Using Measurements to Validate Simulation Models of TCP/IP over High Speed ATM Wide Area Networks

Georgios Y. Lazarou, Victor S. Frost, Joseph B. Evans, Douglas Niehaus

Telecommunications & Information Sciences Laboratory
The University of Kansas
2291 Irving Hill Road
Lawerence, KS 66045-2228
email: frost.@tisl.ukans.edu
http://www.tisl.ukans.edu

Abstract

Predicting the performance of high speed wide area ATM networks (WANs) is a difficult task. Evaluating the performance of these systems by means of mathematical models is not yet feasible. As a result, the creation of simulation models is usually the only means of predicting and evaluating the performance of such systems. In this paper, we use measurements to validate simulation models of TCP/IP over high speed ATM wide area networks. Validation of simulations with measurements is not common; however, it is needed so that simulation models can be used with confidence to accurately characterize the performance of ATM WANs. In addition, the appropriate level of fidelity of the simulation models needs to be determined. The results show that feasible simulations models can accurately predict the performance of complex high speed ATM wide area networks.

1: Introduction

Predicting the performance (that is, throughput) of high speed wide area ATM networks (WANs) is a difficult task. Simulation models enable the systematic performance evaluation of these systems when mathematical models are not available and experiments on the actual systems are impossible or impractical [7]. However, in most cases a very high level of model fidelity is required for accurate performance prediction.

In this paper, the level of model fidelity required to accurately predict measured performance is determined by using measurements from high speed wide area ATM networks to validate simulation results. We present validation results of simulations of TCP/IP over the MAGIC gigabit testbed and the ACTS ATM Internework (AAI) wide area research network. The experimental results presented here also show that today's workstations can fully utilize OC-3 links. These simulation results indicate that network congestion effects can be accuratelly modeled, and thus simulation models can accurately predict the performance of complex high speed ATM wide area networks.

The rest of this paper is organized as follows: Section 2 provides an overview of TCP/IP over ATM; Section 3 presents background information on the simulation software environment, the TCP model, and the MAGIC and AAI networks; Section 4 describes the validation of simulation with experimental results; and Section 5 discusses lessons learned.

This research is partially supported by ARPA under prime contract DABT63-94-C-0068 to Sprint Corporation and under ARPA contract F19628-92-C-0080, Digital Equipment Corporation, the Kansas Technology Enterprise Corporation, and Sprint Corporation.

2: Overview of TCP/IP over ATM

2.1: Importance of TCP/IP over ATM

Probably the most succesful idea in data networking over the past twenty years has been the concept of internetworking [13]. It is a method for interconnecting networks, regardless of the particular networking technology (Ethernet, ATM, HIPPI, Frame Relay, FDDI), used by the individual systems. What makes internetworking possible is the development of protocols like TCP/IP. The TCP/IP protocol suite is the internetworking protocol used on the Internet, a global collection of networks connecting millions of computers and users, and incorporating a large variety of different network technologies [13]. It allows computers of all sizes, from many different computer vendors, running totally different operating systems, to communicate with each other [17].

ATM technology is the emerging standard adopted by telecommunications and computer vendors for high speed networks. The fast-cell switching technology employed by ATM helps to provide scalable (in size and speed) networks [5]. Further, ATM technology provides bandwidth on demand and enables the integration of real-time and data traffic over the same physical medium for wide area networks.

Although TCP/IP and ATM often have been viewed as competitors, their complementary strengths and limitations form a natural alliance that combines the best aspects of both technologies [14]. In the near future, a large portion of the traffic carried by the ATM networks will be generated by applications written to run over a TCP/IP protocol stack [11]. In fact, many of the existing ATM networks employ TCP/IP over ATM technology.

2.2: Previous Studies of TCP/IP over ATM

Numerous simulation and experimental studies have been performed in order to predict the performance of TCP/IP over ATM under congestion and buffer overflow conditions that arise from bandwidth mismatches or multiple sources contending for the same link [2, 5, 8, 11, 15, 16, 19]. The simulation study in [16] examines the performance of TCP over ATM under conditions of network congestion. The results show that the TCP/IP over ATM performance is poor when there is congestion caused by small switch buffers and large TCP segment and window sizes. This performance degradation is caused by a loss-rate multiplier effect caused by the switch dropping cells from multiple packets. In [15], the performance of TCP connections over ATM networks without ATM-level congestion control is investigated. Simulation results of congested networks show that the effective throughput of TCP over ATM can be quite low when cells are dropped at the congested ATM switch. To improve the performance, a mechanism called early packet discard (EPD) which brings throughput performance to its optimal level is proposed.

The work in [11] considers some undesirable interactions between the congestion control scheme used in TCP and the policing mechanisms used in ATM networks that can significantly degrade the throughput of TCP traffic. It is shown that in the presence of policing, once a TCP connection has increased its window size beyond the sustainable cell rate (SCR) times the round–trip time and if the bottleneck capacity exceeds the SCR, the buffer at the site providing the policing mechanism fills up quickly and most of the packets in the TCP window are dropped. This causes the average throughput to be significantly lower than SCR value. In order to improve the performance, the use of smarter policing or cell-level traffic shaping schemes is suggested.

The work in [2, 12] describes performance measurements taken from the MAGIC gigabit testbed relating to the performance of TCP in wide area ATM networks. Results show that the TCP rate control mechanism alone is inadequate for congestion avoidance and control in wide area gigabit networks. It is also illustrated that TCP augmented by cell-level pacing allows the full link capacity to be utilized. Cell level pacing is necessary because the TCP rate control mechanism does not control traffic burstiness sufficiently to avoid congestion-induced cell losses in wide area networks.

These studies separately present simulation and measurement results. However, comparison of performance predictions from measurement techniques and simulation models of TCP/IP-ATM networks is needed to refine our understanding of the network operation. Validation of models against measurements is also required so simulation models can be used with confidence to capture the effects of network control, and to accurately characterize the performance of ATM WANs. However, in order to achieve that goal, the precision level of the models (which is directly proportional to the model execution time) needs to be determined because of the computational complexity of the simulation.

3: Background

This section provides some background information on the simulation software environment, the TCP BONeS model, and the MAGIC and AAI networks.

3.1: The Simulation Software Environment

The simulation software environment used for all our simulations is BONeS DESIGNER [18]. It is a software package for modeling and simulating event-driven systems. A system model can be constructed hierachically and graphically using building blocks from the BONeS model library, or using models written in C or C++.

3.2: TCP BONeS Primitive Module

A TCP BONeS [18] primitive module was created for this study. The source code for this primitive was based on the MIT Network Simulator (NetSim) TCP module [9, 10]. However, using NetSim module we were unable to match measurement with simulation results in the presence of network congestion because specific TCP timer mechanisms were not modeled. The NetSim TCP module is partially based on the Berkeley Standard Distribution (BSD) 4.3 Tahoe version. All TCP implementations that are 4.3 BSD based include two timer functions: one is called every 200 ms (the fast timer) and the other every 500 ms (the slow timer) [17]. The fast timer is used with the delayed ACK timer and the slow timer is mainly used with the retransmission timer. The NetSim TCP model did not include these timers and thus can not accurately predict the performance of TCP over ATM under congestion.

The modified TCP model developed here is based on the 4.3 BSD Reno version. It supports the following major mechanisms:

- fast and slow timers,
- slow start and congestion avoidance,
- fast retransmit and fast recovery,
- window advertisement.

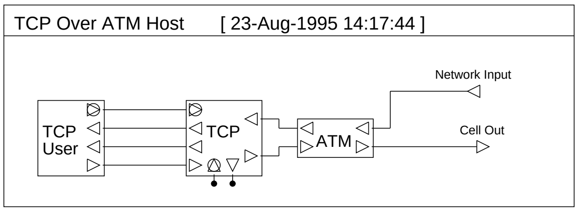

Figure 1. TCP over ATM BONeS host model

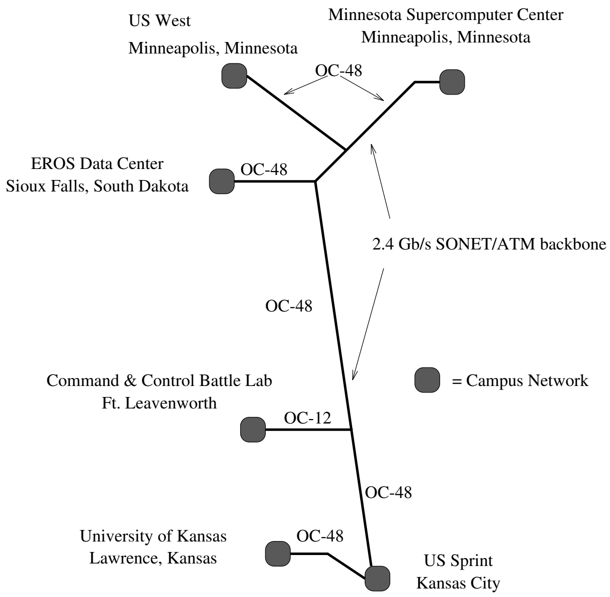

Figure 2. MAGIC Network

The BONeS model of a TCP over ATM host is shown in Figure 1. The TCP User block generates the data buffers for transmission and the ATM block accepts TCP packets and generates ATM cells. It also accepts cells from the network and reconstructs the TCP packets.

From Figure 1 we see that an IP module is not used. IP is the main protocol at the network layer of the TCP/IP protocol suite. It provides an unreliable, connectionless datagram delivery service between hosts attached to an TCP/IP internetwork. In this study, the connectioneless IP datagrams are carried only on connection-oriented ATM networks. Hence, the IP routing functionality is redundant in our simulation models.

3.3: MAGIC and AAI Networks

This study focuses on measurements and models of the MAGIC and AAI networks. The Multidimensional Applications and Gigabit Internetwork Consortium (MAGIC) is a group of industrial, academic, and government organizations participating in gigabit networking research. The MAGIC backbone network (see Figure 2) operates at 2.4 Gb/s and each site on the network includes LANs or hosts communicating at gigabit per second rates [6].

The ACTS ATM Internetwork provides wide area Asynchronous Transfer Mode (ATM) connectivity. It connects several DoD High Performance Computing centers and the MAGIC and ATDnet gigabit testbeds. The ATM service is provided by Sprint. The AAI network is depicted in Figure 3.

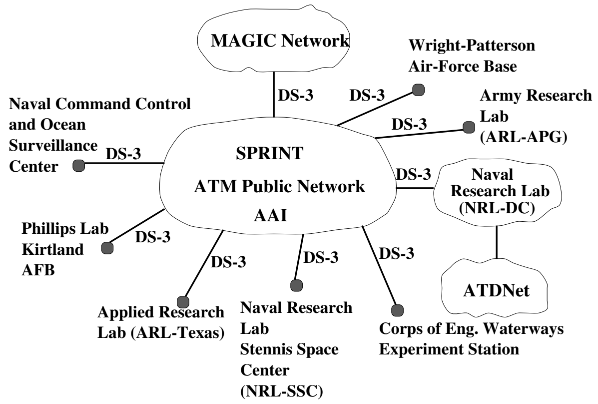

Figure 3. AAI Network

System Parameter	Value
TCP MTU Size	9180 Bytes
TCP Processing & OS Overhead Time	
DEC 3000 AXP	200–300 μs
SGI	550 μs
SPARC 10	550 μs
SPARC 5	700 μs
TCP User Send Buffer Size	64 KBytes
Slow-Timer Period	0.5 s
Fast-Timer Period	0.2 s
Minimum RTO	1.0 s
AAL5 SAR Processing Time	0.2 μs
AAL5 Cell Payload Size	48 Bytes
Switch Processing Time	4 μs
Switch Output Buffer Size per VC	256 Cells
OC–3c Link Speed	155 Mb/s
TAXI Link Speed	100 Mb/s
DS–3 Link Speed	45 Mb/s

Table 1. Run-Time Simulation Parameters

4: Using Measurements to Validate Simulation Models

This section presents the model validation results of TCP over ATM performance for three different cases:

- over the MAGIC network and varying the TCP window size,
- over the MAGIC network and under rate mismatch conditions, and
- over the AAI network.

In addition, this section provides information about the simulation system parameters and network traffic.

4.1: System Parameters

Beside the level of model fidelity, the simulation system parameters need to be correctly specified. The system parameters that are used in our simulation systems are listed in Table 1. The maximum size of the TCP segment is specified by the TCP MTU (maximum transmission unit) size parameter. TCP Processing and OS Overhead Time is the overall time needed by the TCP software to create a segment for transmission or process an incoming segment and for the operating system to handle all system calls and I/O operations during transmission or reception of a TCP segment. In Figure 1, the TCP User module sends data buffers to the TCP module for transmission. The size of these data buffers is designated by the TCP User Send Buffer Size parameter. The retransmission timer is decremented every 0.5 seconds (Slow-Timer Period), and only when the timer reaches 0 a retransmission is performed. A delayed ACK is sent every time the 0.2 second delayed ACK timer (Fast-Timer) expires. The retransmission timer is bounded by TCP to be between 1 (Minimum Retransmission Time-Out) and 64 seconds [17].

Link	Max SONET Layer Throughput	Max TCP Layer Throughput
OC-3	149 Mb/s	135 Mb/s
100 Mb/s TAXI	98 Mb/s	88 Mb/s
DS-3	40.7 Mb/s	36 Mb/s

Table 2. Maximum Throughput for Different Link Speeds

For this study, the host interfaces use the ATM adaptation layer AAL5 for mapping IP datagrams to ATM cell streams. The AAL5 Segmentation and Reassembly Processing Time is the time required for the AAL5 SAR sublayer to map the IP datagrams to ATM cell streams or to reconstruct cell streams to IP datagrams. The AAL5 cell payload is 48 bytes long. The values of the TCP Processing and OS Overhead Time, AAL5 Segmentation and Reassembly Processing Time, and Switch Processing Time parameters are based on experimental measurements [1, 3]. For reference, Table 2 lists the maximum achievable physical and TCP level throughput for SONET links after excluding the SONET, ATM, and IP overhead.

4.2: Network Traffic Characterization

To measure the maximum end-to-end throughput at the TCP layer, a public domain software tool, *ttcp*, was used in all expirements. This tool transfers TCP packets from local memory to memory on a remote host as fast as the operating system, interfaces, and network allow. In each of the experiments performed for this study, *ttcp* used 64 KBytes user data buffers. In order to match the *ttcp* traffic, the data send buffer at the transmitting host always was kept full in all the simulations presented here.

4.3: Validating ATM WAN vs TCP Window Size Performance

An experiment was carried out by Ewy and Evans [2] using the MAGIC testbed in order to study the ATM WAN performance versus TCP window size. The experimental set up is shown in Figure 4. A Digital DEC 3000 AXP with an OTTO OC-3 interface transmits to another DEC 3000 AXP over a SONET OC-3 link with 8.8 round-trip delay. A simulation for this experiment is created using our BONeS TCP/ATM host model. Figure 5 shows the top level of the simulation model. The BONeS model for the two DEC AXP 3000 blocks uses the TCP over ATM host model shown in Figure 1. The two OC-3 links are modeled by delay blocks, while the DIGITAL AN2 Switch is modeled by a simple FIFO queue with a server. The simulation and experimental results are shown in Figure 6. These results match our expectations given the bandwidth-delay product of this link. The results of this experiment clearly show that today's workstations can fully utilize an OC-3c link.

4.4: Validating TCP/ATM Performance Under Rate Mismatch Conditions

Another experiment was conducted by Ewy and Evans [2] in order to study the performance of TCP over ATM under the case of rate mismatch. The configuration for this experiment is shown in Figure 7. A single host transmits to another host with a 155 Mb/s to 100 Mb/s bandwitdh constriction in the path. The simulation system used for this experiment is shown in Figure 8. The TCP

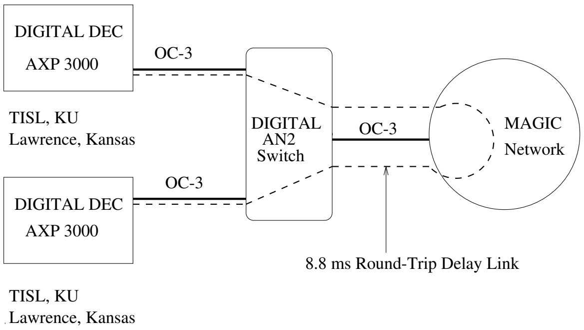

Figure 4. Experimental Set Up for ATM WAN Performance vs TCP Window Size

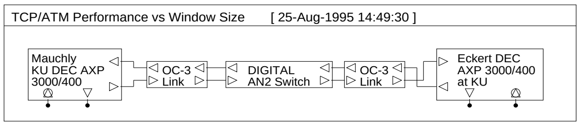

Figure 5. Simulation Model for ATM WAN Performance vs TCP Window Size

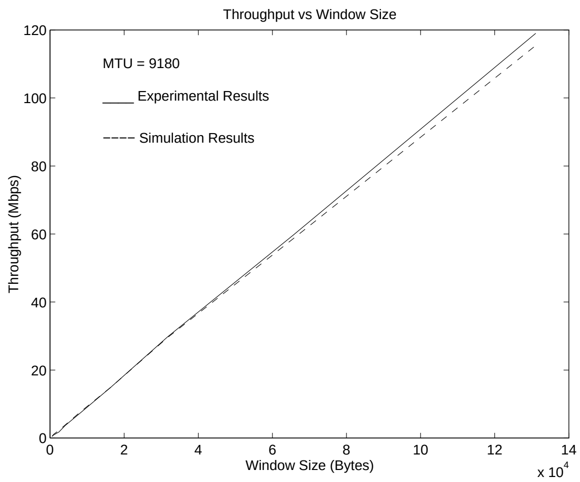

Figure 6. ATM WAN Performance vs TCP Window Size

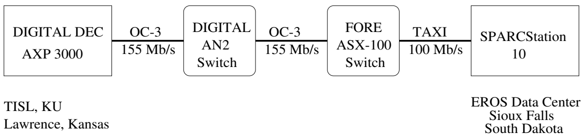

Figure 7. Experimental Set Up for TCP/ATM Performance under Rate Mismatch Conditions

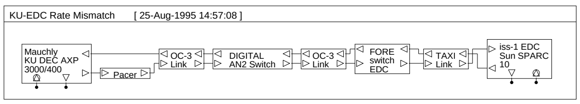

Figure 8. Simulation Model for TCP/ATM Performance under Rate Mismatch Conditions

	No Pacing	Pacing
Experimental Results	1.26 Mb/s	68.20 Mb/s
Simulation Results	1.22 Mb/s	68.59 Mb/s

Table 3. Throughput of TCP over ATM with Rate Mismatch

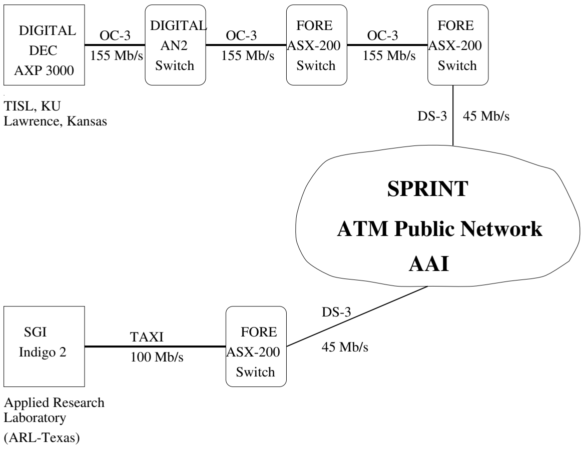

Figure 9. Experimental Set Up of a Single Connection in the AAI Network

window size is set to 128 KB and the cell-level pacing is fixed at a rate of 70 Mb/s. Cell-level pacing is the mechanism which reduces the source cell transmission rate. The Pacer functionality is modeled by an infinite size FIFO queue and a server with service rate equal to the pacing rate. The results are shown in Table 3. This confirms that congestion effects can be accurately modeled.

4.5: Validating TCP/ATM Performance over AAI Network

A simulation model was created for predicting the performance of the AAI Network. The simulation results over a single connection were then validated by measurement. The experimental configuration is shown in Figure 9. A Digital DEC 3000 AXP with an OTTO OC-3 interface transmits to an SGI Indigo 2 host with a 100 Mb/s TAXI interface. The round-trip time for this connection is about 36 ms. The simulation model for this experiment is shown in Figure 10. Note in Figure 9 that the lowest link capacity is DS-3. Hence, a cell-level pacer is used with a rate of 40 Mb/s since the maximum rate of a DS-3 link after excluding the physical layer overhead is 40.7 Mb/s. We used 40 Mb/s at the pacer to avoid any possibility of cell loss due to network congestion. A TCP window size of 256 KB is used. The results are shown in Table 4. Once more, the results show that simulations can accurately predict the performance of complex high speed ATM wide area networks.

A set of three experiments were carried out by Dasilva [4] over the AAI network in order to study the ATM WAN performance versus multiple simultaneous traffic streams between

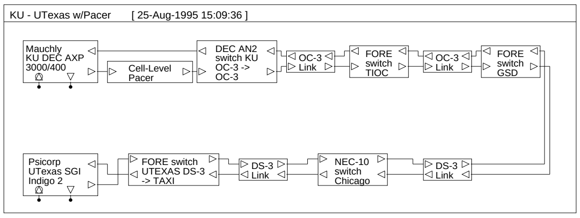

Figure 10. Simulation Model of a Single Connection in the AAI Network

Experimental Results	35.86 Mb/s
Simulation Results	35.97 Mb/s

Table 4. Throughput of TCP over ATM using AAI Network

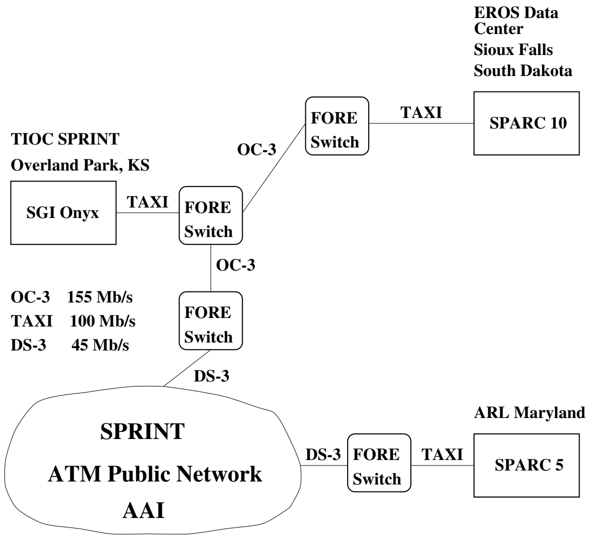

Figure 11. Experimental Set Up of Simultaneous connections over the AAI Network

multiple hosts:

- Experinent 1: an SGI Onyx host with a TAXI interface transmits to a SPARC 10 and SPARC 5 hosts with TAXI interface simultaneously.
- Experiment 2: a SPARC 10 and a SPARC 5 hosts transmit simultaneously to an SGI Onyx host.
- Experiment 3 : experiment 1 and 2 carried out simultaneously.

The experimental set up is shown if Figure 11. In these experiments, the lowest link capacity, DS-3, is between the TIOC and ARL connection. However, a cell-level was not used because the maximum throughput of SPARC 5 with SunOS 4 is less than the DS-3 rate. The round-trip times for TIOC-EDC and TIOC-ARL connections are about 9 ms and 25 ms respectively. For the TIOC-EDC connection, the TCP window size was set to 128 KB, and for the TIOC-ARL it was set to 32 KB. The BONeS simulation model developed for these experiments is shown in Figure 12. The simulation versus experimental performance results are shown in Table 5.

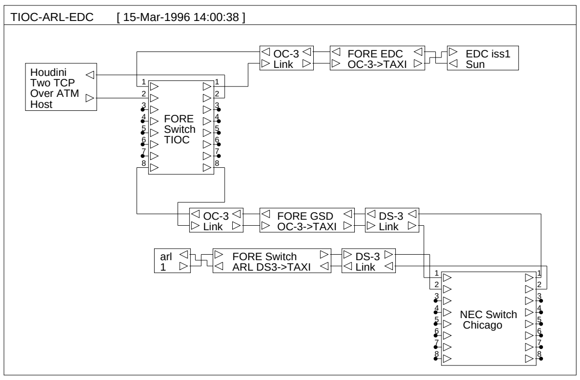

Figure 12. BONeS model of Simultaneous connections over the AAI Network

Connection	Experimental Results	Simulation Results
Baseline Results: Point-to-Point connections		
TIOC-to-ARL	4.2 Mb/s	7.18 Mb/s
TIOC-to-EDC	64.2 Mb/s	65.98 Mb/s
Simultaneous traffic streams: single source, two destinations		
TIOC-to-ARL	4.45 Mb/s	4.60 Mb/s
TIOC-to-EDC	64.36 Mb/s	61.37 Mb/s
Simultaneous traffic streams: two sources, single destination		
ARL-to-TIOC	2.15 Mb/s	4.87 Mb/s
EDC-to-TIOC	52.42 Mb/s	65.01 Mb/s
Simultaneous full duplex traffic streams:		
TIOC-to-ARL	4.34 Mb/s	5.16 Mb/s
ARL-to-TIOC	4.3 Mb/s	5.16 Mb/s
TIOC-to-EDC	22.18 Mb/s	41.80 Mb/s
EDC-to-TIOC	31.18 Mb/s	41.30 Mb/s

Table 5. Simulation & Experimental Results of Simultaneous connections over the AAI Network

5: Lessons Learned

TCP imposes heavy processing overhead when individual packets are retransmitted immediately after they time out and subsequently retransmitted. To reduce this overhead, many TCP implementations use the slow and fast timer mechanisms to handle the retransmission and acknowledgement operations. However, the cost of doing that is a performance degradation. The long delay introduced by these timer mechanisms causes a significant reduction in the performance TCP over ATM when there are cell losses due to network congestion. All workstations used for this study employ TCP implementations which use these timers. Therefore, it was necessary to include them in our TCP model in order to match our simulation results with measurements. For example, in our simulation studies, we first used the NetSim TCP model which does not use these timers. The simulation result (without modeling these timers) for the experiment with no pacing presented in Section 4.4 was 24 Mb/s, which is 20 times greater than the experimental result.

Our results indicate that simulations can be used to accurately predict the performance of high speed wide area networks. However, to enable feasible simulations of such systems, the minimum level of model fidelity for each system element must be used. The simulation's run time is inversely proportional to the level of model fidelity. Hence, simulation's run time can be reduced significantly if redundant model precision is avoided. Table 6 shows the minimum required level of model fidelity used in our simulation systems. Detailed IP and SONET models are not required in our simulation systems; their impact is captured by simply accounting for their information overhead. We avoided the need for a SONET model at the physical layer by reducing the OC-3 link speed from 155 Mb/s to 149 Mb/s. An IP model is not used because the IP routing fuctionality is not needed in ATM networks (see Section 3.2). Here, the ATM switch is modeled only by a single FIFO queue and a server. High precision for this switch model is unnecessary since a single connection with no cross-traffic was considered. Also, the details of the pacing algorithm were not modeled, since a simple FIFO was sufficient to capture its impact on performance. To obtain the results presented in this study, a high level of fidelity was needed for the TCP and ATM modules shown in Figure 1. In these models, we left out only references to Unix and IP specific functions and the open and close connection operations. By using the different levels of model fidelity shown in Table 6, we were able to reduce the simulation time significantly. The run time of the simulation system shown in Figure 8 on a SPARCstation-10 with 120 Mbyte of RAM is about 20 to 30 minutes for each second of real-time. If accurate IP, SONET, and switch models were used, the simulation of these high speed wide area ATM networks would be difficult.

Model	Level of Fidelity
ATM	High
ATM Switch	Medium
IP	Low
Link	Medium
Pacer	Medium
SONET	Low
TCP	High
TCP User (Application Layer)	Low

Table 6. Level of model Fidelity

The ATM Forum is currently in the process of standardizing traffic and congestion control techniques at the cell level for ATM technology. The results presented here show that simulations can be used to capture and evaluate the interactions of these techniques with TCP packet flow and congestion control mechanisms. Questions like "Can TCP packet flow and congestion control techniques cooperate with ATM cell flow and congestion control mechanisms" can be answered using the simulation models developed here.

References

[1] T. E. Anderson, S. S. Owicki, C. P. Thacker, "High Speed Switch Scheduling for Local Area Network," DEC Internal Publication, 1993.

[2] Benjamin J. Ewy, Joseph B. Evans, Victor S. Frost, Gary J. Minden, "TCP/ATM Experiences in the MAGIC Tested," Fourth IEEE International Symposium on High Performance Distributed Computing, Aug. 1995 pp. 87-93.

[3] H. Y. Chen, J. A. Hutchins, N. Testi, "TCP Performance over Wide Area Networks," SAND93-8243, September 1993.

[4] Luiz Dasilva, Personal Communication, December 1995.

[5] Chien Fang, Helen Chen, Jim Hutchins, "Simulation Analysis of TCP Performance in Congested ATM LAN using DEC's Flow Control Scheme and two Selective Celldrop Schemes," in ATM Forum Contribution 94-0119, Jan. 1994.

[6] B. Fuller, I. Richer,"An Overview of the MAGIC Project," MITRE Technical Report M93B0000/73, December 1993.

[7] Hisashi Kobayashi, *Modeling and Analysis: An Introduction to System Performance Evaluation Methology*, Addison-Wesley, Reading, Massachusetts, 1978.

[8] T. V. Lakshman, U. Madhow, "Performance Analysis of Window-based Flow Control using TCP/IP: Effect of High Bandwidth-Delay Products and Random Loss," IFIP Transactions on High Performance Networking, V C-26 1994.

[9] Georgios Y. Lazarou, Victor S. Frost, "BONeS TCP Primitive Module Validation," TISL Technical Report TISL-10980-06, The University of Kansas, 1995.

[10] D. Martin, *Network Simulator User's Manual*, MIT, 1988.

[11] Partho Pratim Mishra, "Throughput Degradation for TCP over ATM in the Presence of Traffic Policing," IEEE ComSoc TCGN Gigabit Networking Workshop 1995.

[12] Iqbal Mohammed, Victor S. Frost, "Validation Measurements for LANSE (DRAFT)," TISL Technical Report TISL-11560-01, The University of Kansas, 1995.

[13] Craig Partridge, *Gigabit Networking*, Addison-Wesley, Reading, Massachusetts, 1994

[14] Guru Parulkar, Douglas C. Schmidt, Jonathan S. Turner, "GIPR: A Gigabit IP Router," IEEE ComSoc TCGN Gigabit Networking Workshop 1995.

[15] Allyn Romanow, Sally Floyd, "Dynamics of TCP Traffic over ATM Networks," IEEE Journal on Selected Areas in Communications, VOL. 13, NO. 4, May 1995.

[16] Allyn Romanow, "TCP over ATM: Some Performance Results," in ATM Forum Contribution 93-784, July 1993.

[17] W. R. Stevens, *TCP/IP Illustrated, Volume 1,2*, Addison-Wesley, Readings, Massachusetts, 1994.

[18] Systems & Networks, *BONeS DESIGNER 3.0 Modeling Guide*, Lawrence, KS, 1995.

[19] Hongbo Zhu, Luiz A. Dasilva, Joseph B. Evans, Victor S. Frost, "Performance Evaluation of Congestion Control Mechanisms in ATM Networks," Computer Measurement Group Annual Conference (CMG'95), Dec. 1995.