks – an introduction

Active networking refers to the placement of user-controllable computing and other resources in the communication network, where they can be utilized by applications that need those capabilities. An active network supports a user-network interface allowing the nodes of the network to be programmed by the application (user) to provide a desired functionality, such as routing. This programming might be done on a per-packet basis (as in the capsule approach of Tennenhouse and Wetherall) or through an out-of-band signaling mechanism (e.g., injection of user-specific programming into the switch, as

in Switchware). Active networks allow users to increase the likelihood that the service offered by the network will be useful to them, by providing an interface that supports multiple (or programmable) services. There are costs associated with such a flexible interface, and they affect all of the network's users whether they have advantage of active network support or not. The (monetary) cost of providing the interface, though likely to be significant, is paid once and can be amortized over all users for a period of time. The performance cost of using the interface should vary with application; this is the end-to-end argument [Sam97]. Applications running on end systems are allowed to inject code into the network to change the network's behavior in their favor [Dan99].
The remainder of this paper is organized as follows. Section 2 outlines the general architecture of our active network system. In Section 3 we give an overview of research work related to the work presented in this paper. Section 4 gives a brief summary of the requirements for a multimedia network system. In Sections 5 and 6 we present our proposed model and our network model respectively. We present our experimental results in Section 7 and conclude the paper in Section 8.

2. Need for active networks

Without architectural support, current network environments impede the revolution of network communication technologies and prevent extension of the Internet services. Several crucial problems within today's Internet infrastructure can be identified. For example, difficulty of integrating new technologies and standards into the shared network infrastructure, difficulty in accommodating new services with the existing architectural model, poor performance due to redundant operations at several protocol layers, etc [Chi00].

In principle, there are three advantages to basing the network architecture on the exchange of active capsules, rather than passive packets [Dav96, Dal96]:

- Exchanging code provides a basis for adaptive protocols, enabling richer interactions than the exchange of fixed data formats;
- Capsules provide a means of implementing application-specific functions at strategic nodes;
- The programming abstraction provides a mechanism for infrastructure innovation, allowing new services to be deployed at a faster pace than can be sustained by standardization.

Thus the process of sending the data file along with the capsule so as to make the capsule do the updating even at the level of the routers, make the whole job easier and faster. In the case of distribution of data among various universities, which are closer by, the following aspects have to be taken into consideration:

- As they are closer by the number of routers the packet (capsule) actually travels will be less.
- The routers might be within a specific region and so the controlling people of the routers might be known. The important point to note here is, when we use the courses that are handled among Brunei, Malaysia and Singapore. All the three countries are very close and so sharing of information will be very easy. As the network length is not that long, the effect of external traffic will be comparatively less.

3. Related Works

Multimedia objects such as audio, video and image are usually very large in size. For example, one MPEG-1 movie in NTSC video quality (which we digitized from a VHS video tape) has an average size of 8617 bytes/frame. The average I-frame size is even larger (17997 bytes/frame). When transporting multimedia data in the networks, the sizes of the payload are usually larger than the Maximum Transmission Unit (MTU) of the underling physical networks. For example, the MTU of Ethernet, the most popular link-layer technology, is 1500 bytes. Therefore it is inevitable that the multimedia data have to be fragmented into smaller-size units when transmitting through physical networks [She98]. For MPEG, we can define two units: an MPEG frame (I, P or B), and an MPEG GOP. MPEG frames are an example of a unit without dependencies, since loss of any part of a frame may render the entire frame useless, while MPEG GOPs are an exmple of a unit with dependencies.

We assume that loss of the I-frame in a GOP renders the P and B-frames in the GOP useless. Note that this is a slight simplification; some P and B-frames may be coded independent of the I-frames [sma96]. When a packet that carries a part of an I-frame is lost or corrupted, P-frames that reference that particular I-frame are decoded erroneously. The same applies to P-frames that are referenced by other P-frames. B-frames are usually not used for real-time transmission, because they increase the end-to-end delay. This is due to the required reordering of video frames for transmission and the buffering of the next I-frame before the B-frame can be encoded. However, they also have dependencies to the previously transmitted P or I-frames and thus similarly suffer from the previously transmitted frames is decoded erroneously when previously transmitted frames are delivered incomplete or with bit errors.

Multimedia communication generally requires high bandwidth, low packet delay, low packet delay jitter and small packet loss rates. In particular mobile multimedia communication demands flexible resource management and highly adaptive data streams that are able to adapt to different packet error rates, sudden network congestion due to terminal mobility and a high packet error rate. Active network technology has shown to be beneficial for applications such as network management, bandwidth adaptation or dynamic protocol deployment. Using dedicated application-specific procedures within a set of active network nodes for application level error recovery also provides promising solutions [Jen00].

Two main research directions in the area of application level feedback for improving video transmission can be pointed out. On the one hand, rate control mechanisms have been developed that attempt to minimize network congestion by matching a certain bandwidth. On the other hand, error control mechanisms have been conceived which attempt to minimize the visual impact of loss of destinations. [Jen00]

4. Performance improvement of multimedia information:

Mainly four error recovery schemes have been proposed and partially implemented in the well known traditional and experimental video conferencing systems are [Jen00]:

1. Full Intra Frame Refresh (FIR): When a receiver experiences substantial packet losses, it sends a so-called “full intra frame refresh” packet to the source. The source will then transmit the next video frame as a complete intra coded frame. The main disadvantage of the FIR scheme is the waste of bandwidth in the forward direction. The probability of loss when receiving the full video frame is much higher than for intermediate frames, because I-frames are much larger. This leads to series of recurring requests of full-intra frames.
2. Negative Acknowledgement (NACK): After detecting subsequent packet losses at the receiver, the receiver may send a negative acknowledgement directly to the sender. The sender will retransmit the packets either directly to the receiver or if needed over multicast to a set of receivers with similar negative acknowledgements. This mechanism forces the sender to buffer transmitted packets for a certain period of time.
3. Forward Error Correction (FEC): Forward error correction is a method, which has recently been implemented in the IVS system. Additional blocks already transmitted are added to the video frame i. The scheme does not allow for error recovery in case of packet loss.
4. Timestamp Control (TT): Similar to the FIR-scheme, the TT-scheme sends a special control packet back from the receiver to the sender application in case packet loss has been detected. Instead of a single refresh request, the control packet hold a table where for each macro block on the receivers display, the timestamp of the associated packet is recorded. The content of the control packet allows the sending application to only refresh those parts of a video in intra mode that is actually damaged. The sender compares its own timestamp table with those received from the receiving applications and re-encodes only that part of the video that has been lost. The TT-scheme has the property that the detailed error recovery procedure has to be tailored to encoder parameters. For example, the range of motion vectors and the usage of P- and B-frames has to be considered. Because a control packet has to carry one timestamp for each macro block, the size of

a control packet is relatively large compared to other schemes. For example, a video in QCIF mode transmitted using the RTP protocol, results in control packets with 396 bytes. To tone down this problem, the granularity of the error recovery scheme can be modified, e.g., by sending only accumulated measures on a Group of Block base.

5. Our proposed model

Between the two available methods in active networks available [as described in section 2], we prefer to select the capsule approach because of the following reasons:

- The method by which we specify the pointer to the data will tell where the data is present. But in our case, the actual data has to be transferred to the universities which share the course, from the original university where the actual class is conducted.
- We need some processing on the capsule and being able to take the data to be traveling along the way towards the destination.
- We can actually specify the required action on the packet. In our case, mostly we do synchronization on the packets.

As discussed in She98, the network traffic is highly utilized when we send multimedia oriented data. The process of MPEG data can be handled by throwing away the unwanted information when an I-frame is lost. Basically, the impact of information on the network is handled by the following means:

- The lecture slides are first transferred to the destination servers, whereby the students can access these materials locally. The information transferred is only the action on the class duration. This is done because, the lecture hand-outs are one of the very important material and if any one slide is lost, the total continuity will be lost.
- As a subject is handled by one single lecturer, we use a speech recognition system, which helps us to convert the total information of his audio into the bits format to be transmitted faster. This one, again will have some delay due to the reverse processing at the destination end. This time delay is reduced by introducing the concept of active networks where the routers on the way itself actually act on the data. Again, the routers posses some data on the cache which are used to be send to the others routers, in case there are many universities sharing the same information.
- The proper mixture of NACK, TT and IP-SQ algorithm will solve issue of error recovery during congestions.

6. Our Network Model

Our network model is made of 4 nodes named from 0 to 3, which are connected to each other through a hub as shown in Figure 1. The host computers handle the packet that carries destination address of the packet. The packet generation is handled by the hosts, which posses the ideal packet generators [***src*** in Figure 2] in them. The generated packets are then processed and transmitted through the transmitter. The hub is made up of a processing node connected to many transmitters and receivers, depending on the number of hosts that ought to be connected to it. The node part of the host and the hub are shown in Figure 2.

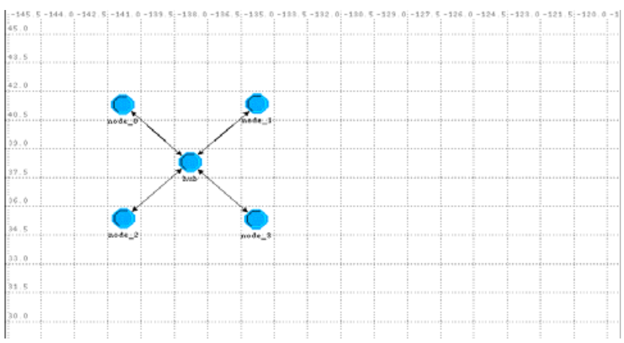

Figure 1. Network diagram of the system

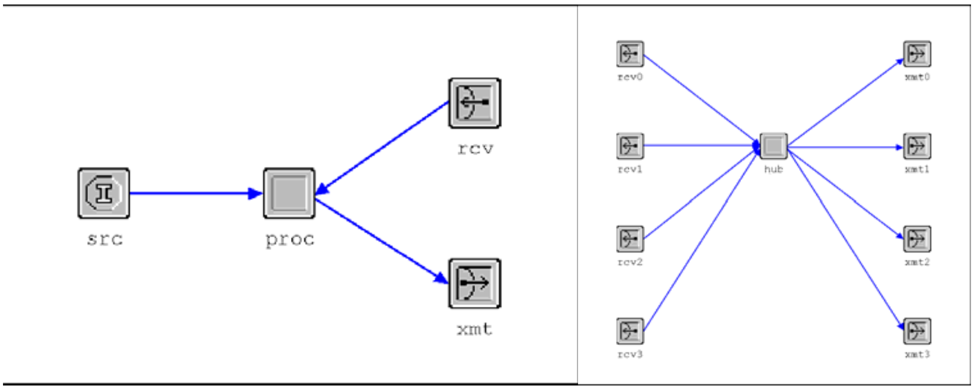

Figure 2. Node diagram of the host and the hub

- **Processing part of host and hub**

The processing portion of the host and the hub are shown in Figure 3. The host's "init" part is used to make the packet destination address generation to be between 0 and 3 because the network has 4 nodes only. The generated packets are then forcibly sent to the idle part. The transmitter then transmits the packet along any one stream depending upon the destination address specified.
At the hub, the checking for the destination address is done and packet is forwarded in the respective transmitter stream. Once the packet reaches the destination, the packet is now processed to identify the time it has taken to reach the destination from the source. Depending upon the time duration taken the SQ message is generated to the source.

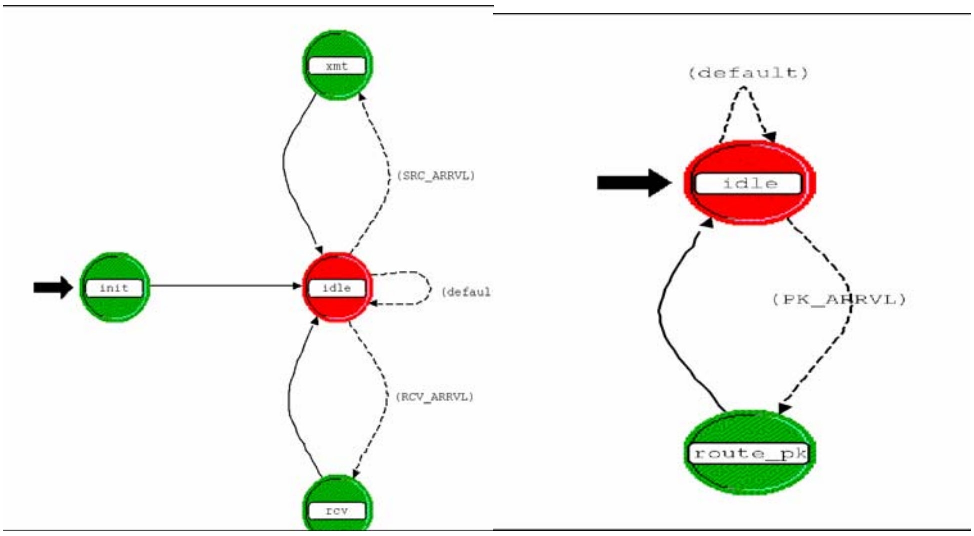

Figure 3. Process module of the host and the hub

To identify the state of congestion, we calculate the time taken for the packet to move from the source to the destination. If it takes long time, that means that the network is congested and there is more likelihood of collision and loss of packets. To handle this situation, we need to check for the time that the packet takes to transfer from source to the destination. Based on the identification of the timing, the source needs to be informed if the destination identifies that the network is congested. After receiving the packet at the destination, the first task done is the calculation of the time taken for the packet to reach the destination from the source. The source time is present in the packet header. This time is subtracted from the current time at arrival to obtain the time taken. The source quench message is generated if the packets are generated at a faster rate by the source. In this case, if the packets are received faster than the receiver can cope with, the receiver will send a SQ message to the source and so the source has to reduce the speed of transmitting packets.

Initial analysis was done by changing the time at which SQ message has to be generated along with varying the bits per second. Later, the processing was done at two levels and compared in terms of the recorded performances. In order to see that the situation of applying the active networks concepts of processing the data on the fly helps the IP/SQ algorithm, we have done the processing at the host level and also at the hub level. The processing time was analyzed.

7. Analysis of Experimental Results

The system was tested with varying inter-arrival time [the time gap between the transmission of two consecutive packets] between packets. There was an increase in the End-to-End (ETE) delay when packets where transmitted with short inter-arrival time. This caused an increase in the network utilization when we transmit packets faster as shown in the Figure 4.

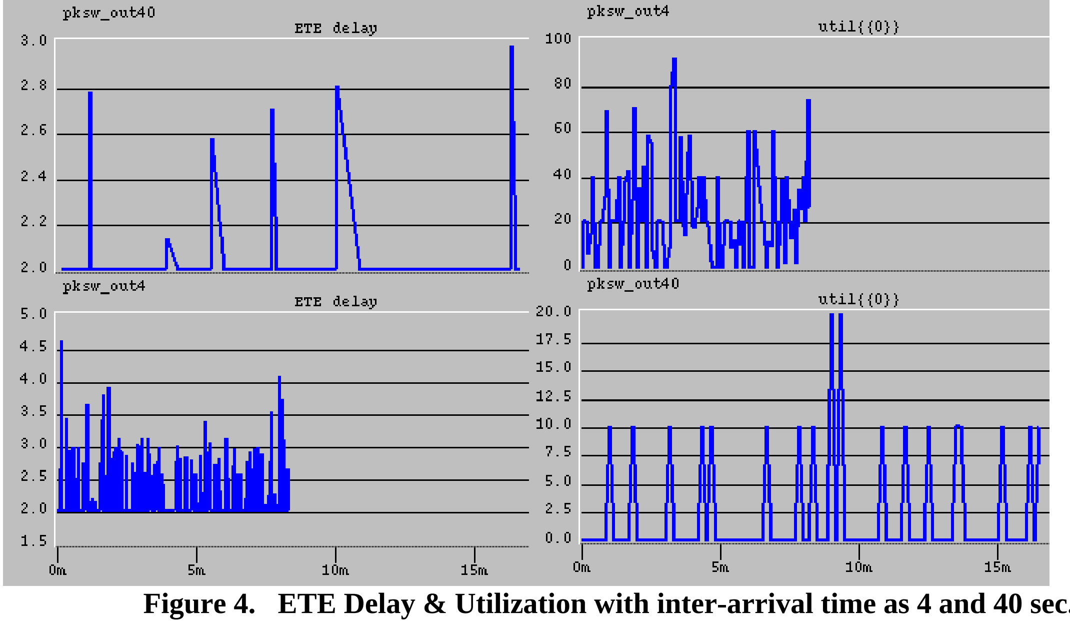

Figure 4. ETE Delay & Utilization with inter-arrival time as 4 and 40 sec.

The above graph [Figure 4] helps us to infer that if the packets are sent with small inter-arrival, there is no more congestion and the network utilization is high. The process of identification of the situation of congestion in the network was done at the destination end as well as at the hub and compared. The impact of the network congestion found at the end hosts has an equivalent impact on the hub too as shown in Figure 5. This makes it clear that we can make an action at the hub on the congestion control that will reduce the impact of congestion throughout the network. The analysis was done when packets where generated at various inter-packet arrival time differences.

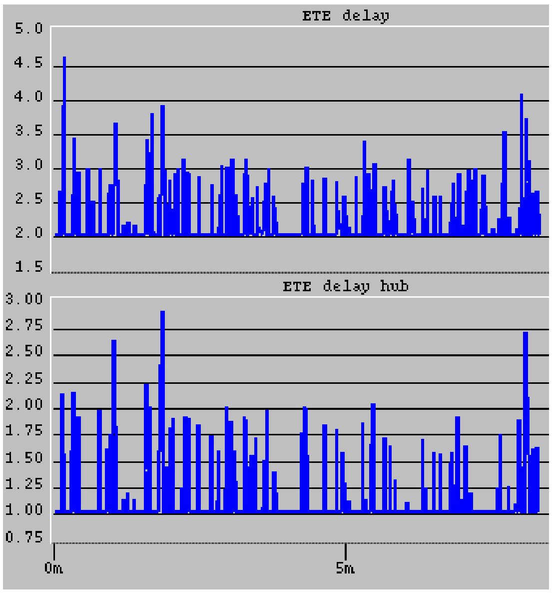

Figure 5. ETE at Hub and Node

ETE delay

After identifying that the presence of some actions at the hub makes an impact on the network, we added the code to send the SQ message from the hub to source in case there occurs a very low ETE delay. Because of the presence of SQ action at the hub itself, we can see from figure 6 that the ETE delay towards the node end is reduced.

The end to end delay varies between the non-active and active system when inter-arrival time is 4 sec

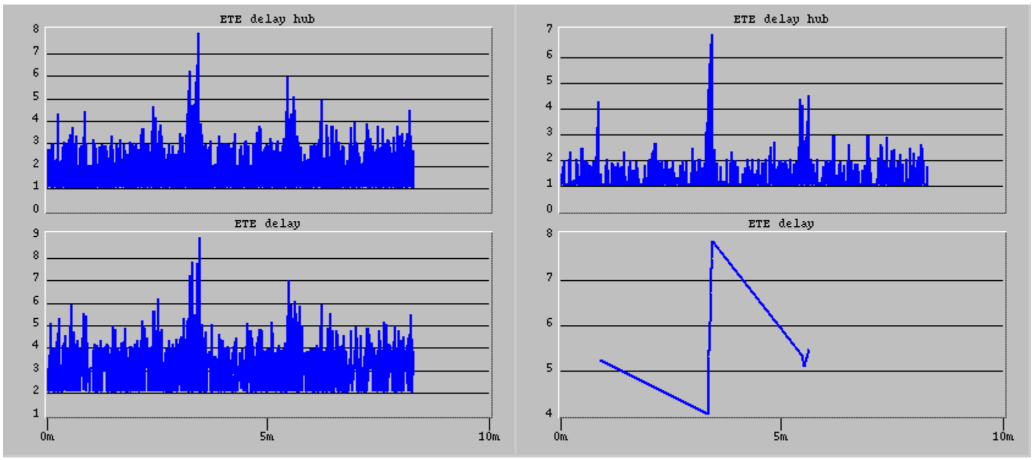

Figure 6. ETE Delay - with inter-arrival time as 4 sec – for non-active and active n/w

The end to end delay and utilization is same for both non-active and active system when inter-arrival time is 40 sec.

The utilization was same with in a subnet, when the system was based as active as well as non-active. The packet transmitted from one subnet to another has a difference in utilization because processing is done at various active nodes.

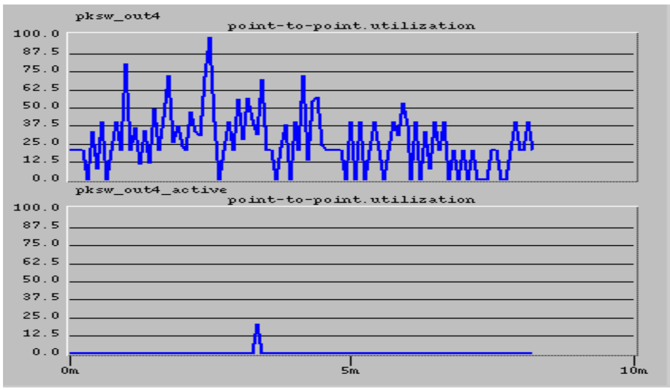

Figure 7. Utilization - with inter-arrival time as 4 sec – for non-active and active n/w – between subnets

A highly congested network may end up in collision and loss of packets. To avoid this situation, we handled the congestion scenario at the router level itself. Therefore, it is clear that the performance of the system improves with the inclusion of Source Quench at the router level.

8. Conclusion

Because of the advancement of computer and digital signal processing technologies, multimedia communication services including image and video signals are expanding very rapidly. Since multimedia data contains a large amount of information, we need to provide techniques for the efficient procession and transmission of this data.
We have presented a novel way of using active networks technology to dynamically adapt the continuous multimedia data transmission rate of the source to the network load thereby minimizing the rate at which congestion occur within the network. This leads to improve performance of the network. Our simulated results using OPNET indicate performance improvement of up to 30% over traditional network setup.

9. References:

1. [Sma96] Samrat Bhattacharjee, “An Architecture for Active Networking”, Technical Report GIT-CC-96/20, College of Computing, Georgia Tech.
2. [Dan99] Dan S. Decasper, Bernhard Plattner, Guru M. Parulkar, Sumi Choi, John D. DeHart, and Tilman Wolf, “A Scalable High-Performance Active Network Node”, IEEE Network, May/June 1999, pp. 8-19.
3. [Chi00] Chih-Lin Hu, and Wen-Shyen E. Chen, “A Mobile Agent-Based Active Network Architecture”, Proceedings of the Seventh International Conference on Parallel and Distributed Systems (ICPADS ’00), Japan, July 2000, pp. 445-452.
4. [Jen00] Jens Meggers, and Michael Wallbaum, “Application Level Error Recovery Using Active Network Nodes”, The Fifth IEEE Symposium on Computers and Communications (ISCC ‘2000), July 2000, pp. 540 – 545.
5. [She98] Sheng-Yih Wang and Bharat Bhargava, “A Fragmentation Scheme for Multimedia Traffic in Active Networks”, The Seventeenth IEEE Symposium on Reliable Distributed Systems, Oct 1998, pp. 437-442.
6. [Dav96] David J. Wetherall and David L. Tennenhouse, “The Active IP option”, 7th ACM SIGOPS European Workshop, Sept. 1996.
1. [Dal96] David L. Tannenhouse and David J. Wetherall, “Towards an Active Network Architecture”, Computer Communication Review, Vol. 26, No. 2, April 1996.