Design of a Label Switch Controller for Differentiated Services in IP and ATM Integrated Networks

JAESUP LEE*+, KEUN HO RYU+, KYOU-HO LEE*, TAE-IL Kim
*Electronics and Telecommunications Research Institute
Yusung P.O. BOX 106, Daeduk Science Park, Taejon, 305-606, Korea
Email {jaesup,kyou}@etri.re.kr
+Dept. of Computer Science, Chungbuk National Univ., Chongju, Korea, khryu@dblab.chungbuk.ac.kr

Abstract: - MPLS is highlightened as the most promising technology for the integrated IP-over ATM backbone networks. Nowadays, one of the important practical issues in MPLS is the capability to provide Differentiated Services. For integrating IP and ATM with scalability, MPLS based ATM switch network can be available, especially for the fast Internet services adding Layer 3 routing module to the existing ATM network, and can provide scaleable Internet services to users with various service level. In this paper, we establish Queuing model for FE(Forwarding Engine) including service differentiation, which performs IP address lookup in MPLS LER, and analyze performance of FE with Internet traffic by statistical Internet protocol analysis. Also we take a closer look into our LC-ATM switch on which we implement MPLS controller that can provide both user and provider with relative proportional differentiated service in a feasible and reliable manner.

Key-Words: - Internet, MPLS, Switching System, Performance, Forwarding Engine, Service Differentiation

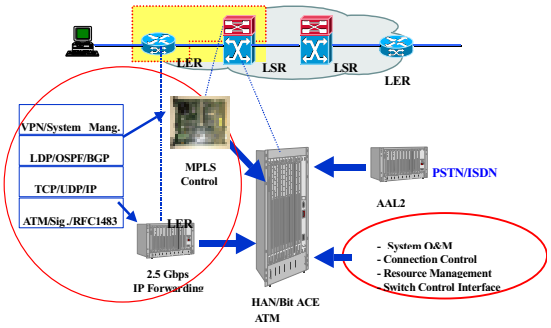
1 Introduction

As existing Internet supports best-effort service, it cannot satisfy user's various requirements. ATM(Asynchronous Transfer Mode) provides several service qualities, but it does not have many services. MPLS(Multiprotocol Label Switching)[1] is one of the methods for providing Internet service based on ATM switch. MPLS networks consists of MPLS control components and LSRs(Label Switched Router), which are composed of ATM switch. LSR performs LDP(Label Distribution Protocol)[2] which negotiates labels based on FEC for transferring IP(Internet Protocol) packet with neighbor LSRs.

In MPLS network based on ATM switch [4][5][6], VPI/VCI of ATM is used as a label, so MPLS network is configured by adding L3 routing module to existing ATM network. Internet ser-vice based on ATM switch is available in the network. Packets are simply transferred with high speed compared with the existing IP network, be-cause, at the end of the MPLS network, L3 lookup is performed and labels according to FEC are as-signed. In the core of the network, L2 switching according to VPI/VCI is performed. We use LDP as signaling protocol in MPLS network. LDP distributes labels, which are used to label switching, to MPLS node, and it associates labels with routing information of network layer in order to establish LSP(Label Switched Path). Recently, IETF and researchers have focused on Differentiated Services(DS) architecture[9]. By extending LDP message, we can keep service differentiation in any service ratios required between classes. In this paper, we study performance analysis of MPLS system architecture based on ATM switch. The performance of our MPLS LSR system has been satisfied with the system requirements we needed by both adding a specific processor for processing IP packet to existing ATM system and mainly using cut through mechanism. But MPLS LER system must perform the label lookup function according to IP address in order to setup LSP. We establish FE Queuing model for IP address lookup in MPLS LER system, and carry out performance analysis based on it. This paper begins with architecture of MPLS LER system and Forwarding Engine block in the second section. The third section describes the Queuing Model of FE plus Differentiated Services, especially for the algorithm for the relative proportional delay differentiation implementation, and the fourth section provides a description of the parameters for performance analysis of MPLS over ATM switch. In the fifth section, we present simulation results and the last section contains the concluding remarks.

2 Architecture of MPLS LER System

Fig. 1 shows the MPLS based ATM networks, which are composed of LERs and LSRs respectively in MPLS domain. For an implementation of MPLS features in our ATM switching system as shown in this figure, we newly developed MPLS control parts and IP forwarding module parts. The interfacing function between MPLS control part and ATM switch is performed using GSMP protocol . We also develop the internet access via PSTN/ ISDN[3]. The overall architecture of our MPLS LER system, which consists



of ATM-LSR and routing module, is depicted in Fig. 2. IPCP(Internet Protocol Control Processor) board for

Fig. 1 MPLS/ATM Networks

supporting IP service is in ATM-LSR, and it has a table for mapping destination IP address to a label in order to establish LSP in MPLS network.

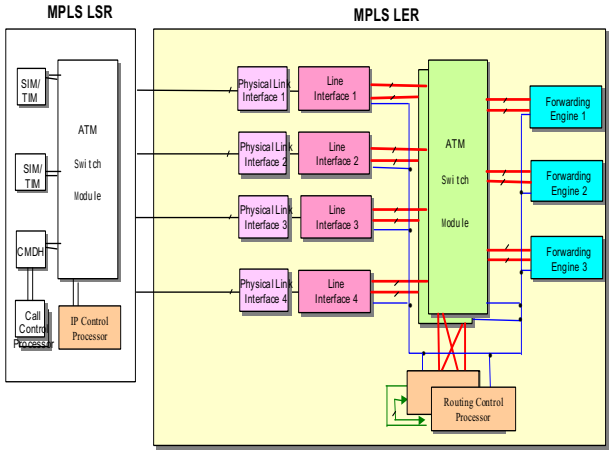


Fig. 2 Architecture of MPLS ATM LER System

Routing module is located in both ends of the MPLS network, and it performs L3 lookup in order to assign a label to the packet with destination IP ad-dress. In this paper, we analyze packet-processing volume from the result of performance analysis of our MPLS ATM system. The important factors of performance in MPLS ATM system are mainly IPCP of LSR, RCP and FE of routing module. However, in case of IPCP and RCP, routing protocol and LDP carried out well in accordance with the specifications as we expected and they don't have any specific bottlenecks. As FE is imposed as the important factor in performance, we analyze processing volume for FE, and estimate performance of MPLS ATM LER system.

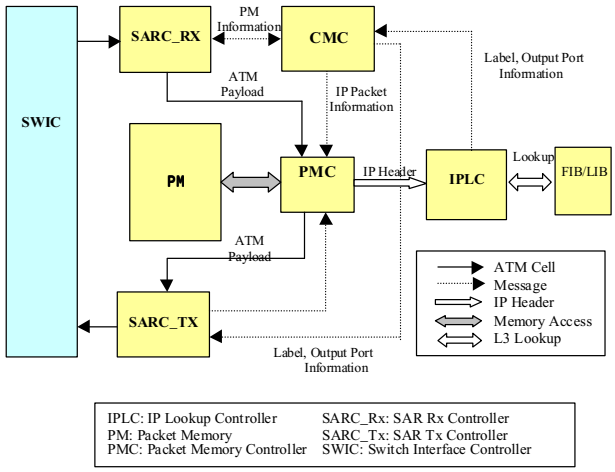


Fig. 3 Composition of IPFA block

Fig. 3 illustrates components for L3 lookup of IP packets received by Non-MPLS network after LDP sends Forwarding Table to IPFA(IP Forwarding Assembly). First, a packet received by Non-MPLS network is transferred to IPFA through IM(Interface Module) board. The packet in IPFA is inputted to SAR without Routing Tag, and then it is stored into Packet Memory in the re-assembled format. At this time, we know the start address and length of the packet. In Port Controller, IPH(IP Header) is written from Packet Memory, LIB(Label Information Base) table pointer is found using destination IP address and netmask, then channel identification, TTL(Time To Live), and CoS(Class of Service) is known by the pointer. Using the TTL/COS, that of IPH stored in Packet Memory is changed, and VPI/VCI, which is swapped, is found using Channel identification. Routing Tag is known by VPI/VCI, and then Payload divided via SAR is assigned to the VPI/VCI, and is configured to 53-byte cell. Routing Tag is added to the Payload and is transferred to MLIA(MPLS Line Interface Assembly). Routing Tag is removed from ATM cell of 64 bytes received at MLIA, and the ATM cell is transferred to MPLS LSR system.

3 Queuing Model of Forwarding Engine

3.1 Queuing Model

Fig. 4 shows Queuing model based on the components of IPFA, which consists of processors of IPFA and

memories for storing packet in order to process one packet per each server. Messages are transferred in order to request processing among processors. Cells are inputted in the receiving part of switch interface with 64-byte length. The time for processing one cell is from 1 cell time to 10 cell times. In case of TX part different from RX/TX part, cells are received from SAR and are checked Port number according to Channel id, so it takes time longer than RX part. SAR-RX/TX receives SWIC output of 53 bytes, and operates each 1-cell time. Packets outputted in SAR-RX part are stored in Packet memory, and in case that the last cell is inputted, IP header is written from packet memory and IP header lookup is performed in the IPLC(IP Lookup Control) block. Therefore, the velocity of read/write of packet memory is the important factor of performance parameters.

For simplicity, we assume following assumptions for the analysis of FE.

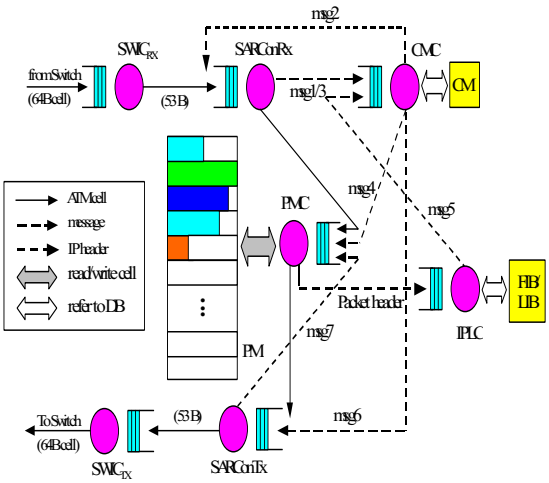


Fig.4 IPFA Queuing Model

- Every IP packet entered to FE is ATM cell type only;
- Traffic patterns and characteristics of IP packet referred to the information surveyed by MCI[8];
- Exponential distribution is commonly used for both the interarrival time for every IP packet cell and the interarrival time between cells of the same IP packet and arrival pattern is assumed as Poisson process;
- Best effort service plus Differentiated Services (Proportional Delay Differentiation) ;
- Other messages in the system except IP processing purpose are not considered.

3. 2 Problem with Diffserv and Solution

For Diffserv model, we implemented three classes on each FE of our MPLS switching system. Each class has one separated queue. A Diffserv domain is defined in[16] as a contiguous set of DS nodes that operate with a common service provisioning policy and set of PHB groups implemented on each node, and marks each packet of each flow with DSCP(Differentiated Services Code Point) code point. All packets marked with the same DSCP are collectively called a Behavior Aggregate(BA). Here, we have to point out some of problems with Diffserv. End-to-end QoS metrics in Diffserv domain are not the same with the summation result of per-Hop metrics. Designing end-to-end services with weighted guarantees at individual hops is also difficult. SLA is designed for static agreements even though both the network topology and traffic are highly dynamic. Dovrolis, and et. al.,[10] proposed proportional Differentiated Services for the relative service differentiation of the Internet. One of their packet schedulers, WTP (Waiting-Time Priority) scheduler is applied to our packet scheduling mechanism, but we extend its delay mechanism and add our modified class adaptation algorithm, which is similar to the one suggested in [11]. In accordance to WTP, the priority of packet increases proportionally with its waiting time as shown in equation (1). The priority of a packet in queue i at time t, $P_i(t)$ is

$$P_i(t) = w_i(t)s_i . \qquad (1)$$

where $w_i(t)$ is the waiting time of the packet at time t. The control parameters $\{ s_i \mid s_i < s_{i+1}, i(1..N-1) \}$ determine the rate with which the priority of packet of a certain class increase with time. We define the following assumptions in order to select packets class and transmit them through the link of LC-ATM.

Definition3.1 Let $D_i(t)$ be the aggregate delay experienced by all packets that have been served and are currently in the queue for next transmission at time t, and $q_i(t)$ is the number of packets queued in i at time t, then

$$D_i(t+\theta) = D_i(t) + q_i(t) \qquad (2)$$

where θ is a time slot unit for updating scheduler variables.

Definition3.2 Let $T_i(t)$ be the number of the packets served from delay class i till time t, and the link capacity is m packets per unit time slot , then

$$T_i(t+\theta) = T_i(t) + I_i(t) \qquad (3)$$

where $I_i(t) = m$ if the corresponding queue class is selected for transmission at time t, or otherwise 0.

Definition3.3 Let $a_i(t)$ be the number of arrivals of i class packets at time t, then

$$q_i(t+\theta) = q_i(t) + a_i(t+\theta) - I_i(t) \qquad (4)$$

Lemma3 Let θ be $mm'(1 \geq m' > 0)$ and all the packets in the queue *i* be transmitted back-to-back starting time at *t* and no other packet arrive into the same queue *i* since , then the minimum average delay $\bar{d}_i(t)$ is

$$\bar{d}_i(t) = \frac{(D_i(t)+(\lceil m'q_i(t)\rceil)(\lceil m'q_i(t)\rceil+1)/2))}{(T_i(t)+q_i(t))} \qquad (5)$$

where θ is a time slot unit for updating scheduler variables.

Proof: If we choose θ=1, then m'=1/*m*. Since the lower bound on the cumulative delay by all packets in the queue, served from time *t* is $q_i(t)(q_i(t)+m)/2m^2$, therefore $\bar{d}_i(t)$ is the the minimum average delay. ■

By using the result discussed above, delay mechanism for Diffserv implementation as shown below can be applied for the service differentiation in our MPLS/ATM network.

Delay mechanism :

consider_arrival-pattern

$q_i(t) = q_i(t) + a_i(t)$;

check-non-empty queue(work-conserving)

for every *i*(1 to N)

$j = \{i \mid q_i(t) > 0\}$;------ (j-1)

select_packets( )

for every *j* (from above (j-1))

calculate( )

$\bar{d}_i(t) = \frac{(D_i(t)+(\lceil m'q_i(t)\rceil)(\lceil m'q_i(t)\rceil+1)/2))}{(T_i(t)+q_i(t))}$;

$k = \arg\max_j \{D_j(t)s_j(t)\}$;

if more than two j(j1,j2) then select

$\{ j2 \mid D_{j2}(t) > D_{j1}(t) \}$;

k= j2;

transmit packets from delay class k;

if $q_k(t) > m$ then

$T_k(t) = T_k(t) + m$;

$q_k(t) = q_k(t) - m$;

else $q_k(t) = m(\mathrm{mod}\, n)$;

$T_k(t) = T_k(t) + n$;

$q_k(t) = q_k(t) - n$;

For every delay class *i*

$D_i(t) = D_i(t) + q_i(t)$

We experiment this algorithm to keep proportional delay ratios between differentiated service classes. The result is dealt with altogether our routing module architecture simulation in the following chapter. In addition, based on algorithm[18], we propose an enhanced and more reliable class adaptation algorithm as shown below.

Class adaptation :

f->estimated_delay = p->estimated_delay; /* p: marker packet, *f*: flow table pointer; measured data->flow base */

$\bar{d}_i$ = *f*->estimated_delay;

i = *f*->delay_class;

if($\bar{d}_i > \bar{RD}_f$) /* $\bar{RD}_f$: average end-to-end delay requirement for flow *f* */

if(*i* < f->max_delay_class)

then *f*->delay_class++; /* shift to higher delay class */

else /* shift to lower delay class*/

$\bar{d}_{i-1} = \bar{d}_i \ \bar{s}_i / \bar{s}_{i-1}$;

if(($\bar{d}_{i-1} \leq \bar{RD}_f$)

&&(($\bar{RD}_f - \bar{d}_{i-1}$)($\bar{RD}_f - \bar{d}_i$)/($\bar{d}_i \ \bar{d}_{i-1}$) $\geq k$)

/* *k* : an arbitrary number between(0..1) generated by uniform(0,1) distribution */

then *f*->delay_class--;

4 Parameters of Performance Analysis

Table 1 and Table 2 illustrate service properties [15] of Internet protocol and size of packet. We know that occupancy ratio of TCP protocol is the largest, so the traffics of UPD, ICMP, Ipv6 and so on can be disregarded. Based on table 1 and table 2, web service, one of the TCP application services, occupies the largest part of Internet service, and average packet length is more than 200 bytes. Based on Table 3, we find that MPLS system based on ATM switch has packets composed of 4 cells.

Class	Occupancy ratio			
	Byte	Packet	Flow	Average Packet Size
TCP	Over 95	85-95	75-85	300 bytes
UDP	Below 5	5-15	15-25	200-500 bytes
ICMP	Disregard			

Table 1. Occupancy ratio of IP protocol

Class	Properties of flow			
	Number of packet	Length of packet	Time of continuance	Average packet size
TCP	16-20	5-8k	12-19 sec.	300 bytes
UDP	5-15	1-2k	10-18 sec.	200-500 bytes
ICMP	Disregard			

Table 2. Properties of IP protocol flow

Length(Bytes)	Occupancy(%)	Accumulative sum
40-44	50	50
45-551	20	70
552	5	75
553-575	1	76
576	10	86
577-1499	4	90
1500	9	99
1501-4500	1	100

Table 3. Packet Length and Occupancy

We performed a simulation for the analysis of FE based on ATM switch using the above facts and generated the input traffics according to the information obtained from Table 3. We cannot exactly measure processing velocity of hardware signal arrived at each processor [12], if the worst case is considered, CMC is 800 ns ~ 200 ns, for PMC, IP Header Read time is 32 TS, and Read-Write Time is almost from 64 TS to 80 TS. Lookup time at IPLC is between 10 and 20 TS.

To evaluate the performance of FE, we ran several numerical simulations. Here we consider the case shown in Table 4. We use a simulation package called AweSIM[7].

Processing time Type	Traffic Processing Duration
ATM cell processing	1 cell time slot(0.68 micro seconds in 622 Mbps)
Message processing	800 ns(CMC); 400 ns(except CMC)
PM memory Access	240 ns/ATM cell
PMC IP header Read	640 ns/IP packet
IPLC Lookup time	Variable to be obtained

Table 4. Traffic Processing Time in FE(Simulation conditions)

5 Results

Input parameters of performance analysis in MPLS ATM LER system are cells of each controller, signal processing time, and L3 lookup time. We analyze average waiting time of each processor using these parameters.

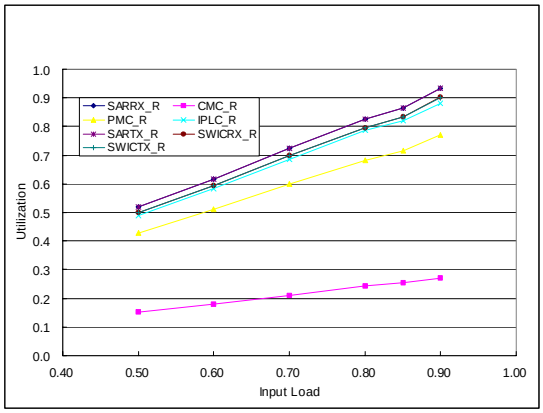


Fig. 5 Utilization of Controllers

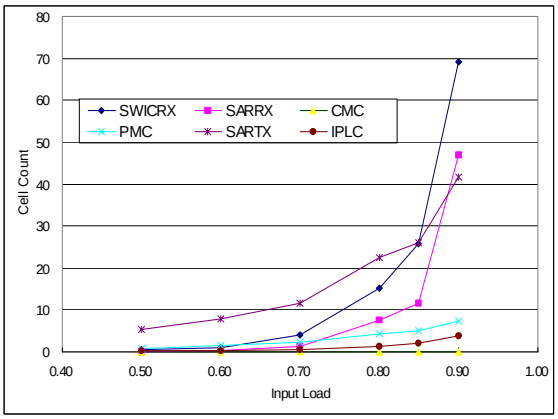


Fig. 6 Queue Length of Controllers

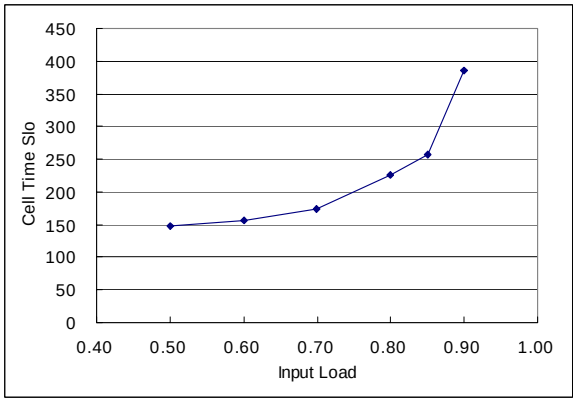


Fig. 7 IP Packet Delay/Offered Traffic Load

Fig. 5 shows the utilization of each controller in FE as the offered load is increased. As shown in Fig. 5, both SARRX and SARTX have the highest value in utilization than that of IPLC. IP Packet Delay/Offered Traffic Load will be a major bottleneck as an offered load is increased. We know that SWIC_RX and SAR_TX of Forwarding Engine have more bottlenecks than any other parts considering the average waiting time of MPLS system based on our MPLS-ATM switch. Fig. 6 shows the queue length corresponding to highly utilized controllers as the offered traffic load changed. We can see that the queue length is drastically increased as the offered load is beyond 0.85. The simulation result in packet delay against offered traffic load came out a similar result as shown in Fig.7. This means that, in our system based on FE queuing model, an offered load of pure IP packet data traffic except control information cells must be set to below 0.85. From this standpoint, we can estimate our maximum IP packet processing capacity(a) and desirable IP lookup time limits(b) in IPLC as a=152,7000, 5 TS< b<8TS respectively.

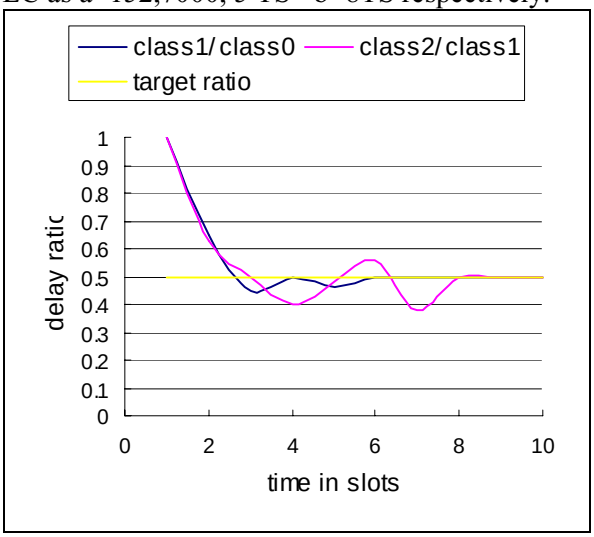


Fig. 8 Delay Ratio between Delay Classes

In Fig. 8 shows that delay ratio between classes by our delay mechanism, where $\bar{s}_i / \bar{s}_{i-1}$ equals 2, approaches to the target ratio(0.5) through the whole time spanning required except the first initial time period.

6 Conclusion

This paper proposed Queuing model for Forwarding Engine, which performs IP address lookup in MPLS LER, and analyze performance of FE with Internet traffic by applying statistical Internet protocol analysis method. Estimating the time slot limit for IP lookup and IP packet processing capacity of Forwarding Engine is useful to plan upgrading and modification of the system and network deployment involved. Including Diffserv function, we are in the completion phase for pilot system operated in ships-in-the-night based on this result. In addition to this, in order to increase the performance factors, we will design an independent LER system, which is operated in a separate fashion from LSR, considering buffer space for SWIC part and SAR. And we'll try to consider traffic management framework for a better performance of our whole system as the future study.

References:

[1] Eric C. Rosen, Arun Viswanathan, Ross Callon, Multiprotocol Label Switching Architecture*, RFC3031, IETF*, Jan. 2001

[2] Loa Anderson, Paul Doolan, Nancy Feldman, Andre Fredette, Bob Thomas, LDP Specification, RFC 3036 *, IETF*, Jan. 2001

[3] Jaesup Lee, B. Kim, et al., A Novel Switching System architecture for the Internet Access, *ICT'98*, Greece, May 1998, pp.244-248

[4] B. Kim, Jaesup Lee, et al., A Structure of Multi-protocol Label Switching-based HANbit ACE Switching System, *ATM Technology Workshop*, Bejing, Nov. 1998

[5] Jaesup Lee, B. I. Kim, H. H. Lee, K. H. Ryu, A Functional Architecture for Service Classification based on Multiprotocol Label Switching Network, *ICT'99*, Cheju Korea, June 1999, pp.426-430

[6] Jaesup Lee, An Implementation of MPLS Controller for ATM and IP Integration Networks, *MPLS'99*, Paris France, June 1999

[7] Pritsker, A. Alan B, *Simulation With Visual Slam and Awesim.* John wiley & Sons(Sd), 1997

[8] Kevin Thomson, et. al., Wide-Area Internet Traffic Patterns and Characteristics, *IEEE Network*, Nov/Dec.1997, pp. 10-23

[9] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, An architecture for differentiated services, *IETF RFC 2475*, Dec. 1998

[10] C. Dovrolis, and D. Stiliads, Relative Differentiated Services in the Internet: Issues and Mechanisms, *ACM SIGMETRICS*, May 1999

[11] T. Nandagopal, and et. al., Delay Differentiation and Adaptation in Core Stateless networks, *Infocom 2000*, Mar. 2000, pp. 651-660

[12] *Cell Maker Chip Manual*, Lucent Technology