Experimental and Simulation Performance Results of TCP/IP over High-Speed ATM over ACTS*

Charalambous P. Charalambos, Georgios Y. Lazarou, Victor S. Frost
Joseph Evans, Roelof Jonkman

Information and Telecommunication Technology Center
Department of Electrical Engineering & Computer Science
The University of Kansas
2291 Irving Hill Road
Lawrence, KS 66045-2228
Phone: (913) 864-4833
Fax: (913) 864-7789
E-mail: frost@ittc.ukans.edu

September 10, 1997

Abstract

Simulation and practical experiments in a satellite network environment assist in the design and understanding of future global networks. This paper describes the practical and simulation experiences gained from a TCP/IP on ATM network over a high speed satellite link and presents performance comparison studies of such networks with the same host/traffic configurations over local area (LAN) and wide area (WAN) networks. These performance comparison studies on the LAN, WAN, and satellite environments increase our understanding of the behavior of high bandwidth networks. It was found that the satellite systems deliver results similar to the terrestrial fiber-optic systems regardless their path latencies in cases where the communication channels are not noisy and exhibit low bit error rates (BER). NASA's Advanced Communications Technology Satellite (ACTS), with its special characteristics and high data rate satellite channels, and the ACTS ATM Internetwork (AAI) were used in these experiments to deliver broadband traffic. Network performance tests were carried out using application-level software (ttcp, Netspec) on OC-3 (155.54 Mbps) and OC-12 (622.08 Mbps) ATM satellite links.

*This research is partially supported by DARPA under prime contract DABT63-94-C-0068.

1 Introduction

Communication satellites, with their broadcast characteristic provide an effective and useful platform for establishing links between areas that are inaccessible by terrestrial communication facilities. Advanced satellite systems will compete with terrestrial fiber-optic networks in terms of high transfer rates, very low bit error rates (BER), and will become significant players in the Global Information Infrastructure (GII) in the future. Asynchronous Transfer Mode (ATM) cell switching technology offers the flexibility to handle and transfer advanced broadband services (voice, data, video) and integration of those in the same network at high transfer rates. The combination of ATM and satellite technologies, both using the most widely used protocol suite in the computer communication world, that is the Transmission Control Protocol/Internet Protocol (TCP/IP), form an internetwork architecture that has the potential to provide seamless networking. Such networks will enable the transfer of data and multimedia communications on the same network, everywhere in the world.

This paper presents results from experiments conducted as a part of the ACTS ATM Internetworking (AAI) project. The AAI testbed provides wide area ATM connectivity to several DoD High Performance Centers, and the MAGIC and ADTnet gigabit testbeds. Our experiments focus on performance measurements on OC-3 and OC-12 ATM using standard TCP/IP hosts over LAN, WAN, and ACTS high data rate (HDR) channels. These experiments show that the performance of TCP/IP on ATM networks over NASA's ACTS, with its special characteristics and low BER channels, are comparable with similar architectures on terrestrial fiber networks. Also, simulations modeling these experiments were constructed using the BONeS DESIGNER software. Performance predictions based on these models were validated by the experiments and thus showed that these are appropriate tools to gain additional understanding of network behavior.

In geosynchronous earth orbit (GEO) satellite systems, performance is affected by the inherent latency due to the speed of light impairment and the distance of the satellite from the earth's surface, as well as the probability of bit errors on the satellite links. Since we are running TCP over ATM, different enhancements to TCP [14, 15, 19, 24], can achieve high throughput over satellite links, comparable with those on terrestrial fiber networks.

The rest of this paper is organized as follows: Section 2 provides background on TCP; Section 3 describes TCP/IP over ATM; Section 4 describes the ACTS architecture; Section 5 presents the experimental scenarios; Section 6 presents the experimental results; Section 7 analyzes the results and discusses TCP pitfalls; Section 8 presents the simulation results; Section 9 states the conclusions of our work.

2 Transmission Control Protocol

TCP is the reliable, connection-oriented, end-to-end error, flow and congestion control protocol in the transport layer of the TCP/IP reference model. It is the most widely used protocol in the Internet, providing reliable transfer of data (packets). It was designed to work independently of the lower layer implementation for transferring data, such us ATM. It is sensitive to the host operating environment [8]. Its basic implementation

[23] is unsuitable for high speed and high delay networks, and therefore modifications and additions were added to enhance the performance of the protocol over such networks [14, 15, 19, 24]. This is one of the reasons for the variety of TCP versions available today. *TCP Reno* or TCP Reno like implementation versions used in the hosts under test in our experiments. Since we believe that the performance results from our experiments are affected greatly by the transport protocol, a description of the major characteristics of TCP Reno (slow start, congestion avoidance, fast retransmit, fast recovery, large windows) follows.

- **Slow Start and Congestion Avoidance** [24]

  TCP uses the slow start technique to recover from congestion. TCP uses slow start whenever it establishes a new connection, in order to avoid these connections instantaneously contributing to congestion. It is an algorithm for controlling the rate at which the sender transmits data into the network. The sender maintains a variable called congestion window (CWND) to measure the networks capacity. When a new connection is established with another host, the sender sets CWND to one segment and begins by transmitting one segment and waiting for its acknowledgment (ACK). When the ACK is received, it transmits two segments, and thus the CWND size is increased by one segment. From now on, two segments are injected into the network for every ACK received and thus CWND doubles every round trip time until the the window size advertised by the receiver is reached. This leads to an exponential growth of the window size and the time needed to utilize the full bandwidth (i.e. to reach the receiver's advertised window) is given by the formula [8]:

$$Slow\_Start\_Time = RTT * log_2 W \tag{1}$$

where RTT is the round trip time between the two hosts, and W is the number of segments fit in the receiver window size.

When a lost segment is detected, the CWND size is set to one segment, and the slow start algorithm starts over until the sender reaches half of the original CWND. From thereafter, TCP enters the congestion avoidance phase and slows down the rate of increment. During this phase, the sender transmits to the network one additional segment for each round trip time, until the receiver's advertised window is reached.

- **Fast Retransmit and Fast Recovery** [24]

  The TCP fast retransmit algorithm detects a segment loss quickly by counting three duplicate ACKs, and then immediately retransmits the lost segment without waiting the retransmission timer to expire.

  After fast retransmission, fast recovery is used instead of slow start, and is followed by congestion avoidance. During the fast recovery phase, the sender halves the CWND and gets to the congestion avoidance phase, without using slow start.

- **Large Windows** [14, 15]

  The significant parameter in high bandwidth networks with long delay paths is the product of bandwidth (bits per second) and round trip delay (seconds). This product is the number of bits it takes to fully utilize the network capacity, or to "fill the pipe". The bandwidth-delay product is actually the amount of unacknowledged data outstanding at any moment on the network, keeping the link or pipeline full, and it corresponds to the minimum buffer size or window size on the receiver host as given by the formula:

$$Window(Buffer\_size) = (Round\_Trip\_Time) * (Throughput) \qquad (2)$$

The upper limit of the TCP window is determined by the socket buffer space in the source and receiver UNIX operating system kernels.

The initial implementation of the TCP protocol [23] had the capability to provide only 65535 bytes of window sizes, by using 16 bits in the TCP header. According to equation (2), this was inappropriate for high bandwidth, high delay networks, like ATM over satellite links. Therefore extensions were added [15] to increase the window option to 32 bits in the TCP header. This enhancement is included in TCP Reno systems.

3 TCP/IP on ATM networks

ATM is a scalable cell switching and multiplexing technology that was chosen by ITU-T to be the transport technology for the Broadband Integrated Services Digital Network (B-ISDN). It is widely implemented in WANs and is used in LANs as well. Studies [3, 4, 6, 9] have shown that TCP (with the enhancements discussed in the section above) can achieve maximum performance over ATM in cases that there is no congestion and rate mismatch in the path between the sender and the receiver hosts. Classical IP has been implemented on ATM networks according to the recommendations in [17] to enable connection with other machines over internetworks.

In our implementations, ATM is transported by Synchronous Optical Network (SONET) protocols, a synchronous transport technology with extensive support for operations, administration, and maintenance (OAM) [26]. The fact that SONET permits irregular ATM cell arrivals and then transports them via its Synchronous Payload Envelope (SPE) made this system very popular for ATM transfer. We are also using AAL 5 (ATM Adaptation Layer 5) in our network architecture, which offers unreliable data transfer services with error detection.

4 ACTS

ATM transported over SONET systems is rapidly emerging as the transport mechanism for future high speed networks [6]. Broadband communication satellite systems provide an effective platform for world wide communications. Thus, ATM/SONET over satellite channels is the next step towards the implementation of high speed global networks [6].

NASA's ACTS is a unique satellite system providing SONET STS-3 (155.54 Mbps) and SONET STS-12 (622.08 Mbps) point-to-point and point-to-multipoint services, with possible direct support of STS-1 (51.84 Mbps) with the necessary multiplexing equipment in the ground stations [21]. The satellite is operating in the Ka band with a 20 GHz downlink frequency and a 30 GHz uplink frequency. The ACTS, the gigabit earth stations (GES), and the Network Management Terminal (NMT) form the gigabit satellite network (GSN), which uses the high data rate (HDR) section of ACTS with its dynamic beam-hopping and microwave switch matrix (MSM) capabilities. MSM is a 4-by-4 switch, with only a 3-by-3 subset to be usable at any given time, with the ability to support an aggregate user data rate of 622 Mbps on each of these three simplex channels through terrestrial interfaces at OC-12/12c and OC-3/3c rates.

All the GESs are equipped with standard SONET OC-3/3c and SONET OC-12/12c fiber interfaces to ensure the interoperability of the satellite network with the terrestrial network. These interfaces are provided by the Digital Terminal. The high data rate is possible because of the 348 Msymbols/s burst modem. Offset Binary Phase Shift Keying (BPSK) modulation, with 1 bit per symbol, generates a 348 Mbps maximum data rate, while offset Quadruple Phase Shift Keying (QPSK) modulation, with 2 bits per symbol, generates a 696 Mbps maximum data rate. Without the overhead for Forward Error Correction (FEC) and Time Division Multiple Access (TDMA) synchronization and alignment, ACTS HDR channels can support 311.04 Mbps (with BPSK) and 622.08 Mbps (with QPSK) of user data bit rate [13]. The BER specifications on the satellite links are similar to those on terrestrial fiber-optic networks. Reed-Solomon forward error correction encoding provides a BER less than 10^{-12} in clear sky conditions, and 10^{-11} for at least 99% of the time in the presence of rain-fade [7].

Transmissions to the satellite by the GESs are performed using Satellite Switched Time Division Multiple Access (SS-TDMA) techniques, coordinated by a reference terminal, which can be programmed by the NMT to be any of the GESs. Three uplink and three downlink antenna beams can be active simultaneously, with maximum 696 Mbps bursts per beam, resulting to an aggregate system bit rate of 2 Gbps. Forward Error Correction (FEC) and overhead utilizes about 10% of the total bandwidth, resulting in a user aggregate throughput of 1.8 Gbps (3 $\times$ OC-12) [7].

5 Experimental Scenarios

5.1 System Overview

The different machine architectures used throughout the experiments are shown in Table 1. Refer to this table when reference to machine names is made throughout the paper.

Name	NIC	Architecture - Clock Speed	Operating System
faraday(KU)	OC-12c	SUN UltraSPARC 1 - 167MHz	Solaris 2.6beta
mckinley(GSFC)	OC-12c	SUN UltraSPARC 1 - 167MHz	Solaris 2.5.1
hartley(KU)	OC-3c	SUN SPARC 20 - 125MHz	Solaris 2.4
wiley(KU)	OC-3c	DEC 3000/700 - 225MHz	Digital Unix 4.0A
elmer(KU)	OC-3c	DEC 3000/700 - 225MHz	Digital Unix 4.0B
galaga(KU)	OC-3c	DEC 3000/700 - 225MHz	Digital Unix 4.0
nrl(NRL)	OC-3c	DEC 3000/700 - 225MHz	Digital Unix 4.0A

Table 1: Workstations used throughout the experiments and their specifications.

The ATM switches used in the experiments are FORE ASX-1000 and FORE ASX-200BX models. These switches provide a shared buffer space of 8192 cells for Unspecified Bite Rate (UBR) traffic for each network module (four ports for SONET OC-3c or one port for SONET OC-12c). The UBR buffer space is allocated per virtual circuit (VC) dynamically on an as needed basis [22]. These switches also support the Early Packet Discard (EPD) algorithm, which in case of congestion, and therefore switch buffer overflow, discards the entire sequence of ATM cells belonging to a single packet, thereby not loading the link with unnecessary cells that will be retransmitted by TCP (in the packet level).

5.2 Description of scenarios

- **Scenario 1** is illustrated in Figure 1. In this scenario an OC-3c equipped Alpha workstation (wiley) at KU transmits to an OC-12c equipped UltraSPARC workstation (mckinley) at GSFC (Goddard Space Flight Center) via an OC-3 ACTS link. The purpose of this experimental scenario is to note the maximum TCP over ATM throughput that we can obtain on a SONET OC-3 satellite link. The two hosts are running TCP/IP on ATM/SONET interfaces. Congestion is not present

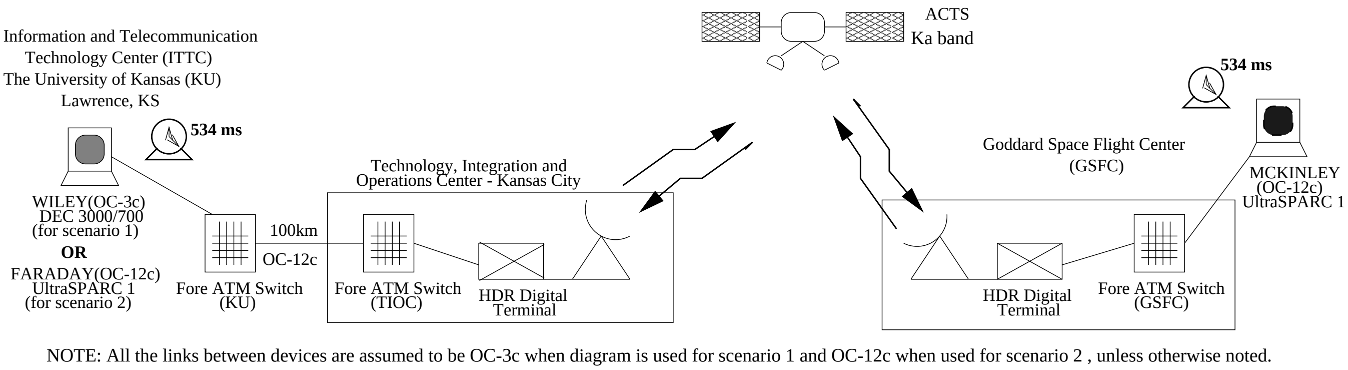

Figure 1: Diagram of scenario 1 and scenario 2; Congestion free TCP/IP over ATM over ACTS networking, for SONET OC-3c and OC-12c rates.

since the maximum rate of each machine is less than the rate of the ATM switch interfaces or the satellite link.

- **Scenario 2** is illustrated in Figure 1 as well. In this scenario an OC-12c equipped UltraSPARC workstation (faraday) at KU transmits to an OC-12c equipped UltraSPARC workstation (mckinley) at GSFC via an OC-12 ACTS link. The purpose of this experimental scenario is to note the maximum throughput that we can obtain on a SONET OC-12 satellite link. The two hosts are configured as above, and congestion is not present for the same reason as in scenario 1.

- **Scenario 3** is illustrated in Figure 2. This scenario represents tests in which congestion is present. Three OC-3c equipped workstations (elmer, galaga, wiley)

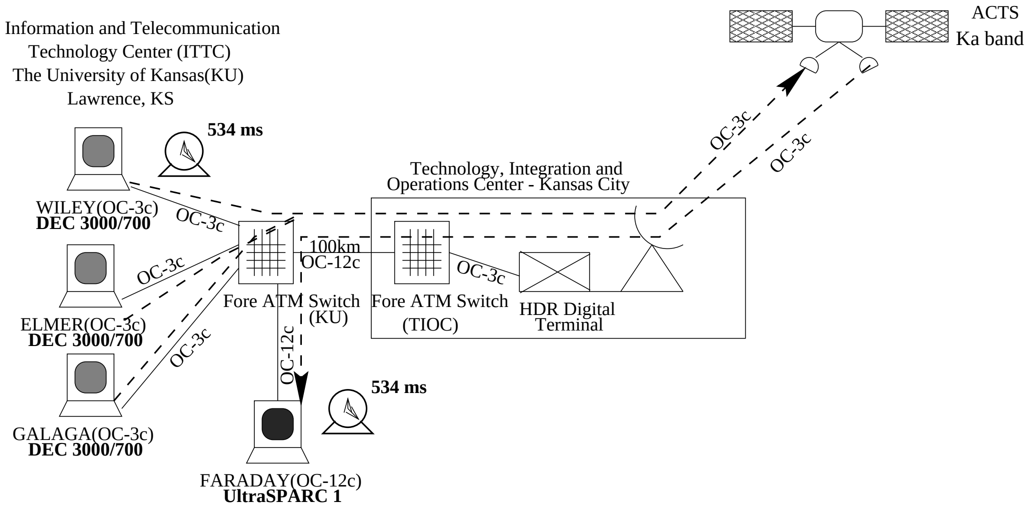

Figure 2: Diagram of scenario 3; TCP/IP over OC-3c ATM over ACTS networking under congestion conditions.

at KU transmit to an OC-12c workstation (faraday) at KU via an ACTS OC-3 link in loop-back mode. The transmitting stations are injecting data into the link at a faster rate than the link can handle, and therefore congestion is present. The purpose of this experimental scenario is to note the maximum throughput that we can obtain on a SONET OC-3 satellite link under congestion conditions. All the stations are configured as in scenario 1.

- **Scenario 4** is illustrated in Figure 3, where performance measurements are taken from the KU ATM LAN and the AAI WAN, for performance comparisons between

LANs, WANs, and the satellite system of scenario 1 (shown in Figure 1). This test was carried out under conditions of no congestion. All the hosts are configured as in scenario 1.

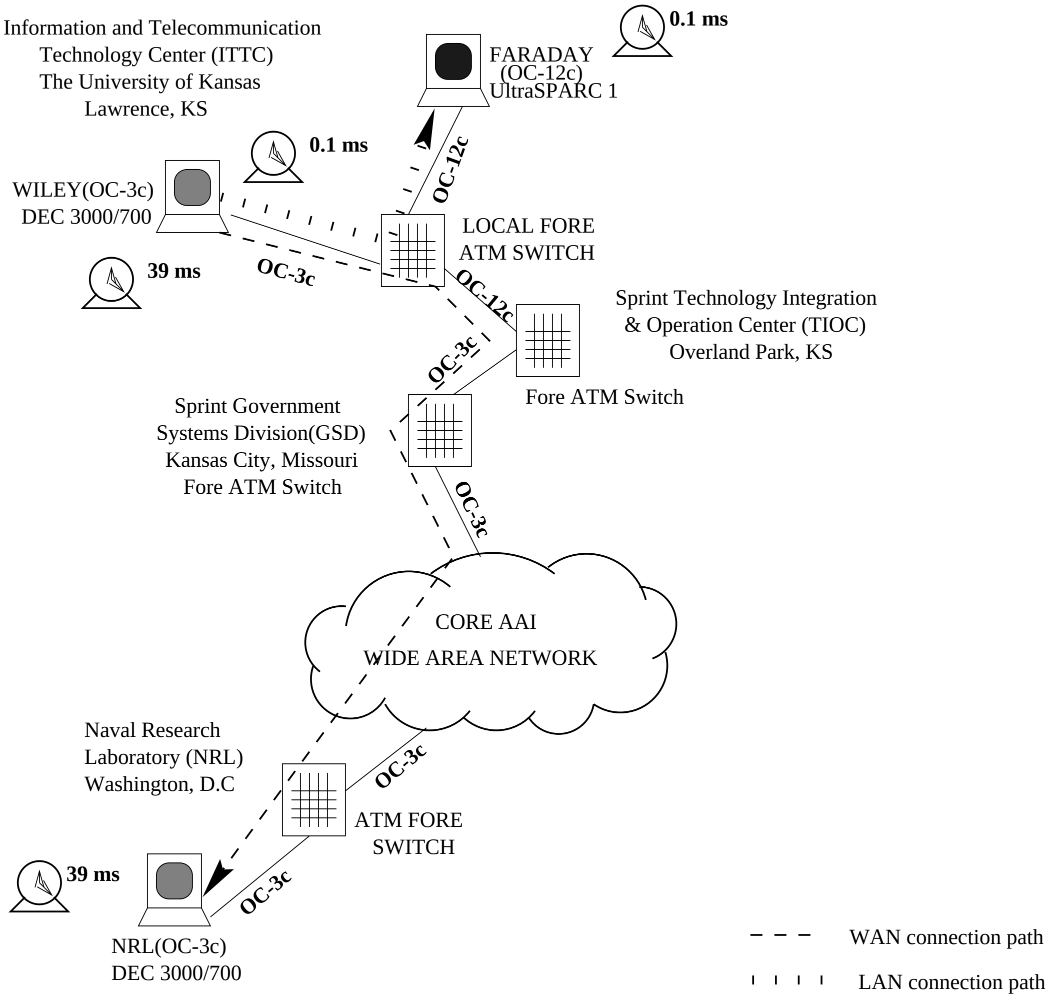

Figure 3: Diagram of scenario 4; Congestion free, TCP/IP over OC-3c ATM in the local ATM network (LAN) and the AAI WAN.

- **Scenario 5** is illustrated in Figure 4, where performance measurements are taken from the KU ATM LAN and the AAI WAN, for performance comparisons under congestion conditions on the LAN, WAN, and the satellite system of scenario 3 (shown in Figure 2). Congestion is generated when all three workstations shown

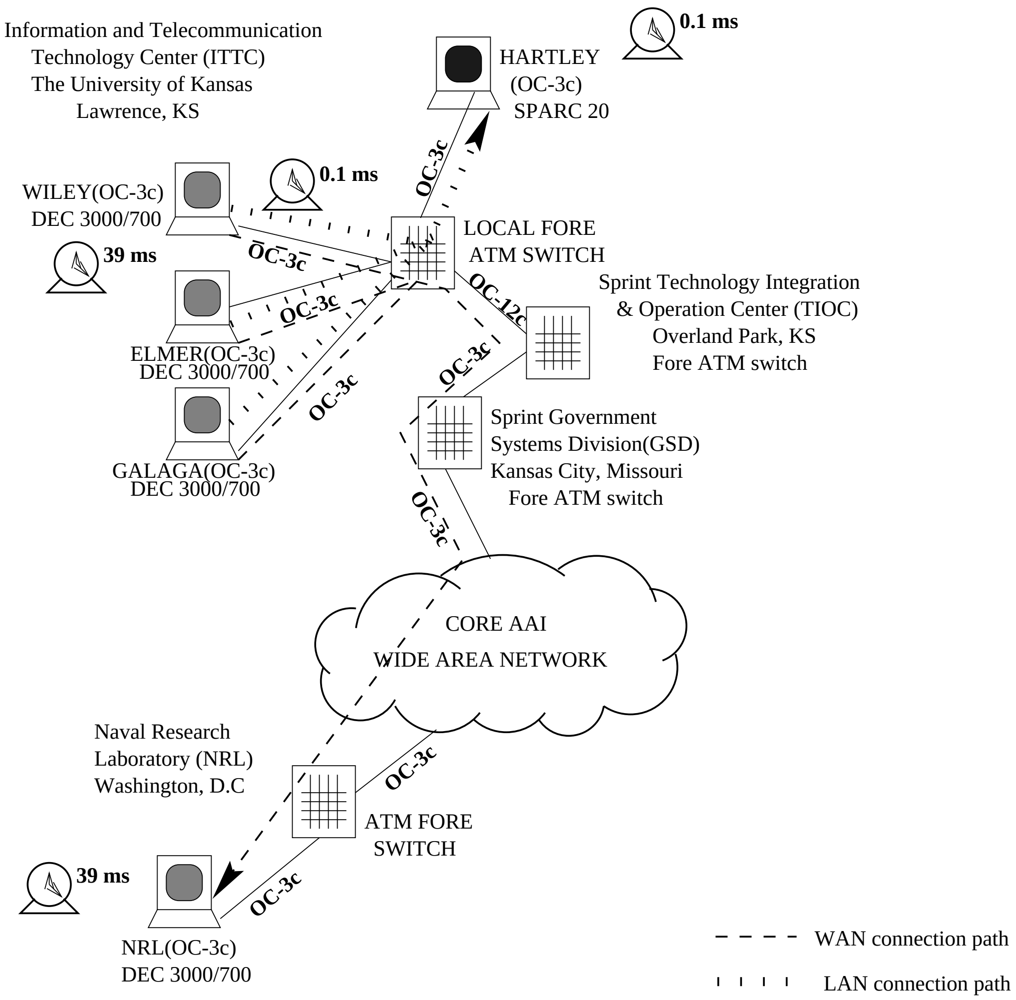

Figure 4: Diagram of scenario 5; TCP/IP over OC-3c ATM in the local ATM network (LAN) and the AAI WAN, under congestion conditions.

in the diagram transmit at a faster rate than the OC-3c link can handle. All the workstations are configured as in scenario 1.

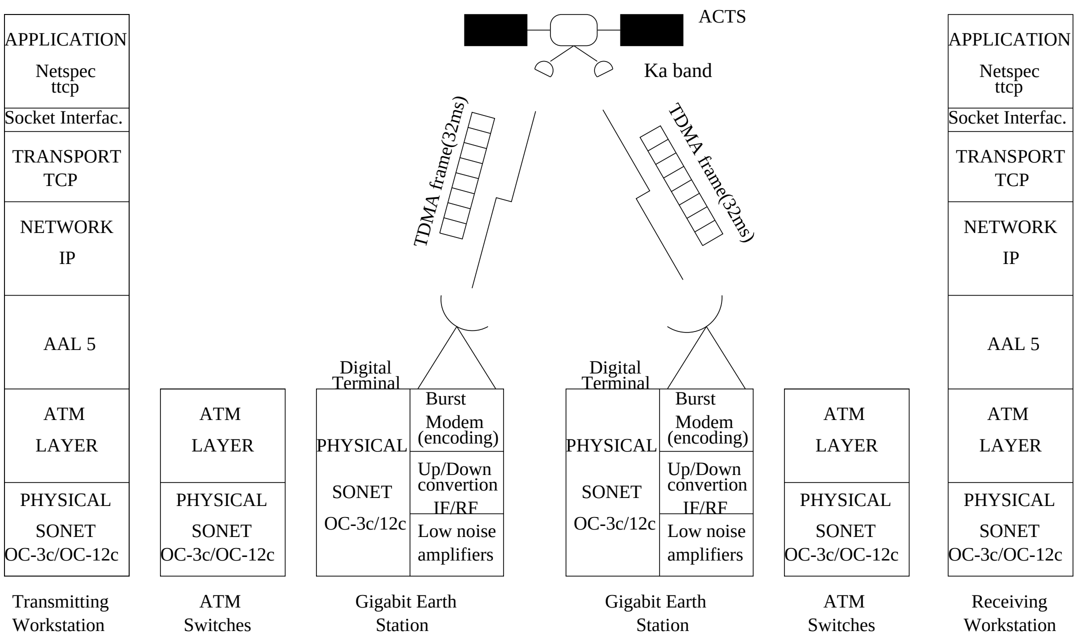

Figure 5: Diagram of the layered protocol architecture used throughout the experiments.

5.3 Performance of TCP/IP over ATM/SONET

The protocol layered implementation used in the experiments is shown in Figure 5. The ATM/SONET functions are provided by the Network Interface Cards (NIC) in each machine and by the ATM switch fabrics. The physical layer of our implementations is based on 155.52 Mbps SONET OC-3c and 622.08 Mbps SONET OC-12c interfaces. TCP/IP functions are provided by the TCP/IP programs implemented in the kernel of each machine. Classical IP over ATM is implemented according to RFC-1577 [17], where IP datagrams are encapsulated using IEEE 802.2 LLC/SNAP and segmented into ATM cells using AAL 5 (48 bytes). ATM layer adds another 5 bytes of header information. The default Maximum Transmission Unit (MTU) size for Classical IP over ATM networks is 9180 bytes [5], with a SNAP header of 8 bytes, [17, 12] and a maximum AAL 5 Protocol Data Unit (PDU) of 65535 bytes [26]. TCP and IP add another 20 bytes of header information each. Table 2 presents the theoretical expected rates on each layer for both OC-3c and OC-12c ATM/SONET systems.

	OC-3c ATM/SONET	OC-12c ATM/SONET
Line Rate	155.540 Mbps	622.080 Mbps
SONET Rate	149.760 Mbps	599.040 Mbps
AAL 5 Rate	135.632 Mbps	542.527 Mbps
IP Rate	135.102 Mbps	540.408 Mbps
Application Rate	134.513 Mbps	538.053 Mbps

Table 2: Theoretical maximum rates on each level of the network architecture used for OC-3 and OC-12 ATM/SONET systems.

5.4 Performance of TCP/IP on ATM/SONET over ACTS

The ACTS network, from the end-user point of view, was designed to replicate the functions of terrestrial SONET-based fiber networks. The GES performs the functions of SONET line terminating equipment (LTE), where the section and line overheads of the SONET frames are terminated locally at the GESs and the SONET SPE bytes are transported over the satellite [7]. Therefore, the throughput shown in Table 2 for SONET OC-3c and SONET OC-12c should be achievable over the satellite network.

6 Experimental Results

All the experimental scenarios were carried out with the default MTU size for Classical IP over ATM networks, which is 9180 bytes [5]. This configuration results in a TCP Maximum Segment Size (MSS) of 9140 bytes. This means that messages coming from the application level that are larger than the MSS will be fragmented into smaller packets equal to the MSS size. The most important parameter that had to be calculated for the experiments, using equation (2), was the upper limit of the TCP window or equivalently the send and receive socket buffer sizes. This parameter was passed to the operating system kernel via the *setsockopt* system level function by the application level network performance tools (Netspec, ttcp) that were used to test the network capacity.

Netspec and *ttcp* were the performance evaluation tools used throughout the experiments. TTCP was created at the US Army Ballistics Research Laboratory and can measure the throughput of a connection between two hosts using TCP or UDP (User Datagram Protocol). It is a useful and easy to use program. It is able to perform full rate experiments between two hosts only.

Netspec is a more sophisticated tool created by researchers at the University of Kansas [16]. It provides a block structured language for specifying network experiments and can support several connections (both TCP and UDP), as well as different types of traffic (CBR, VBR, MPEG, FTP) at full and constant rates.

6.1 Parameters and results from scenario 1

In this scenario, as shown in Figure 1, an OC-3 satellite link was established between the TIOC (Technology Integration and Operation Center) and GSFC. The BER in the satellite links was measured to be in the range of 10^{-11} at the TIOC and 10^{-13} at GSFC throughout the experiments. This is an almost error free OC-3 link, similar to the specifications of terrestrial fiber networks. The round trip time (RTT), that is the path latency between the two hosts, was measured by the program *ping* to be 534 ms on an average of ten trials. For comparison, let us note here that the average RTT in the local ATM network, shown in Figure 3, is only 0.1 ms, while the average RTT in the AAI WAN, shown in the same figure, is 39 ms. The satellite round trip latency is the sum of the propagation delay over the satellite link, the transmission delay of ATM cells, the ATM segmentation and reassembly (SAR) delay, the processing delay within the TCP/IP stack, and the propagation delay through the terrestrial fiber networks used to connect the hosts under test with the ground stations. Of these factors, the satellite propagation time is by far the dominant.

The send and receive buffer sizes to fully utilize the OC-3 ATM link and for optimal performance, were calculated according to equation (2). As shown below, the minimum window size required for maximum throughput is about 9 MB. In our experiments we used a transmit and receive window size of 10 MB.

$$Window\_Size = 534ms \times 134.513Mbps = 71.829942Mbits = 8.563MB$$

We obtained the performance results shown in Table 3 with a standard deviation (σ) of 5.5442 Mbps.

Trials	Throughput (Mbps)
1	100.973
2	105.270
3	105.986
4	106.124
5	106.550
6	111.600
7	112.923
8	114.068
9	114.679
10	119.000

Table 3: Results obtained from experimental scenario 1 in ten trials (placed here in ascending order).

6.2 Parameters and results from scenario 2

In this experimental scenario, as shown in Figure 1, an OC-12 satellite link was established between the TIOC and GSFC. At the transmitting end an OC-12c UltraSPARC workstation (faraday), and at the receiving end another OC-12c UltraSPARC workstation (mckinley) were used. The RTT is dominated by the speed of light delay, thus it was found to be the same as in scenario 1 (534 ms).

The transmit and receive socket buffer sizes were again calculated according to equation (2). As shown below, the minimum window size required for maximum throughput is about 34 MB.

$$Window\_Size = 534ms \times 538.053Mbps = 287.320Mbits = 34.251MB$$

Unfortunately, due to technical problems in the OC-12 digital terminal at the GSFC ground station at the time of the experiment, we managed to run only two successful tests with a 10 MB window size in the sender and receiver hosts. According to equation (2), the theoretical throughput that we should obtain with the 10 MB window size on an OC-12 link is 157.088 Mbps. Due to a relatively high BER (10^{-8} at TIOC and 10^{-9} at GSFC) reported on the satellite links, we observed 121.835 Mbps and 127.870 Mbps instead.

6.3 Parameters and results from scenario 3

When different OC-3c traffic sources are competing for the same OC-3c link, the offered traffic is too high for the switches to handle, resulting in cell overflow. As a result of this, cells will be dropped and TCP retransmissions will occur. A TCP packet includes about 192 ATM cells. One cell loss will result in the loss of the rest cells in that packet. Under these conditions, the goodput will be dramatically decreased.

In the case shown in Figure 2, the satellite network is tested under congestion conditions. Three OC-3c sources are transmitting over an OC-3 satellite link. This will result in congestion and the throughput will be decreased. Netspec with constant rate traffic was used to gradually increase the offered load, and measurements were taken for the aggregate throughput of the source machines. Also the available throughput was measured when the source machines were transmitting full rate traffic. The results from this experiment are shown in Figure 6, and compared with similar conditions over LANs and WANs.

6.4 Parameters and results from scenario 4

In this experimental scenario, as shown in Figure 3, measurements were taken in the local ATM network and in the AAI WAN on OC-3c connections under no congestion conditions, in order to compare these results with the ones obtained from the satellite network of scenario 1 and shown in Table 3. The hosts participating in the LAN and WAN environments were configured with the default IP over ATM settings, like the hosts in the satellite experiment of scenario 1, and the tests were run with a window size of 800 KB, which guarantees maximum throughput. Table 4 shows the average throughput obtained from the LAN, WAN and satellite network experiments. It is obvious from this table that the throughput results obtained in TCP/IP on ATM networks over terrestrial LANs and WANs, and satellite links with low BER, are very close regardless of the large difference in their network path latencies. The maximum throughput obtained in the LAN was 126.955 Mbps, in the WAN it was 126.109 Mbps, and in the ACTS environment it was 119 Mbps.

6.5 Parameters and results from scenario 5

In the experimental scenario shown in Figure 4, measurements were taken in the local ATM network and in the AAI WAN on OC-3c connections under congestion conditions in order to compare these results with the ones obtained from the satellite network of scenario 3. All the hosts participating in the LAN and WAN environments were configured as in scenario 3 discussed above. Figure 6 shows the offered load versus aggregate throughput obtained in the experiments under congestion conditions over

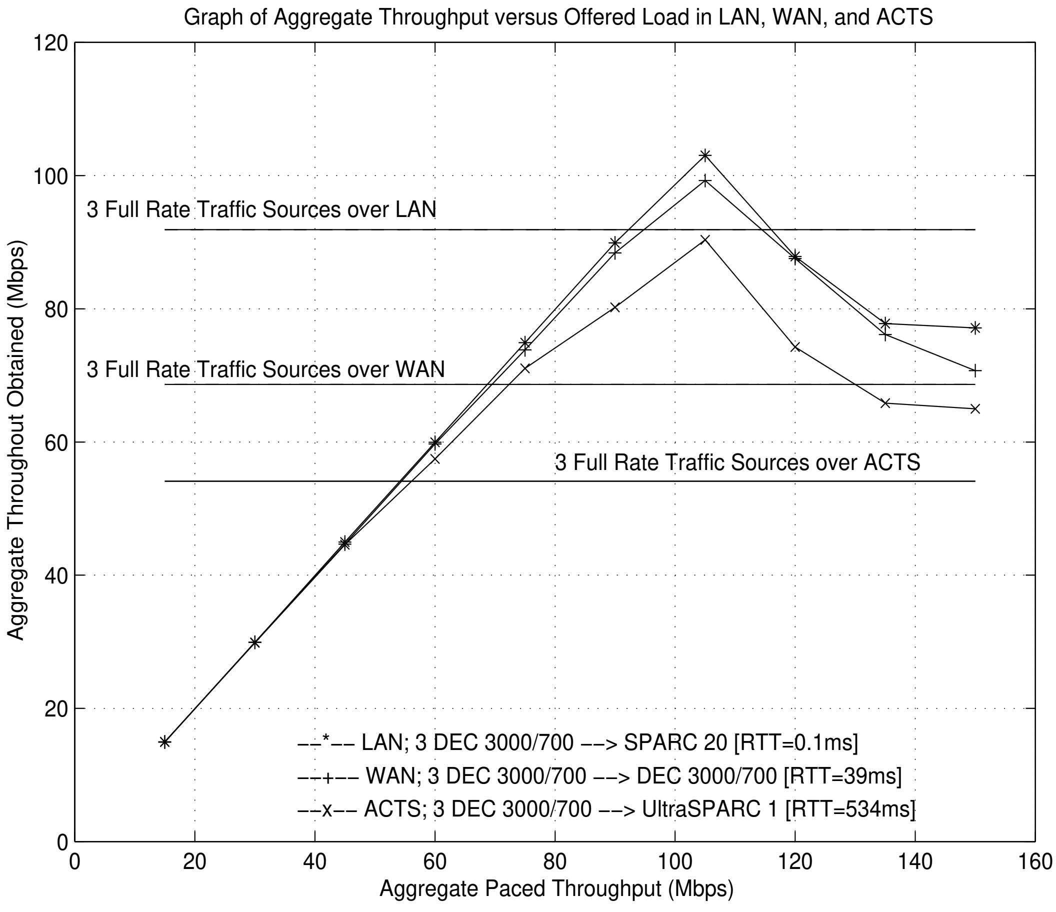

Figure 6: Graph illustrating throughput obtained versus offered load with three TCP connections over a LAN, WAN and a satellite environments.

LANs, WANs, and satellite environments. One can observe from these results that the maximum value of throughput in all environments under test was obtained when the aggregated offered load was 105 Mbps. At that load, the switch buffers overflowed, cell losses occured, and the throughput dropped sharply. It also continued to drop while we were increasing the offered load. The peak throughput under congestion for the LAN was 103.023 Mbps, for the WAN 99.265 Mbps, and for the satellite environment 90.342 Mbps. When all three sources transmit simultaneously as fast as they can (full rate), the obtainable throughput is much more lower, and it was measured as 91.887 Mbps, 68.646 Mbps, and 54.105 Mbps for the LAN, WAN and satellite environments, respectively.

	LAN	WAN	ACTS
Throughput (Average)	119.838 Mbps	114.058 Mbps	109.717 Mbps
σ	6.398 Mbps	7.079 Mbps	5.544 Mbps
RTT	0.1 ms	39 ms	534 ms

Table 4: Table showing the average throughput, standard deviation, and round trip time on OC-3c connections over a LAN, WAN, and ACTS environments in ten trials.

7 Analysis of results and TCP performance pitfalls

The results obtained from the experiments where no congestion conditions were present are a bit less than the theoretical ones shown in Table 2, for all three environments. This is due to the protocol implementations under different operating system kernels in the transmitting and receiving hosts, as well as due to the congestion algorithms implemented in the TCP protocol.

The slow-start and congestion avoidance algorithms in the TCP protocol have a negative effect on throughput, especially for long delay networks. Equation (2) shows that maximum throughput will be achieved if the send and receive maximum window sizes are set accordingly. This is of course true in case where the increase in the window size takes effect instantaneously and not gradually, as in the case of TCP with slow start and congestion avoidance mechanisms [6]. According to equation (1) and under no congestion or segment losses, with a window size of 10 MB as used in our satellite experiments (thus about 1143 MTUs of 9180 bytes in one window) and a RTT of 534 ms, it takes about 5.4 seconds ($10.2 \times RTT$) to fill the pipe. In our experiments, we were transmitting an average of 1 GB (about $102 \times Window\_Size$) of data per trial. Therefore, one window of data was transferred in 5.4 seconds, and after that each remaining window of data was transferred in one RTT (534 ms); so in our case we transferred 102 windows of data in about $10.2 \times RTT + 101 \times RTT = 111.2 \times RTT$ seconds (about 59.3 seconds). This is equivalent to $(102 \times Window\_Size)/(111.2 \times RTT) = 0.917 \times (Window\_Size/RTT)$, and means that there is an 8.3% reduction in throughput caused by the TCP slow-start mechanism in our set of satellite experiments even when no congestion or segment losses were present.

When there is a segment loss, TCP assumes that is caused by congestion [8, 23, 24], and therefore the transmitter has to reduce the rate of injecting data into the network. In the congestion avoidance phase, as discussed in Section 2, the CWND halves, and increases linearly (approximately one segment per RTT) until it reaches its original value. In the satellite experiments we ran, with 10 MB window size and default MTU sizes (9180 bytes), it will take 571.5 segments(5 MB) $\times$ RTT = 305.2 seconds to fill the pipe using this algorithm. In the fast recovery phase, as discussed in Section 2, the CWND halves (plus 3 segments due to the three duplicated ACKS received) and congestion avoidance follows. Thus, for our satellite experiments it will take (571.5-3) segments $\times$ RTT = 303.5 seconds to fill the pipe. TCP Reno with fast retransmit and fast recovery improves performance over the basic TCP implementation, but it exhibits another pitfall. Studies [4, 11] have shown that when more than one loss occurs within one window, fast retransmit and fast recovery will be triggered several times in one RTT, resulting in reduction of the CWND several times and then linear growth. This leads to throughput reduction.

Due to these TCP pitfalls, the satellite links must have a very low BER and no losses, otherwise the throughput will drop dramatically. If the satellite links are noisy with high BER and cell losses, then the probability of having more than one segment drop within one window is large, resulting in throughput degradation as discussed above. In the OC-3 experiments, the satellite links exhibited very low BER, comparable with those in fiber-optic terrestrial networks, therefore lost segments and retransmissions were limited, and throughput was comparable to that obtained over LANs and WANs. In the OC-12 experiments, the BER in the satellite links was relatively high, resulting in a degradation in throughput which can be justified using the same logic as above. Even if we have only one segment loss per window which is detected by fast retransmit, it will take 303.5 seconds for the TCP end host to utilize the available bandwidth in the satellite environment. Using the same logic, in the AAI WAN this time is reduced to

22.18 seconds and for the local ATM network to 0.057 seconds.

The experiments conducted under congestion conditions show that TCP congestion algorithms are not efficient for TCP over WAN and satellites, and result in throughput degradation in common cases. TCP end traffic sources competing for the same link will cause throughput to drop because of switch buffer overflow and cell losses. Traffic shaping is essential under these circumstances to prevent losses. Increasing the offered load above a certain level will cause the throughput to drop sharply. In the satellite environment, throughput drops faster than in the LAN or the WAN environments. This is because of the TCP mechanism, where bandwidth is wasted due to the long time needed to reach the receiver advertised window size after a segment loss occurs. For the same reason, when no traffic shaping is present, and all sources inject data in the network simultaneously as fast as they can, the aggregated throughput obtained over ACTS is lower than that obtained over the WAN or LAN.

8 Simulation of TCP over satellite networks

The creation of simulation models is usually the only means of predicting and evaluating the performance of high speed networks, since mathematical models of such networks are not yet feasible [18]. Simulation is an excellent way of investigating and understanding the behavior of the network architecture under test, as well as an efficient methodology for observing the effects of network parameter changes on overall performance.

In this section we investigate simulation models for TCP/IP end hosts over satellite environments and we use measurements obtained from the experimental scenarios 1, 2, and 3 to validate our simulation results. The simulation software used for our simulations is *BONeS DESIGNER* [25]. It is a software package for modeling and simulating event-driven systems, where a system model can be constructed hierarchically and graphically using building blocks from the BONeS model library, or using models written in C or C++ [18].

8.1 Simulation model primitives and parameters

The core of the simulation models developed is the TCP primitive module. We used the TCP primitive module developed by researchers in the University of Kansas, as described in [18], which is based on the 4.3 BSD Reno version and supports the functions covered in Section 2. In our simulation models, the network (IP) and physical (SONET) layers are not included since they are of little significance to the outcome of the results. The impact of these layers is captured by accounting their information overhead; the rates displayed in Table 2 were used. Also, due to the long run-time of the ATM segmentation and reassembly process in the simulated satellite environment, an ATM model was not used either in the simulation models of scenarios 1 and 2. In these models, we are actually simulating TCP over the satellite path and it is to be shown that the measurements obtained by experiments using TCP/IP on ATM/SONET networks over satellite can validate the simulation results of TCP over satellite models. In the simulation model of scenario 3, where congestion with lots of cell losses and TCP retransmissions is present, the ATM module is required for more accurate results.

The simulation system parameters used in our models are shown in Table 5. The

MTU size used is the default IP over ATM MTU size, which was used throughout the real experiments. TCP processing time is the overall time needed by the TCP module 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. The value of 62 μs shown in Table 5 is for a DEC 3000/700 machine with Digital Unix Operating System and it was provided by the Data Stream Driver Interface (DSKI), a tool developed at the University of Kansas to collect event traces and data from the operating system kernel [27]. The same value was assumed for the Sun Solaris machines used throughout the experiments. The TCP buffer size or window size used was 10 MB and line rates were used without the SONET overheads; thus a SONET module was not necessary in the simulation models. The retransmission timer is decremented every 0.5 seconds and retransmission occurs when the timer reaches zero (Slow-Timer period). A delayed ACK is sent every time the 0.2 second delayed ACK timer (Fast-Timer period) expires [18].

System Parameter	Value
TCP segment size	9140 bytes
MTU size	9180 bytes
TCP processing time	62 μs
TCP buffer size	10 MB
Slow-Timer period	0.5 s
Fast-Timer period	0.2 s
Minimum RTO	1.0 s
Switch processing delay	6.0 μs
Satellite propagation delay	125 ms per link
OC-3c Link speed	149.760 Mbps
OC-12c Link speed	599.040 Mbps

Table 5: Simulation parameters used in our models.

8.2 Simulation Results

- **Simulation of scenario 1, OC-3 satellite connectivity**

  The block diagram of the model simulating scenario 1 is shown in Figure 7 and the results are shown in Table 6. The data flows from end host wiley (KU) to end host mckinley (GSFC). The whole path through the satellite is bidirectional to allow ACKs from the receiver to the transmitter. The switches, High Data Rate Terminal (HDRT), and ACTS blocks are modeled by a FIFO (First In First Out) queue and a server. The link propagation delay is represented by the Link blocks (OC-3 or OC-12) and the OC-3 or OC-12 service times are included in the output port of every TCP end host and switch. The total path round trip delay through all the blocks is 534 ms, which is the round trip delay found in the real scenarios. The switch buffer sizes were set to a large value, since there is no congestion and thus their values are irrelevant to performance. The error blocks in the downlink paths of ACTS are random switches from BONeS library and allow data to pass through with a certain probability (passed as a parameter). This parameter was calculated by equation (3),

  $$P_{error\_free\_packets} = (1 - BER)^{(Packet\_Size)_{bits}} \quad (3)$$

  where $P_{error\_free\_packets}$ is the probability with which packets are passing through without any errors, BER is the bit error rate for each link as given by NMT during the experiments, and Packet_Size is the MTU used, 9180 bytes multiplied by 8 bits.

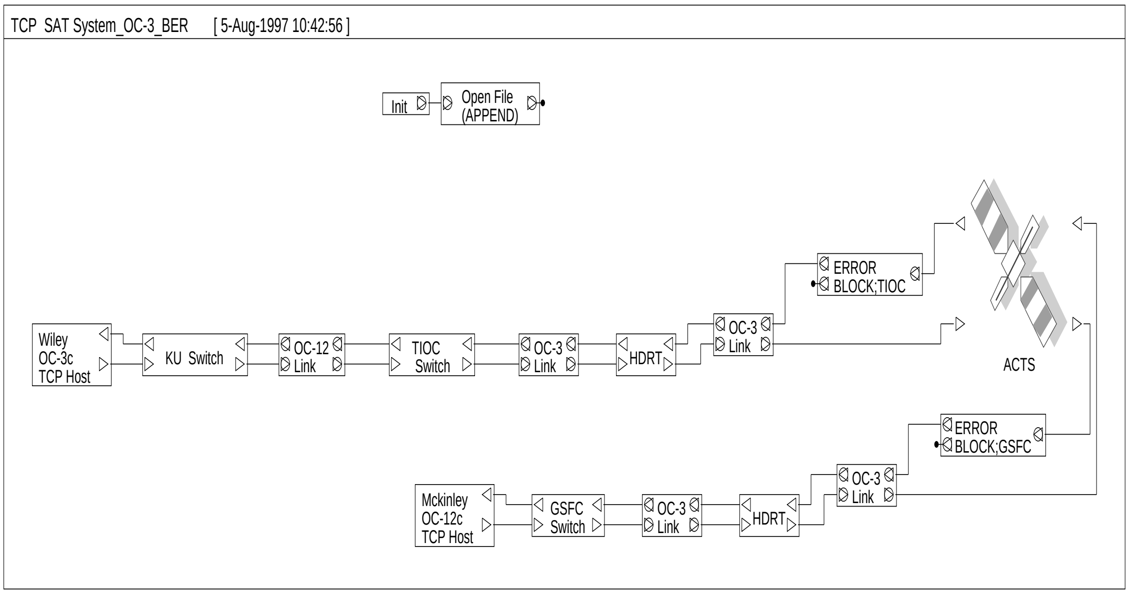

Figure 7: Simulation model for experimental scenario 1, OC-3 satellite connectivity.

- **Simulation of scenario 2, OC-12 satellite connectivity**

  The block diagram of the model simulating scenario 2 is shown in Figure 8 and the simulation results are shown in Table 6. The blocks and parameters are the same as in the simulation model of scenario 1, with the only differences being the OC-12 service times in all paths, and the appropriate probability value, calculated by equation (3), for the error blocks in the satellite links.

- **Simulation of scenario 3, OC-3 satellite connectivity under congestion conditions**

  The block diagram of the model simulating scenario 3 is shown in Figure 9 and the simulation results are shown in Table 6. This model represents an OC-3 ACTS bent pipe loop-back connection between three TCP/ATM traffic source blocks (no

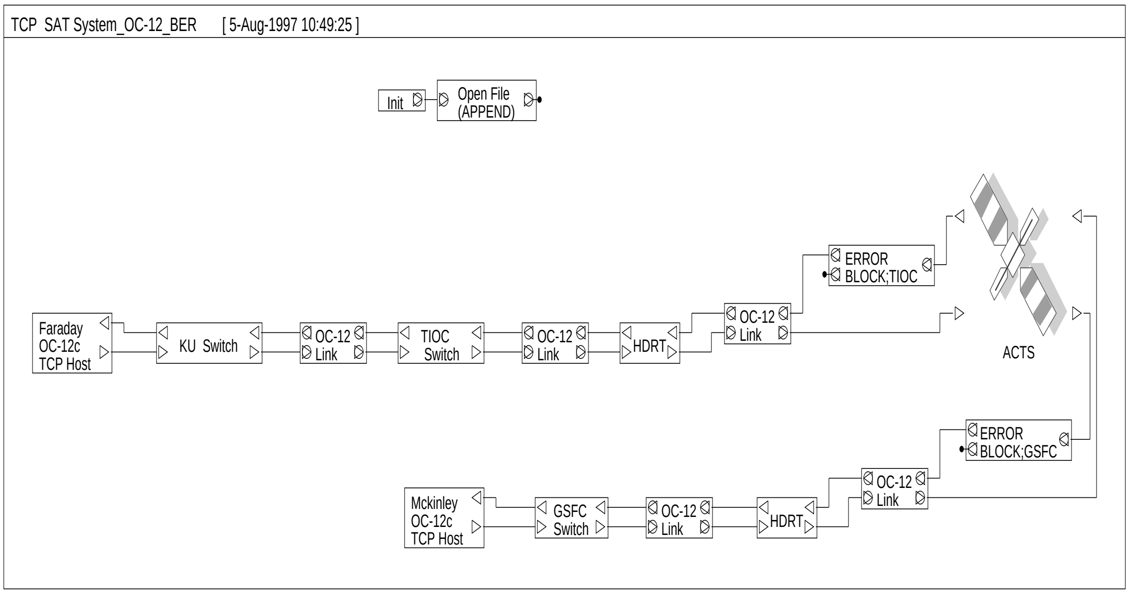

Figure 8: Simulation model for experimental scenario 2, OC-12 satellite connectivity.

traffic shaping is present) and a block accommodating three TCP/ATM hosts in order to be able to accept data from them. The connection paths between the source and receiver blocks are established through the "KU ATM Switch" block using connection identifiers, which are equivalent to the virtual circuit identifiers (VCI) used in the experiments, and treated at the receiving end with the use of a FIFO queue with a server which allows TCP packets to pass through towards the specified receiver TCP/ATM host at specific time intervals, specified by the TCP processing time. The probability with which cells are passing through is calculated for this model by equation 4, and is passed as a parameter in the error block located on the downlink satellite path,

$$P_{error\_free\_cells} = (1 - BER)^{(Cell\_Size)_{bits}} \tag{4}$$

where $P_{error\_free\_cells}$ is the probability with which cells are passing through without any errors, BER is the bit error rate for each link as given by NMT during the experiments and Cell_Size is the ATM cell size in bits, which is 424 bits.

The rest of the blocks are the same as in the simulation model of scenario 1, with the only differences being the processing of ATM cells instead of TCP packets, and the switch buffer size, which was set to 8192 cells per virtual circuit (VC) to model the maximum possible buffer space a VC can have in the FORE ATM switches. However note that the UBR buffer space in the FORE ATM switches is allocated per VC on an as needed basis [22]. For the simulation purposes, setting the switch buffer space to the maximum value of 8192 cells per VC is a close approximation.

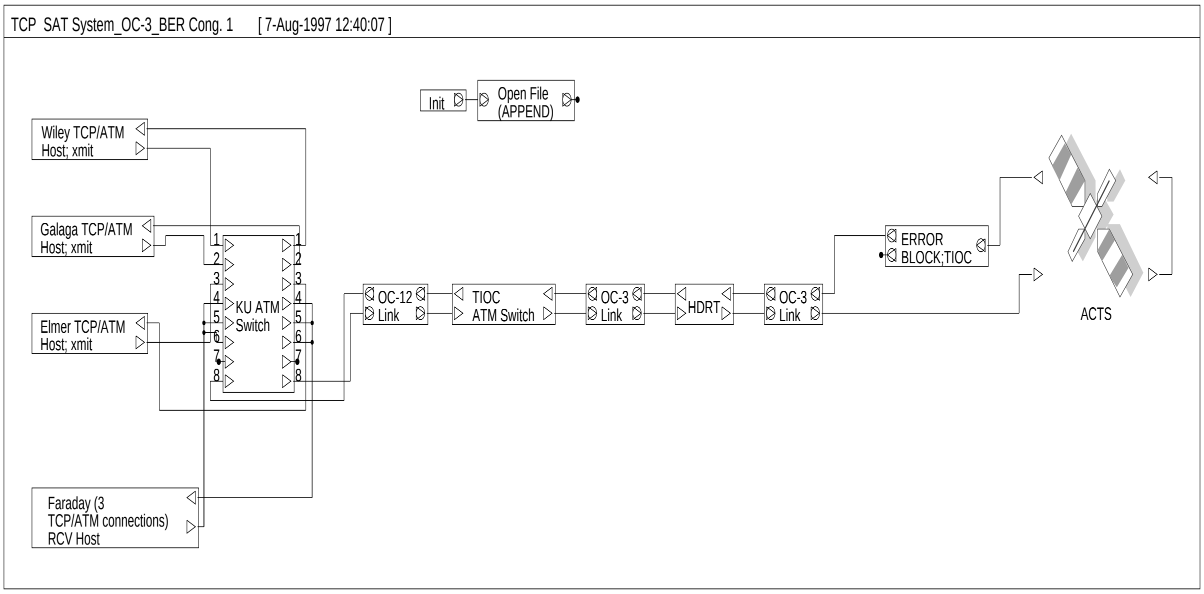

Figure 9: Simulation model for experimental scenario 3, OC-3 satellite connectivity under congestion conditions.

	BER	Experimental Results (Max.Values)	Simulation Results (Max.Values)	% Error
Scenario 1 (OC-3 link) (No ATM simulated)	TIOC/GSFC $10^{-11}/10^{-13}$	119.000Mbps	115.295Mbps	3.1%
Scenario 2 (OC-12 link) (No ATM simulated)	TIOC/GSFC $10^{-8}/10^{-9}$	127.870Mbps	131.360Mbps	2.6%
Scenario 3 (full rate) (OC-3 congestion) (ATM simulated)	TIOC 10^{-11}	54.105Mbps	49.641 Mbps	8.2%

Table 6: Table showing the experimental and simulation results (maximum values obtained over trials) for scenarios 1, 2, and 3, as well as the BER on the satellite links during the experiments.

9 Conclusions

In this paper we studied the performance of TCP/IP end hosts on ATM networks over high speed satellite links. We also compared throughput obtained from experimental scenarios over the local ATM network, over the AAI WAN, and over ACTS. We did the same using simulation and validated the simulation results by the experimental measurements. The basic results of our study indicate:

- Throughput results for TCP/IP hosts on ATM/SONET networks over LANs, WANs and satellite environments with very low BER and high speed channels (like ACTS) are similar to each other regardless of the large differences in path latencies they exhibit. Of course this is achievable if TCP supports the necessary extensions, with the most important being the extension of scaling window sizes beyond 64 KB. Even if we assume that the communication channels are ideal, the throughput will be lower than the theoretical values because of the TCP slow-start algorithm.

- In cases where noisy high speed satellite links are established, throughput obtained by TCP Reno end systems will be degraded, because of retransmissions and the fact that bandwidth will be wasted while the source tries to reach the receiver's advertised window size using the fast recovery or congestion avoidance algorithms. A possible solution to this that might increase the performance obtained by TCP over noisy satellite links is to modify TCP to handle retransmissions in a more efficient way. Simulation studies [11, 4], have shown that the implementation of a selective acknowledgment (SACK) TCP protocol can improve the performance over TCP Reno.

- When traffic sources are competing for the same link and no traffic shaping is used, throughput drops dramatically. In this case, cells are dropped due to overflowed switch buffers and thus TCP retransmissions occur. This is much worse over the satellite environment, since TCP Reno will decrease the window size when multiple segment drops occur within one window, and fast recovery will not be fast enough due to the time needed for the sender to ramp up and reach the receiver's window size.

- Traffic shaping will help to achieve higher peak values of throughput compared with unregulated transmissions from the traffic sources under congestion conditions.

- Using experimental measurements we validated the simulation results for models representing scenarios 1, 2 and 3 with a small error. That means, simulation is an excellent way of predicting the performance of communication networks.

Further performance experiments must be carried out to obtaining a better understanding of the OC-12 connectivity through ACTS. Also new TCP implementations must be applied in order to investigate how efficiently they can handle retransmissions caused by congestion or segment losses on noisy high speed satellite channels.

References

[1] Mark Allman. Improving TCP Performance over Satellite Channels, *Master of Science Thesis*, Ohio University, 1997.

[2] Mark Allman, Chris Hayes, Hans Kruse, Shawn Ostermann. TCP Performance over Satellite Links. *Proceedings of the 5th International Conference on Telecommunication Systems, March 1997.*

[3] I. Andrikopoulos, T. Örs, M. Matijasevic, H. Leitold, S.P.W. Jones, R. Porsch, TCP/IP Throughput Performance Evaluation for ATM Local Area Networks. *Proceedings of the IFIP TC6 "Fourth Workshop on Performance Modeling and Evaluation of ATM Networks", July 1996*

[4] Furquan A. Ansari. Adapting TCP/IP over ATM, *Master of Science Thesis*, University of Kansas, 1996.

[5] R. Atkinson. Default IP MTU for use over ATM AAL5, May 1994. RFC 1626.

[6] S.M. Bajaj, C. Brazdziunas, D.E. Brooks, D.F. Daly, S.M. Srinidhi, T. Robe, F. Vakil. Performance Characterization of TCP/IP-On-ATM over an ATM/SONET High Data Rate ACTS Channel. *Proceedings of the 16th AIAA International Communications Satellite Systems Conference, February 1996.*

[7] M.A, Bergamo. Network Architecture and SONET Services in the NASA/ARPA Gigabit Satellite using NASA's Advanced Communications Technology Satellite (ACTS). *Proceedings of the 15th AIAA International Communications Satellite Systems Conference, February 1994.*

[8] Douglas E. Comer. *Internetworking with TCP/IP, Volume I, Principles, Protocols, and Architecture.* Prentice Hall, 3rd edition, 1995.

[9] L. DaSilva, J. B. Evans, D. Niehaus, V. S. Frost, R. Jonkman, B. Lee, G. Lazarou. ATM WAN Performance Tools, Experiments, and Results. Accepted for publication in IEEE Communications Magazine, expected publication date August 1997.

[10] C.E. Fair. TCP Performance over ACTS. *Proceedings of the 1996 IEEE Transport Protocols for High-Speed Broadband Networks workshop*, IEEE Globecom '96 Workshop.

[11] Kevin Fall, Sally Floyd. Simulation-based Comparisons of Tahoe, Reno, and SACK TCP. *Computer Communications Review, July 1996.*

[12] Juha Heinanen. Multiprotocol Encapsulation over ATM Adaptation Layer 5, July 1993. RFC 1483.

[13] D. Hoder, B. Kearny. Design and Performance of the ACTS Gigabit Satellite Network High Data-Rate Ground Station. *Proceedings of the 16th AIAA International Communications Satellite Systems Conference, February 1996.*

[14] Van Jacobson, Robert Braden. TCP Extensions for Long-Delay Paths, October 1988. RFC 1072.

[15] Van Jacobson, Robert Braden, David Borman. TCP Extensions for High Performance, May 1992. RFC 1323.

[16] Roelof Jonkman. Netspec: A Network Performance Evaluation and Experimentation Tool.
Available at: http://www.ittc.ukans.edu/netspec.

[17] M. Laubach. Classical IP an ARP over ATM, January 1994. RFC 1577.

[18] Georgios Y. Lazarou, Victor S. Frost, Joseph B. Evans, Douglas Niehaus. Using Measurements to Validate Simulation Models of TCP/IP over High Speed ATM Wide Area Networks. In *Proceedings of the 1996 IEEE International Conference on Communications, June 1996.*

[19] Matthew Mathis, Jamshid Mahdavi, Sally Floyd, Allyn Romanow. TCP Selective Acknowledgment Options, October 1996. RFC 2018.

[20] J. Mogul, S. Deering. Path MTU Discovery, November 1990. RFC 1191.

[21] NASA System Handbook. *Advanced Communications Technology Satellite*, NASA, Cleveland, 1995.

[22] Vector General Description Guide. Northern Telecom, Canada, 1996.

[23] Jon Postel. Transmission Control Protocol, September 1981. RFC 793.

[24] W. Richard Stevens. TCP Slow Start, Congestion Avoidance, Fast Retransmit, and Fast Recovery Algorithms, January 1997. RFC 2001.

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

[26] Andrew S. Tanenbaum. *Computer Networks, 3rd edition*, Prentice Hall PTR, New Jersey, 1996.

[27] Yulia I. Wijata. Netspec Data Stream Daemon.
Available at: http://www.ittc.ukans.edu/~ywijata/projects/nsdstrd.