Interworking of TCAP/SS7 and TCP/IP Networks

by

Jingsheng Qin

A Thesis presented to the University of Manitoba in

partial fulfillment of the requirements for the degree of

Master of Science

in the

Department of Electrical and Computer Engineering

Winnipeg, Manitoba

August 2001

0-612-76853-8

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION

INTERWORKING OF TCAP/SS7 AND TCP/IP NETWORKS

BY

JINGSHENG QIN

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of

Manitoba in partial fulfillment of the requirement of the degree

of

MASTER OF SCIENCE

Abstract

This thesis addresses the SS7 and TCP/IP interworking area, a popular research topic in the telecom industry for building a seamless network between the circuit-switched and packet data networks. This thesis is specifically focused on the TCAP in SS7 interworking with UDP/IP. The approach proposed in the thesis employs a signaling gateway to work on the edge of the two different networks in order to transfer the SS7 messages across UDP/IP network. The details of implementations for a signaling gateway to perform the protocol conversion are provided. The thesis also discusses how the management and control functions are implemented by mapping the ICMP messages of TCP/IP and the SCCP Unitdata Service messages of SS7. Examples for the interworking are given to demonstrate the solutions and products, which some telecom companies have developed.

Based on the interworking system proposed in this thesis, the performance parameters are simulated with OPNET, where different arrival rates of the traffic towards the signaling gateway are presumed. The simulation results are given in the graphs of CDF (cumulative distribution function), which show that most of the measured performance parameters meet the performance requirements as described in SS7 specifications.

Acknowledgements

I wish to thank my supervisor Dr. Pawlak, who has been contributing a lot of his time supervising my graduate study and research and discussing with me the issues addressed in this thesis.

Also I wish to thank TR*Labs* for providing me with financial support for two years, which makes the completion of this thesis possible. I wish to thank Dr. Rueda in TR*Labs* for proposing many constructive ideas to my research work that this thesis is based on.

Finally, I wish to thank my wife and my parents for giving me the warm family supports throughout my graduate study and my preparation for this thesis. Without their supports, it would have taken me much longer to finish this thesis.

Acronyms

AAL	ATM Adaptation Layer
AIN	Advanced Intelligent Network
ASE	Application Service Element
ATM	Asynchronous Transfer Mode
CDF	Cumulative Distribution Function
CPE	Customer Premises Equipment
DAS	Dial Access Solution
DPC	Destination Point Code
GT	Global Title
GTI	Global Title Indicator
ICMP	Internet Control Message Protocol
IP	Internet Protocol
ISP	Internet Service Provider
ISUP	ISDN User Part
MSU	Message Signal Unit
MTP1	Message Transfer Part Level 1
MTP2	Message Transfer Part Level 2
MTP3	Message Transfer Part Level 3
MTU	Maximum Transfer Unit
NAS	Network Access Server
OAM	Operations, Administration, Maintenance
OPC	Origination Point Code

PC	Point Code
PCI	Point Code Indicator
POP	Point of Presence
PSTN	Public Switched Telephone Network
QOS	Quality of Service
SCCP	Signaling Connection Control Part
SCP	Service Control Point
SCTP	Simple Control Transport Protocol
SG	Signaling Gateway
SL	Signaling Link
SLS	Signaling Link Selector
SNI	Subsystem Number Indicator
SONET	Synchronous Optical Network
SS7	Signaling System No. 7
SSN	Subsystem Number
SSP	Service Switching Point
STP	Signaling Transfer Point
SU	Signal Unit
TC	Transaction Capabilities
TCAP	Transaction Capabilities Application Part
TCP	Transmission Control Protocol
TDM	Time Division Multiplexing
UDP	User Datagram Protocol

UNI	User Network Interface
VCI	Virtual Circuit Identifier
VPI	Virtual Path Identifier
WAN	Wide Area Network

Table of Contents

List of Figures

List of Tables

List of Graphs

Chapter 1. Purposes of the Interworking & Objectives of the Thesis

1.1 Summary

In this chapter, the objectives of this thesis and the purposes of the interworking of SS7 and TCP/IP networks are provided. An overall description of TCAP and SS7 will also be provided as an entry point to understand interworking theory.

1.2 Why Interworking?

The SS7 or PSTN network has been a dominant communication network for many years. The investments in it are huge because a single T1/E1 line usually costs thousands of dollars. With the internet becoming more and more popular, attention as to how to utilize the internet to perform similar functions of the SS7 or PSTN networks has increased. The major motivation is to save the big costs of the infrastructure of SS7 or PSTN. For example, with the internet telephony technology, voice can be transferred via internet or TCP/IP network. Nowadays, it has become a trend in the telecommunications industry to converge the data and voice networks into a smooth and seamless network at a low cost. How to employ the existing protocols for the convergence of the voice and data networks has been studied for a few years by many R&D researchers and engineers. In spite of the high complexity and the many different aspects of the two protocols, a lot of issues have been addressed, and many constructive proposals have come out. Major companies in telecommunications industry, such as Lucent, Nortel, Cisco, etc., have been trying very hard to develop the protocols and the products, and have made considerable progress.

1.3 Objectives of the Thesis

One of the objectives of this thesis is to provide a solution to the interworking issues mentioned in Section 1.2. The solution proposed has three features:

a. It utilizes the existing TCAP/SS7 and UDP/IP infrastructures, and does not introduce any new protocols in order to achieve cost-saving.

b. The interface equipment working on the edge of the two networks are able to transfer a TCAP/SS7 message across TCP/IP network successfully to the destination nodes as defined in the routing label of the message.

c. The management and control functions are able to control the conditions of a remote interface equipment. The mechanisms to handle the abnormalities which occur at a remote interface equipment are provided.

The interface equipment in the proposed interworking system is called "signaling gateway", which is used throughout the thesis.

Another objective of this thesis is to provide simulation results based on the solution proposed for some important performance parameters as defined in SS7 specifications. These results can provide good references for the readers to better understand the interworking performance issues. In the following, some basic concepts and functions of SS7 are introduced to give readers some ideas of the TCAP and SS7.

1.4 What is SS7?

SS7, signaling system number 7, was originally designed to support the PSTN (public switched telephone network) in the call setup, management, and release between telephone offices and CPE (customer premises equipment) to simply transport voice traffic. But now it has been further developed for the establishment of connections between the service provider offices and CPE to transport not only voice but also data/video traffic [Blac97]. In the original SS7, the "in-band" signaling mechanism was used, which means that the signaling information and the user data (e.g. voice) are carried on the same channel. Nowadays, many new applications of SS7 have emerged, such as SS7 intelligent network. Because these applications require a high quality transfer of the signaling information, the signaling in SS7 has changed to the "out-of-band" signaling. In the "out-of-band" signaling, the signaling information is carried in a separate and dedicated physical channel, and the user data is carried in other channels, as shown in Figures 1.1 & 1.2.

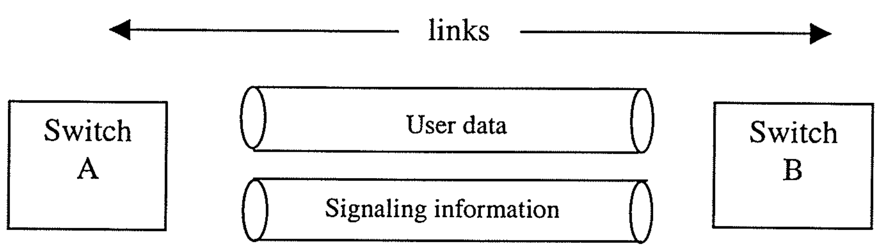


Figure 1.1: Out-of-band Signaling

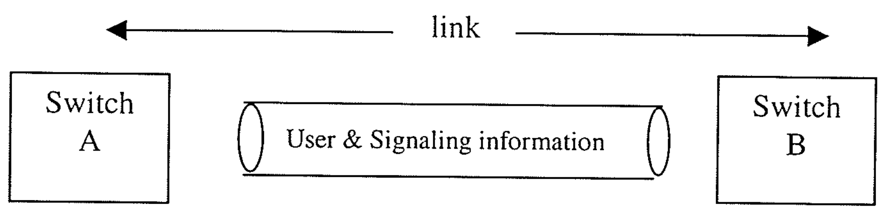


Figure 1.2: In-band Signaling

As shown in Figure 1.3, the SS7 protocol stack consists of many different layers. In the following, a brief introduction to each of the layers that are involved in this thesis is given.

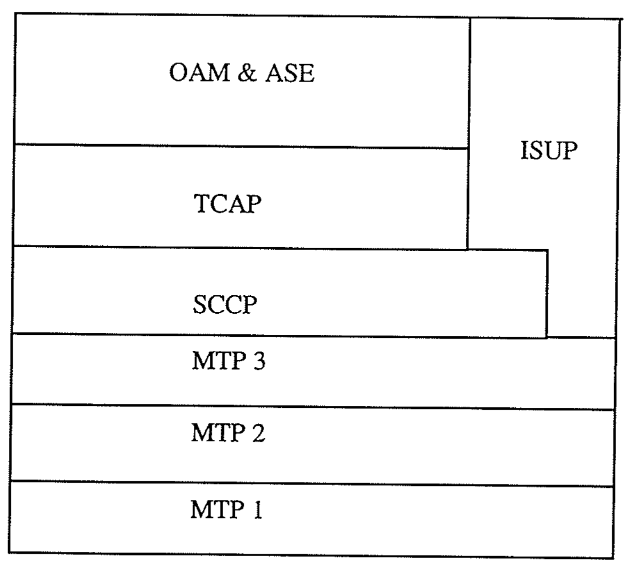


Figure 1.3: SS7 Protocol Layers

The MTP1 (message transfer part level 1) is the physical layer, providing a transport for an individual signaling data link. It defines the physical, electrical, and functional characteristics of a signaling data link and the means to access it. In a digital environment, 56 or 64 kbps digital paths are used for the signaling data link [Ansm92].

The MTP2 (message transfer part level 2) is the data link layer, which provides the functions and procedures to transfer signaling messages over an individual signaling data link. It is primarily responsible for error free transmission of data for some basic communication functions. Once information enters the network, it must be transferred in the proper sequence and without error between network nodes over each segment of the transmission path [Ansm92].

The MTP3 (message transfer part level 3) is the network layer, providing functions and procedures that are common to (but independent from) the operation of individual signaling links. It provides services that transmit the data through the network from originator to its final destination. The message routing is implemented at this layer through the routing labels maintained in MTP3 signal units and at a specific signaling point. The OPC (origination point code) or DPC (destination point code) in a routing label consists of three bytes which are used to identify the network, network cluster and network cluster member. The SLS (signaling link selector) in a routing label is used to balance the traffic load among signaling links. MTP3 also provides the means to establish, maintain, and terminate connections between systems, such as signaling link failure detection and link recovery [Ansm92].

On the top of MTP3 is the SCCP (signaling connection control part), which is a combination of parts of the network layer and parts of the transport layer. It relies on the MTP level 1, 2, 3 to provide additional functions for both connectionless and connection-oriented network services to transfer circuit-related and non-circuit-related signaling information between switches or exchange centers [Anss92]. When coupled with the MTP layers, SCCP provides specialized routing and management and control functions for the transfer of higher level messages between the MTP layers and the SCCP users, for example, TCAP (Transaction Capabilities Application Part).

There are four classes of service in SCCP, two for connectionless services and the other two for connection-oriented services [Anss92]:

a) basic connectionless class (class 0);

b) sequenced connectionless class (class 1);

c) basic connection-oriented class (class 2);

d) flow control connection-oriented class (class 3).

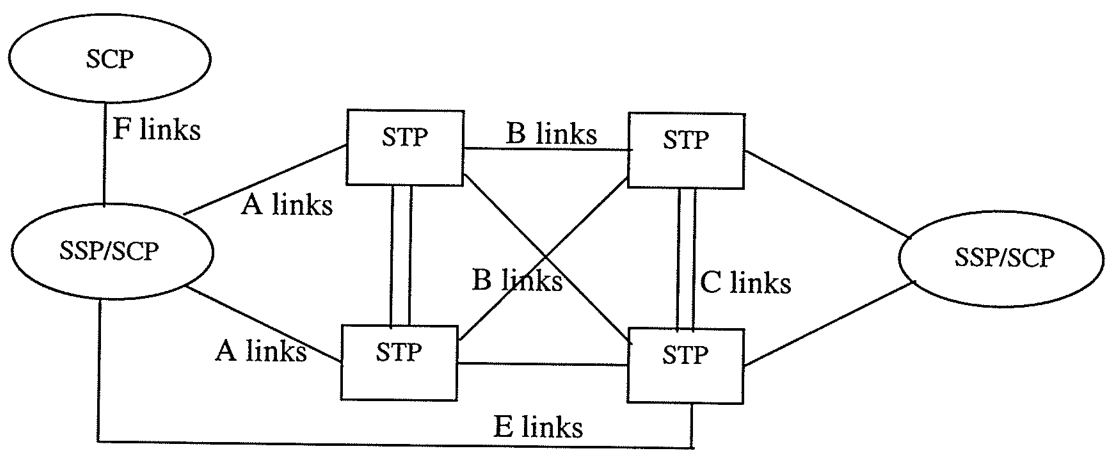


Figure 1.4: SS7 Network Topology

Figure 1.4 shows a typical topology of an SS7 network, where the major physical components include SSP (service switching point), STP (signaling transfer point), SCP (service control point) and SL (signaling link) [Blac97].

SSP is the local exchange to the subscriber and the interface to the telephone network. The SS7 signaling information is produced at the sending SSP and is handled at the receiving SSP. It converts voice signaling into the SS7 signal units, and vice versa. It also supports database access queries, such as for 1-800 service.

STP is the network node. It works like a router to perform the routing of messages to other signaling points. It is usually paired to provide redundancy for a reliable message transfer. It is adjunct to a voice switch, and might stand alone as a separate machine.

SCP works as the interface with the databases, which contain the information of the subscribers, such as 1-800 service, calling cards, fraud data, etc. When receiving a request, it is triggered to handle the database queries, and returns a response with the results of the queries to the originating SSP.

SL is the link interconnecting the signaling points of SS7 network. A SL is made up of digital transmission channels, and in each channel, the standard bit rate on a bearer service is 56 or 64 kbps. But these channels may be multiplexed into a transmission rate of 1.544, 2.048, or 8.448 Mbps. All the links between two signaling points (e.g. SSP, STP, SCP) compose a "link set". The network traffic between two signaling points are evenly distributed to reach a load sharing within a link set, and these different link sets can provide the redundancy to ensure the high reliability of message transfer.

There are six types of SL: A, B, C, E, F links which are shown in Figure 1.4, and D link which is not shown in Figure 1.4 (D links connect the primary level STP pairs with secondary level STP pairs).

1.5 What is TCAP?

The TCAP, transaction capabilities application part, provides application level functions for special SS7 services. Service information exchanged between the SSP and the network database would typically be defined within the TCAP. The TCAP protocol layer resides between the application layer (e.g. a 1-800 service), and the SCCP layer. An ASE (application service element) at a SSP representing an application uses the TCAP to initiate a transaction by sending a query message to a remote SCP, while an ASE at the SCP uses the TCAP to end the transaction by returning a TCAP response message to the originating SSP, which carries the result of the query [Anst92].

The TCAP are supported by SCCP, and the TCAP messages are encapsulated into the SCCP header part within an MSU (message signal unit) for its transfer. As shown in Figure 1.5 [Blac97], the TCAP layer is composed of two sublayers: the Transaction sublayer, and the Component sublayer. A TCAP message is composed of a "transaction portion" and a "component portion".

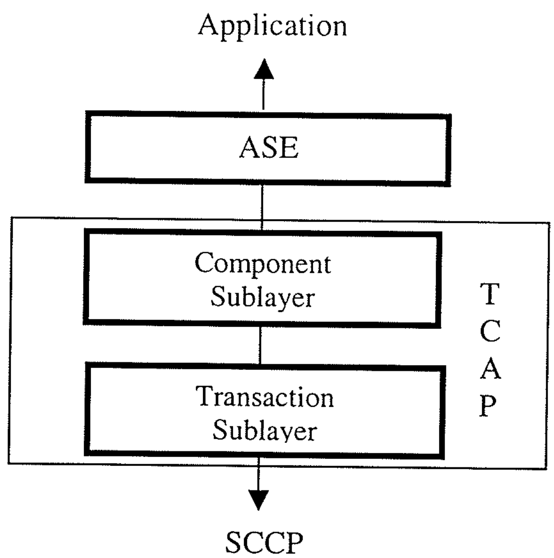


Figure 1.5: TCAP Layer Structure

The transaction portion contains the Package Type Identifier. There are seven package types [Anst92]:

- Unidirectional: transfers component(s) in one direction only (no reply expected).
- Query with Permission: initiates a TCAP transaction (e.g. a 1-800 call service). The destination node may end the transaction.
- Query without Permission: initiates a TCAP transaction. The destination node may not end the transaction.
- Response: ends the TCAP transaction. A response to query with permission may contain the information requested.
- Conversation with Permission: continues a TCAP transaction. The destination node may end the transaction.

- Conversation without Permission: continues a TCAP transaction. The destination node may *not* end the transaction.
- Abort: terminates a transaction due to an abnormal situation.

The transaction portion also contains the Originating Transaction ID and Responding Transaction ID fields which can associate the TCAP transaction with a specific application at the origination and destination signaling points respectively.

The component portion contains six kinds of components [Anst92]:

- Invoke (Last): invokes an operation. The component is the "last" component in the query.
- Invoke (Not Last): similar to the Invoke (Last) component except that the component is followed by one or more components.
- Return Result (Last): returns the result of an invoked operation. The component is the "last" component in the response.
- Return Result (Not Last): similar to the Return Result (Last) component except that the component is followed by one or more components.
- Return Error: reports the unsuccessful completion of an invoked operation.
- Reject: indicates that an incorrect package type or component was received.

Components also contain additional parameters for application-specific data unexamined by TCAP.

Chapter 2. An Example of Convergence

2.1 Summary

The well-known "voice over IP" is a typical application example of converging SS7 and IP networks. In this chapter, we will give another example of convergence: internet dial-up problem.

2.2 Internet Dial-up Problem

The PSTN, designed to handle voice traffic with an average holding time of three to five minutes, is being bombarded with dial-up access data traffic with 20-minute average holding times. The inevitable consequence is that the internet access via dial-up modem causes congestion problems within the PSTN. The following presents the solutions to this problem as provided by some telecom companies.

2.3 Telcordia's Solution

In a white paper from Telcordia (former Bellcore) [Atai97], five architectures are recommended to extract the traffic off the PSTN and direct them into the Internet, three of them are post-switch architecture and the other two are pre-switch architecture.

In a post-switch architecture, the internet calls pass through the ingress switch before being redirected off the PSTN and onto a packet network for final delivery to an ISP POP. Its main advantage is that the internet calls bypass the PSTN's interoffice trunks and egress switches. But in such an architecture, the ingress switches are still involved.

In a pre-switch architecture, the involvement of PSTN is minimal compared with the post-switch architectures because the internet calls are intercepted and redirected onto a packet network before they enter the ingress switch. The advantage of such a structure is that the internet calls bypass almost all the PSTN elements (ingress switches, trunks, and egress switches).

Some of the above architectures must be supported by the intelligent network capabilities of SS7 for signaling and routing the internet calls.

2.4 3Com's Solution

3Com has proposed several solutions for this dial-up problem [3com98]. The first solution provides the access concentrator with the ability to interface directly to SS7. The advantage of this approach is that it keeps all the functionality of the SS7/IP integration contained within a single device, making it the most manageable. But this approach has limitations for scalability.

The second solution is to give several access concentrators an SS7 connection by using an external converter to handle the translation of the incoming traffic into SS7 signaling. But this approach still does not overcome completely the limitations for scalability.

The third solution is to use SS7 signaling gateway. The gateway works as a bridge between the existing PSTN and IP networks, translating the signaling information between the two incompatible network types. Unlike the above two solutions, gateways provide added intelligence for security and control. The other advantage of gateways over other solutions is the tremendous scalability and cost-efficiencies that gateways can provide in larger networks, since a gateway can interface to multiple access concentrators on the IP side and share a single connection on the SS7 side.

2.5 Lucent's Solution

Lucent has built the SS7 Signaling Gateway solution, which enables service providers to cost-effectively offload internet traffic from PSTN and deliver this traffic directly into broadband data networks [Luce98]; in other words, this solution gives service providers the immediate ability to handle dial-up internet calls without costly investment in additional voice circuits.

Based on this solution, Lucent has developed two products: Ascend SS7 Gateway and the MAX TNT WAN access switch. The strengths of the two products are combined to deliver the solution: the SS7 Gateway enables data calls to be taken from the originating central office and the MAX TNT provides the means for internet calls to immediately be placed on a data network. The SS7 Gateway connects to the SS7 network via multiple A links and to the data network via multiple Ethernet connections. The MAX TNT connects directly to the originating ingress switch, bypassing most of the PSTN. The SS7 Gateway receives SS7 signals for call setup, circuit establishment, and call release from the originating central office; converts these signals to enhanced Q.931 messages; and relays these messages to MAX TNT to further handle the call. The MAX TNT then uses this information to establish a connection with the originating switch, thus freeing up the PSTN.

2.6 Nortel's Solution

Nortel Networks has developed the CVX SS7 Gateway and CVX 1800 dial access switch to assist service providers and telecommunications carriers to keep up with the accelerated demand for internet access [Nort98]. The Gateway supplements the CVX 1800 by enabling the use of SS7 network infrastructure with dial access switch. It terminates SS7 signaling links and communicates with multiple CVX 1800 to terminate incoming data calls. Multiple CVX 1800 appears as a single "virtual switch" and point code to the PSTN network. As a result, the Gateway efficiently minimizes internet traffic across the PSTN, and decreases network congestion by optimizing call routing and processing.

2.7 Cisco's Solution

Cisco has developed SS7 Dial Access Solution (DAS), which provides a full integration of dial access capabilities within PSTN infrastructure, and provides significant savings on switching interface costs [Cisc98]. The DAS consists of a Signaling Controller working together with the NAS (network access server) to create a system that emulates a terminating or originating end-office telephone switch in PSTN. The SS7 network routes SS7 messages to the Signaling Controller, which converts the SS7 messages to a protocol recognized by the NAS. The Signaling Controller appears as a signaling point on SS7 network, while the NAS appears as a telephone switch to the PSTN, thus bypassing the local exchange carrier and the need to purchase ports on the telephone central office switch.

In addition, the Signaling Controller and NAS communicate via an IP network, using an extended Q.931 protocol. This allows the Signaling Controller to provide call control for multiple NAS, which can be located in the same or different geographical sites.

Chapter 3. Comparison of SS7 and TCP/IP/ATM

3.1 Summary

In this chapter, a layer-to-layer comparison is provided between SS7 and TCP/IP/ATM. The comparison shows some similarities and dissimilarities.

3.2 MTP Level 1 versus ATM Physical Layer

MTP1 is the physical layer in SS7, which can provide 56 or 64 kbps transmission rate in each of its 24 or 32 TDM (time division multiplexing) channels [Blac97]. The PMD (physical medium dependent) sublayer of the ATM physical layer can support SONET-based ATM UNI (user network interface) where the transmission can reach a very high rate [Mill94]. In addition, the transmission convergence sublayer of the ATM physical layer is responsible for HEC (header error check) for ATM cell headers. MTP1 does not perform this function as the error detection is performed at MTP2.

3.3 MTP Level 2 versus ATM

MTP2 is the data link layer, which provides point-to-point functions, such as forming signal units, performing error correction and detection. It uses a "go-back-N" technique of retransmission in error correction. But it is not responsible for the routing and addressing of the messages. The routing function is performed at higher layers. An MTP2 signal unit has a variable length and can have a maximum of 280 octets [Blac97].

The combination of the ATM layer and the AAL (ATM adaptation layer) corresponds to part of the data link layer [Mill94]. The addressing and routing functions are performed at the ATM layer using VPI (virtual path identifier) and VCI (virtual circuit identifier). It is not responsible for error correction, but the SAR (segmentation and reassembly) sublayer of the AAL performs sequencing and error detection in addition to segmentation and reassembly. An ATM cell has a fixed length of 53 bytes.

3.4 MTP3/SCCP versus IP

The combination of MTP3 and part of SCCP in SS7 corresponds to the IP layer in the TCP/IP model. One of the basic functions of MTP3 is to provide network routing for the transfer of messages between the source and destination nodes. The user data information may be passed down from the application layer to SCCP and MTP3, and is encapsulated into MTP3 messages. The header of an MTP3 message contains the OPC (origination destination code) and DPC (destination point code), and the message routing is determined by DPC. The address of a signaling point in the form of a GT, such as a 1-800 number, is first translated by SCCP into the routing address containing DPC. Then MTP 3 uses this DPC for further routing. Like IP address, a DPC or OPC is also hierarchical, in the form of: network identifier, cluster identifier, and cluster member identifier [Blac97]. But a DPC or OPC consists of 3 bytes, while an IP address consists of 4 bytes [Tane96].

At each signaling point (STP or SCP), there is a routing label for message routing, containing its PC (point code), a primary link selector and alternative link selector. When a message arrives at a signaling point, all the identifiers in the header are examined with the routing label at this node to determine whether the message is routed to itself or another node. A STP performs the routing functions. Just like a router, it routes the messages to the destination node in the same network (with the same network identifier) or a different network (with a different network identifier).

3.5 SCCP versus TCP/UDP

The other part of SCCP in SS7 corresponds to TCP/UDP. TCP provides a reliable connection-oriented data transmission. Similar to TCP, the functions of flow control (sliding window), sequencing, multiplexing, segmentation and reassembly, are also performed at SCCP in SS7. SCCP protocol class 3 is connection-oriented with sequencing and flow control. It uses the sliding window mechanism, allowing a maximum of 127 signal units to be transmitted continuously without acknowledgement of any single signal unit. SCCP protocol class 3 can multiplex a number of SS7 connections into a single MTP 3 connection. Due to the limitation on the size of an MTP2 signal unit or the limited SCCP message receiving buffer [Blac97], it segments a large message into smaller segments. Likewise, TCP divides user data into multiple segments, each of which has a size fit for the physical transmission medium. TCP and SCCP both do reassembly in the receiving node [Tane96].

UDP supports connectionless transmission for small messages, or the messages over reliable channels. UDP increases transmission efficiency by ignoring the loss of data [Burk97]. A UDP datagram overhead is very short as all the fields in a TCP segment for flow control and sequencing are removed in the UDP datagram. Similar to UDP, the SCCP protocol class 0, providing basic connectionless connection, does not perform flow control and sequencing either.

3.6 ASE/TCAP versus Application Layer

The ASE and TCAP in SS7 supports non-circuit-related message transfer between exchanges and databases [Boss97], and correspond to the application layer in TCP/IP model. The TCAP usually requires the support of SCCP connectionless protocol classes (class 0 and 1). In TCP/IP stack, the user (application layer) is supported by either TCP or UDP for data transfer, depending on the requirement of the transmission reliability.

Chapter 4. Interworking Mechanism

4.1 Summary

In this chapter, we will provide the details of our proposed solution for the interworking of TCAP/SS7 and TCP/IP, which will be defined as "TCAP over IP".

4.2 General Architecture of a TCAP Application

For a TCAP application in SS7, a SSP initiates a remote transaction, such as 1-800 toll free call, by sending a TCAP query message to a SCP. It ends the transaction by receiving a TCAP response message from the SCP which carries the information requested. In the architecture of our interworking solution, it is recommended that the transfer of the TCAP messages across TCP/IP network to its destination in SS7 networks be through a SG (signaling gateway) [Mato98]. The SG works as an interface between the two different networks resolving the protocol conversion (as shown in Figure 4.1).

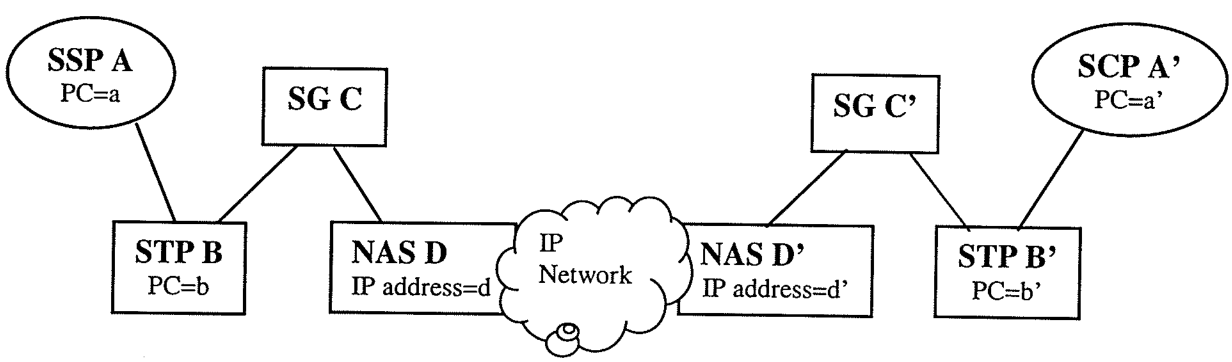


Figure 4.1: SS7-TCP/IP Interworking Topology

4.3 Inter-layer Interactions at a SSP

Before we discuss the TCAP over IP, a good understanding of how TCAP works in SS7 will be very helpful for understanding the interworking solutions.

(Note: all the following codes are the national specifications in North America unless specified otherwise).

4.3.1 Interactions between ASE and the TCAP

As shown in Figure 4.2 [Boss97], when an ASE at SSP A, identified by the SSN q, needs to query the information at a remote SCP A', it passes a TC Primitive containing an Invoke component to the TCAP layer, notifying the TCAP layer to invoke a transaction.

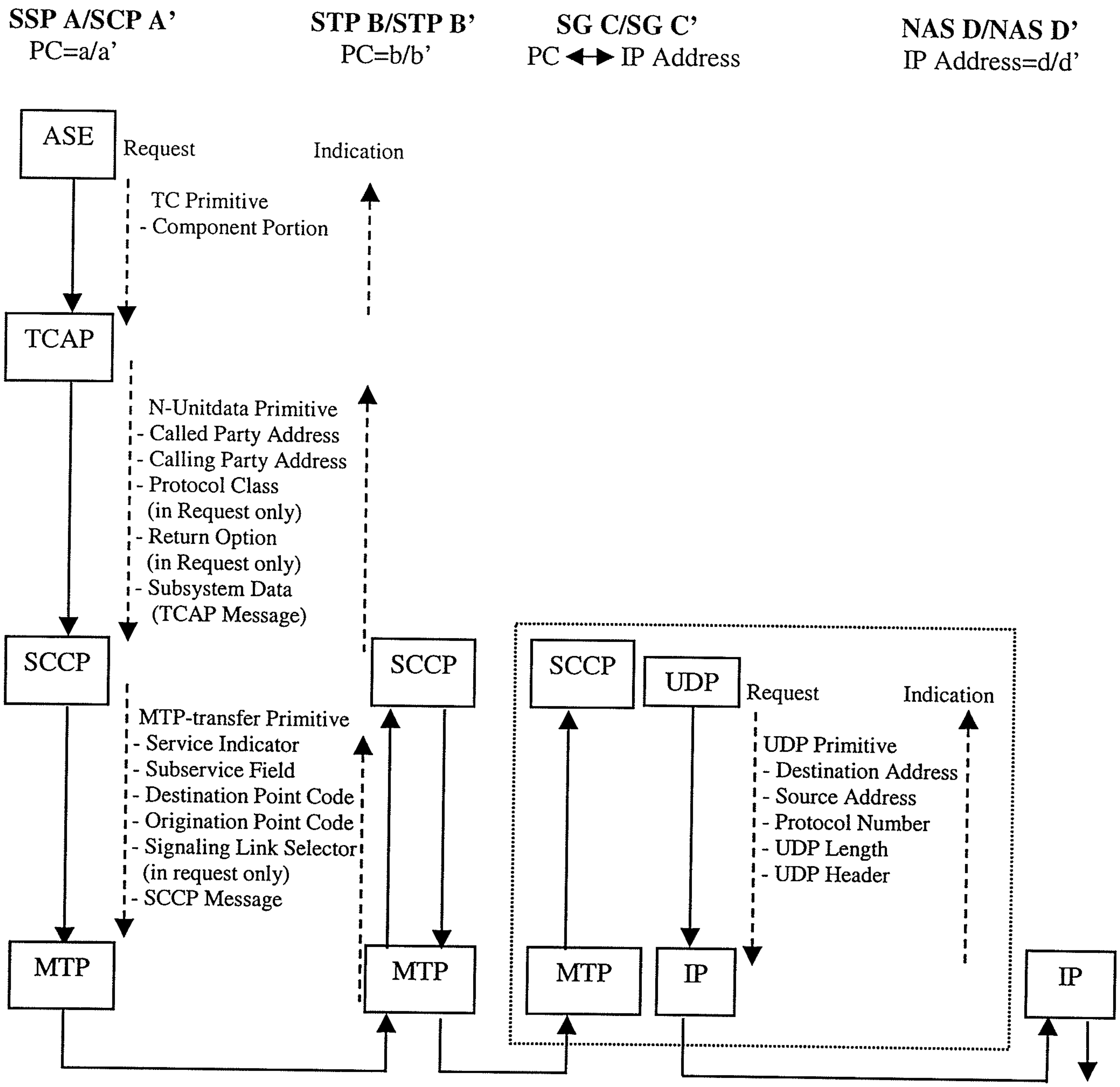


Figure 4.2: Interworking Interactions

The format of an Invoke (Last/Not Last) component usually is given by [Anst92]

8 7 6 5 4 3 2 1
Component Type Identifier Component Length
Component ID Identifier Component ID Length Component t IDs
Operation Code Identifier Operation Code Length Operation Code
Parameter Set/Sequence Identifier Parameter Set/Sequence Length Parameter Set/Sequence

Figure 4.3: Invoke Component Format

where

- Component Type Identifier: is 11101001 for Invoke (Last), and is 11101101 for Invoke (Not Last).
- Component Length: is the total number of bytes in this Invoke component.
- Component ID identifier: is 11001111.
- Component ID Length: is the total number of bytes in Component ID, ranging from 0 to 2.
- Component ID: is Invoke ID of one byte long, and assigned by the origination node to identify an invoke in a transaction.
- Operation Code Identifier: is 11010000.
- Operation Code Length: is the total number of bytes in Operation Code, usually 2 byte long.

- Operation Code: specified by an application.
- Parameter Set Identifier: is 11110010, indicating that a set of parameters is to follow.
- Parameter Sequence Identifier: is 00110000, indicating that a sequence of parameters is to follow.
- Parameter Set/Sequence Length: is the total number of bytes in Parameter Set/Sequence.
- Parameters: specifies the operands for the requested operation. Digits is a parameter containing the Calling/Called Party Number or Routing Number, which has the following format

8 7 6 5	4 3 2 1
Digit Identifier	
Type of Digits	
Nature of Number	
Numbering Plan	Encoding
Number of Digits	
Digit 1 (0001)	Digit 0 (0000)
Digit 3 (0011)	Digit 2 (0010)
Digit 5 (0101)	Digit 4 (0100)
Digit 7 (0111)	Digit 6 (0110)
Digit 9 (1001)	Digit 8 (1000)
...	...

Figure 4.4: Digits Parameter Format

where

- Digit Identifier: 10000100.
- Type of Digits: 00000001 - Called Party Number, the digits dialed by the customer;

  00000010 - Calling Party Number;

  00000100 - Routing Number, the digits associated with network routing.
- Nature of Number: xxxxxxx0, where x is either 1 or 0.
- Numbering Plan: 0xxx, where x is either 1 or 0.
- Encoding: 0000 - not used; 0001- BCD (binary coded decimal).
- Number of Digits: total number of binary bits in the Digits field.

(Note: there are many other parameters in a Component Portion, here we just list the parameter "Digits" as an example)

4.3.2 Interactions between the TCAP and the SCCP

The Transaction Portion - Query With/Without Permission is added to the Component Portion to form a TCAP Begin message. The format of Query with/without Permission is given by [Anst92]

8 7 6 5 4 3 2 1
Package Type Identifier Total TCAP Message Length
Transaction ID Identifier Transaction ID Length Originating Transaction ID
Component Sequence Identifier Component Sequence Length

Figure 4.5: Query Permission Transaction Portion Format

where

- Package Type Identifier: is 11100010 for Query With Permission, and is 11100011 for Query Without Permission.
- Total TCAP Message Length: is the total number of bytes in Transaction Portion and Component Portion.
- Transaction ID Identifier: is 11000111.
- Transaction ID Length: is the total number of bytes in Transaction IDs.
- Originating Transaction IDs: assigned by the origination node, 4 bytes long.
- Component Sequence Identifier: is 11101000, indicating a sequence of one or more Components that follow.
- Component Sequence Length: is the total number of bytes in the Components contained in the Component Portion.

The TCAP Begin message has the format

8	7	6	5	4	3	2	1
Transaction Portion (Query with/without Permission)							
Component Portion (Invoke component)							

Figure 4.6: TCAP Begin Message Format

The TCAP layer passes an N-unitdata Request primitive to the SCCP [Anss92]. The information in this primitive includes:

- Called Party Address: its format is given by

8	7	6	5	4	3	2	1
1	R I	GTI				SNI	PCI
Point Code							
Subsystem Number							
Global Title							

Figure 4.7: Called Party Address Format

where

- PCI (point code indicator): 1 - point code is present; 0 - point code is absent.

- SNI (subsystem number indicator): 1 - subsystem number is present; 0 - subsystem number is absent.

- GTI (global title indicator): 0000 - No global title included;

0001 - Global title includes translation type, numbering plan and encoding scheme;

0010 - Global title includes translation type, used in U.S. networks.

- RI (routing indicator): RI=0 - Global title translation should be performed, and routing should be based on the translation result;

RI=1 - Global title translation should not be performed, and routing should be based on the destination point code in the routing label of the message, and the subsystem number in the Called Party Address.

- Point Code: MTP 3 point code. When GTI is non-zero, its value is zero.

- Subsystem Number: indicates the SCCP user, values ranging from 0 from 127. The Bit 0 means that the subsystem number is not known/used; the numbers from 1 to 10 have been assigned to some SCCP users; the numbers from 11 to 127 are spare or reserved for expansion. For example, in 1-800 number service, the subsystem number is 00000010.

- Global Title: consists of translation type and GTA (global title address). In 1-800 service, the global title address is the dialed 1-800 number. The format of a Global Title is given in Figure 4.8.

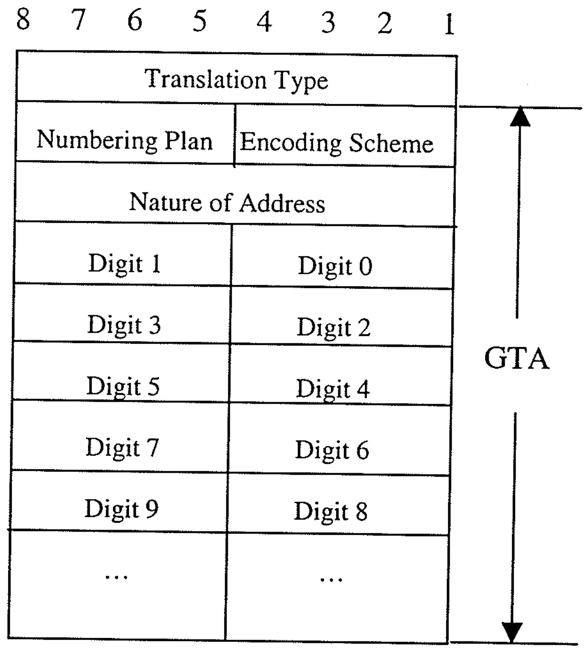


Figure 4.8: Global Title Format

where

- Translation Type: specifies the type of translation to be performed.

  Encoding Scheme: is BCD (binary coded decimal), and is absent when GTI=0010.

  0000 - not used; 0001 - the number of digits is odd; 0010 - the number of digits is even.

- Numbering Plan: 0xxx, where x is either 1 or 0. It is absent when GTI=0010.
- Nature of Address: 00000110.

- Calling Party Address: its format is the same as Called Party Address except that global title is only present in the transactions between different networks. The point code in Calling Party Address is that of SSP A, i.e., PC=a.

- Protocol Class: is 0000, representing SCCP basic connectionless protocol class 0, where the SCCP messages can be delivered out of sequence.
- Return Option: is 4 bits long, and is an optional QOS parameter present in N-unitdata Request primitive only, used to determine the handling of messages encountering transport problems. 0000 - discard message on error; 1000 - return message on error.

4.3.3 Interactions between the SCCP and the MTP3

After the SCCP layer receives an N-unitdata Request primitive, it is aware that the transfer of TCAP messages is not related to individual trunks, and it must provide the GT (global title) translation when necessary [Anss92]. It then forms an SCCP Unitdata message with the format

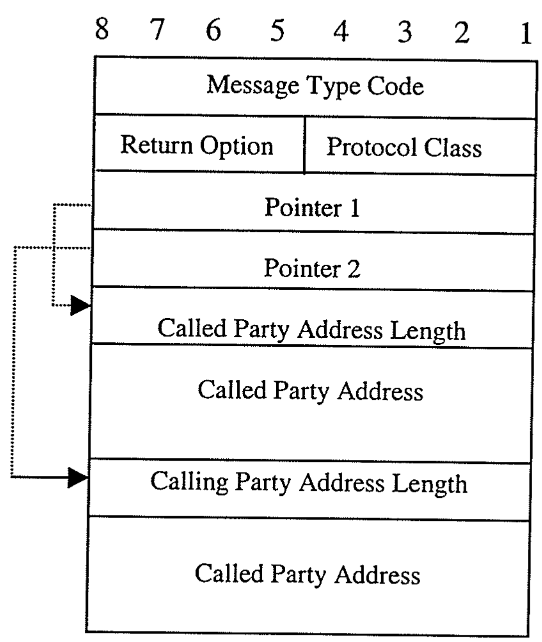


Figure 4.9: SCCP Unitdata Message Format

where

- Message Type Code: is 00001001 for Unitdata message.
- Pointer 1 (Pointer 2): is 1 byte long, used to indicate the beginning position of the Called (Calling) Party Address field.
- Return Option: the same as aforementioned.
- Protocol Class: is 0000.
- Called (Calling) Party Address Length: is total number of bytes in the Called (Calling) Party Address.

The SCCP layer passes an MTP-transfer Request primitive to MTP3 layer [Ansm92]. This primitive contains the following information:

- Service Indicator: is 4 bits long, performing message distribution and indicating MTP3 user. When its value is 0011, it indicates the MTP3 user is SCCP.
- Subservice Field: is 4 bits long, identifying international and national network. 0010 - national network.
- DPC/OPC: is 3 bytes long, consisting of network ID, network class ID and class member ID.
- SLS (signaling link selector): is 5 bits long. But in SCCP protocol class 0, it can be random.

4.3.4 MTP Layers

After MTP3 layer receives MTP-transfer Request primitive, it forms an MTP3 SU (signal unit), which has the format [Ansm92]

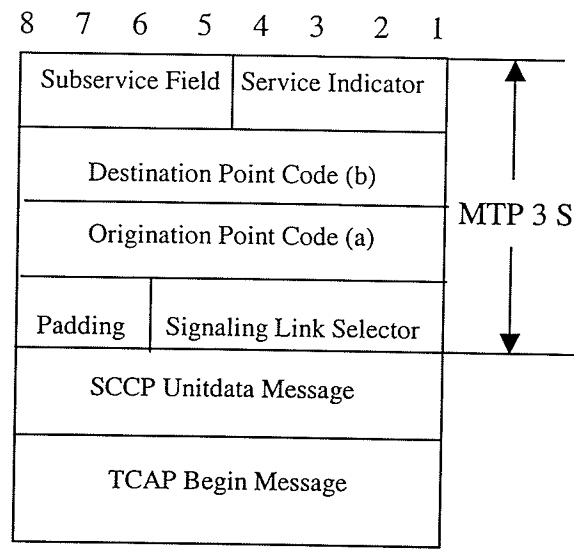


Figure 4.10: MTP3 SU Format

Finally, MTP2 encapsulates an MTP3 SU into a frame, and places it into MTP1 for transmission.

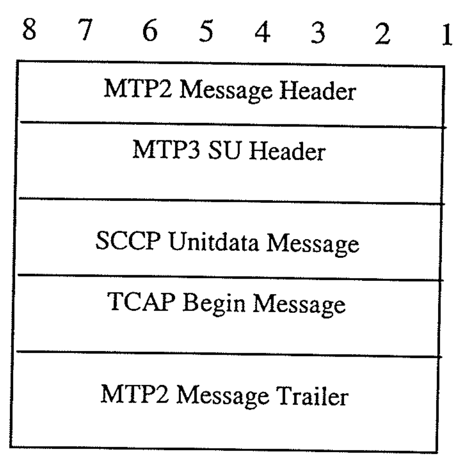


Figure 4.11: MTP2 Message Format

4.4 Message Routing at a STP

When STP B receives an MTP message, its MTP3 layer examines the routing label where the DPC=b, and the OPC=a. Since the DPC is b, this MTP3 SU should be routed within STP B and so passed to its higher layer. If the service indicator identifies SCCP as its user, an MTP-transfer indication primitive is then passed to the SCCP layer [Boss97].

After receiving an MTP-transfer indication primitive, the SCCP examines the GTI contained in the Called Party Address of the Unitdata message (in the applications like 1-800 service, a non-zero GTI indicates that a global title translation is required). Then SCCP translates the global title contained in the Called Party Address into PC=a', SSN=r where a' is the point code of a destination node (i.e. SCP A'), and r is the subsystem number at SCP A'. Then the other parameters contained in the Called Party Address of the outgoing Unitdata message change the values - say PCI changes from 0 to 1, SNI changes from 0 to 1, PC changes from 0 to a', subsystem number changes from 0 to r.

The SCCP then passes an MTP request primitive with OPC=b, and DPC=a' down to MTP3 layer, which are used in the routing labels for routing the MTP3 SU. From the DPC, STP B is aware that this MTP3 SU must be sent across the SS7 network to another STP for further routing. It then passes this MTP3 SU to SG C.

If the SCCP at STP B cannot deliver the received Unitdata message, it sends a Unitdata Service message back to SSP A [Anss92]. This message contains Return Cause, a one-byte parameter indicating why the Unitdata message is being returned (if the Return Option in this Unitdata message is Return Message on Error). A Unitdata Service message has the format

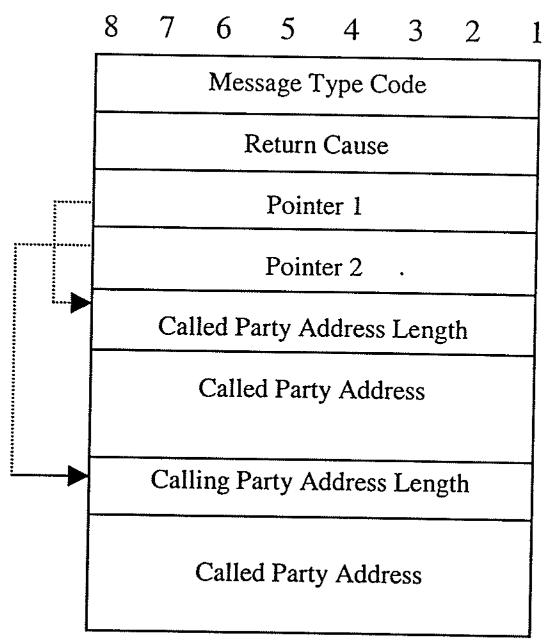


Figure 4.12: SCCP Unitdata Service Message Format

where

- Message Type Code: 00001010.
- Return Cause: 00000000 - No translation for an address of such nature;

  00000101 - Network failure;

  00000001 - No translation for this specific address;

  00000111 - Other failures.

The Pointer 1 & 2, and the Called/Calling Party Address are the same as the ones in the Unitdata message.

4.5 Protocol Conversion at a SG

A SG normally has a look-up table maintained inside it for the address mapping, which contains the mapping between PC and IP address, and the mapping between the SSN and UDP port numbers. Figure 4.13 shows how the protocol mapping inside a SG works.

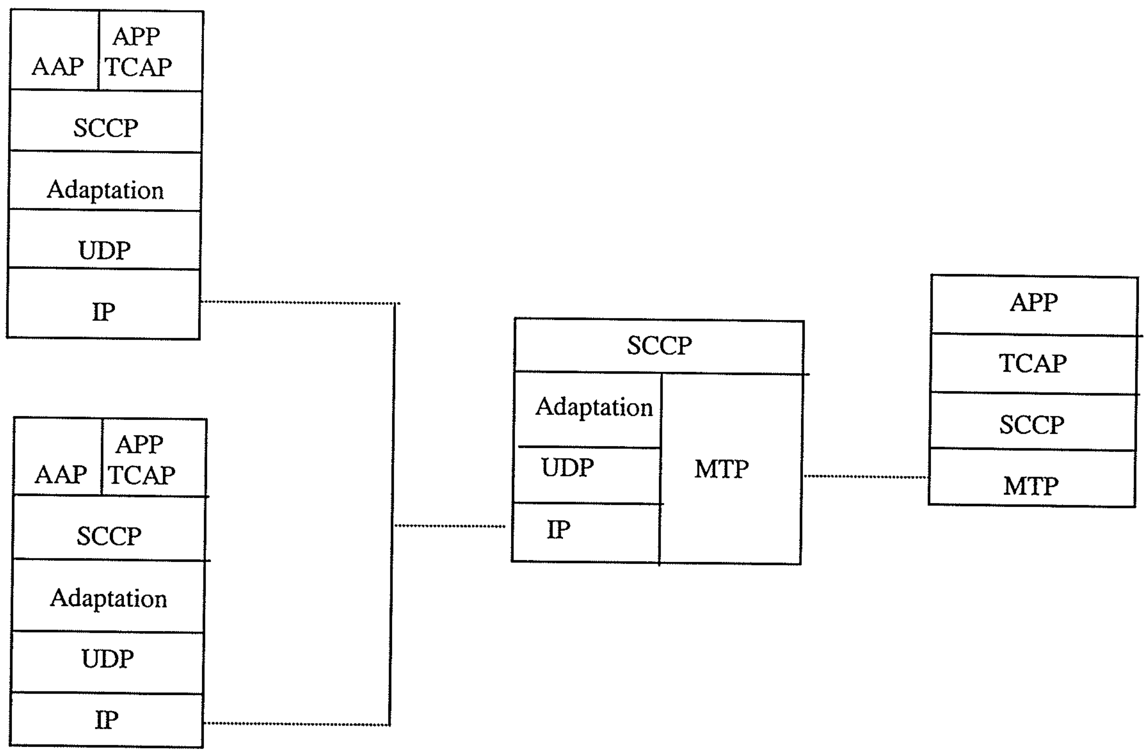


Figure 4.13: Protocol Mapping within a SG

The TCAP is supported by the SCCP connectionless classes, so the UDP is chosen as the transport layer on the TCP/IP side for the proposed TCAP over IP, in order to maintain compatibility with this connectionless feature. The TCP is not suitable for the proposed TCAP over IP, because if the TCP is used, many transactions would need to establish and maintain many connections, this would use too many resources of the network.

4.5.1 Adaptation Layer

From Figure 4.13, it can be seen that there is an Adaptation layer in between UDP and SCCP layers. An Adaptation layer is designed to provide adaptation between SCCP and UDP in such a way that it performs the address conversion between the PC and the IP address, and between the SSN and the UDP port number. It also provides the network management and control functions to detect failures and congestion at remote SG by employing SCCP Unitdata Service messages and IP ICMP messages. The discussion of the management and control will be provided in a later chapter.

SG C is the termination of SS7 network on the side of SS7, and is the termination on the side of IP network:

1) On the SS7 side, once an MTP SU is received, the MTP layer passes an MTP-transfer Indication primitive to the SCCP layer. The SCCP layer then passes a primitive of Adaptation-transfer Request to the Adaptation layer, carrying the DPC, OPC and the destination and origination SSN. The Adaptation layer converts the DPC/OPC to an IP address, and converts a SSN to a UDP port number, by triggering a table look-up.

Finally, an IP packet is formed, encapsulating UDP header and TCAP/SCCP message, and is delivered across IP network to a remote SG.

2) On the IP side, once an IP packet is received, the IP layer passes an Indication primitive to UDP layer, carrying the destination and origination IP addresses, and destination and origination UDP port numbers. After the Adaptation layer obtains this information, it converts an IP address to a PC, and converts a UDP port number to a SSN, by triggering a table look-up. Then the Adaptation layer passes an Adaptation-transfer Indication containing all the converted information to the SCCP layer.

Finally, an MTP message is formed, encapsulating the TCAP/SCCP message, and is delivered across SS7 network to the signaling point that the message is intended for.

4.5.2 Look-up Table for the Conversion between PC and IP Address

A PC consists of 3 bytes, but the IP address is 4 bytes long, therefore, the mapping of SS7 PC to IP address cannot be direct, because one cannot insert eight 0 bits into a PC to generate an IP address since it may result in a conflict with IP addresses that are already in use elsewhere. But one can establish a look-up table where a PC of an SSP/SCP is mapped to an unassigned IP address of a geographically nearest ISP's (internet service provider) POP (point of presence), or vice versa. Such mappings are possible to implement because the number of SSP/SCP in each city or area is quite limited and one just needs a limited number of unassigned IP addresses for the mapping.

Once an SG C receives an MTP message, its table look-up is triggered. Using PC=a' of the destination node (i.e., SCP A'), the IP address of the NAS D' is retrieved from the table. The PC of the origination node (i.e., SSP A) contained in the calling party address is converted to the IP address of NAS D after a table look-up is finished.

4.5.3 Look-up Table for the Conversion between the SSN and the UDP Port Number

A SS7 subsystem number ranges from 1 to 10. A UDP port number is 2 bytes long, so its value ranges from 0 to 65535, many of which are unassigned. So it is quite feasible to choose 10 out of them for mapping to a SSN, or vice versa. Such a look-up table can also be established inside SG C. Once an SG receives an SCCP message, a table look-up is triggered and the corresponding port number can be retrieved from the table, according to the SSN contained in this SCCP message.

4.5.4 Encapsulating TCAP/SCCP Messages into an IP Packet

After finishing the above procedures, an IP packet will be formed by encapsulating the TCAP+SCCP message with the corresponding UDP and IP, as shown in Figure 4.14:

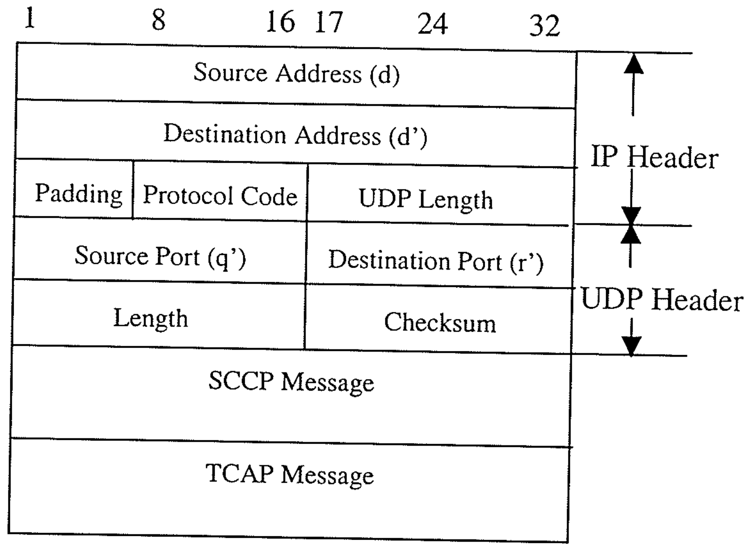


Figure 4.14: An Encapsulated IP Packet Format

where the source address is the IP address of NAS D with the value of d. The destination address is the IP address of NAS D' with the value of d'. The source port number is q' which is converted from the SSN at SSP A with the value of q, and the destination port number is r' which is converted from the SSN at SCP A' with the value of r.

4.6 Routing at NAS

NAS D then routes this IP packet to NAS D' using its IP destination address. After examining the destination port number, NAS D' is aware that the application layer of UDP is SCCP+TCAP and that this packet carries SS7 SCCP+TCAP messages. It routes this packet to SG C' where the packet is converted to an MTP3 SU.

4.7 Query Processing at a SCP

SG C' converts an IP packet to an MTP3 SU after finishing the table look-up to map the destination/origination IP addresses to the PC of SCP A'/SSP A. The port number within the UDP header indicates to SG C' that the MTP3 user is SCCP, so the Service Indicator in this MTP3 SU shall have the value of 0011. Then this MTP3 SU, which has the value of a' and a as its DPC and OPC respectively, is routed to STP B'. The message relay at STP B' is the same as that at STP B. Finally this MTP3 SU is routed to its destination node (i.e. SCP A').

At SCP A', the MTP3 layer passes an MTP-transfer indication primitive to the SCCP layer. The contents in this primitive are almost the same as that in MTP-transfer request primitive except for the SLS. The SCCP layer then passes an N-unitdata indication primitive to the TCAP layer. The contents in this primitive is almost the same as that in N-unitdata request primitive except for the Protocol Class and Return Option. The TCAP layer then uses the SSN r contained in the Called Party Address to pass a TC indication primitive to the ASE specified by r. The ASE r then implements the transaction. In 1-800 service, the ASE maps the 1-800 number to a regular telephone number of a company's nearest service center to SSP A, and puts this number in the Digits parameter field of Routing Number (Type of Digits is coded as 00000100) which is contained in the Component Portion [Boss97].

4.8 The Status of the Transaction at a SCP

A transaction at a SCP could be either successful or unsuccessful, and the two situations are handled in different ways:

- Successful Transactions [Boss97]:

If the transaction at SCP A' is successful, the TCAP layer forms a TCAP End message, carrying the information requested by SSP A. In this TCAP End message, the Component Portion is Return Result whose format is similar to the one in Figure 4.3, but it does not contain the Operation Code (Identifier and Length). The Component ID is the Correlation ID with the same value as the Invoke Component received from SSP A, where the parameters hold the result of the operation of this transaction. The Component Type Identifier is 11101010 for Return Result (Last), or 11101110 for Return Result (Not Last). The Transaction Portion is Response whose format is similar to that in Figure 4.5, except that the Transaction ID is the Responding Transaction ID with the same value as the Originating Transaction ID received from SSP A. The Package Type Identifier is 11100100.

The Called Party Address and Calling Party Address in the outgoing SCCP Unitdata message are the same as those in the incoming SCCP Unitdata message. All further routings for outgoing MTP3 SU are then performed, as previously described in Section 4.3.

- Unsuccessful Transactions [Boss97]:

If the transaction at SCP A' is not successful, the ASE r passes a TC indication primitive containing Return Error Component or Reject Component to the TCAP layer. The TCAP layer then forms a TCAP End Message where the Transaction Portion is Abort (P-Abort) or Abort (User Abort). All the subsequent procedures are the same as those described in Successful Transactions.

The Return Error Component has the format

8 7 6 5 4 3 2 1
Component Type Identifier Component Length
Component ID Identifier Component ID Length Component IDs
Error Code Identifier Error Code Length Error Code
Parameter Set/Sequence Identifier Parameter Set/Sequence Length Parameter Set/Sequence

Figure 4.15: Return Error Component Format

where

- Component Type Identifier: is 11101011.
- Component ID Identifier: is 11001111.
- Component IDs: is the Correlation ID with the same value as the Invoke ID in the Invoke Component received from SSP A.
- Error Code Identifier: is 11010011.

- Error Code: specified by the ASE, providing the reason why a specific operation could not be completed successfully.
- Parameter Set/Sequence Identifier: is the same as the ones in Figure 4.3.
- Parameters: provide the information of the cause of the failure of the operation.

The Reject Component has the format

8 7 6 5 4 3 2 1
Component Type Identifier Component Length
Component ID Identifier Component ID Length Component IDs
Problem Code Identifier Problem Code Length Problem Code
Parameter Set/Sequence Identifier Parameter Set/Sequence Length Parameter Set/Sequence

Figure 4.16: Reject Component Format

where

- Component Type Identifier: is 11101100.
- Component ID Identifier: is 11001111.
- Component IDs: is the Correlation ID with the same value as the Invoke ID in the Invoke Component received from SSP A.
- Problem Code Identifier: is 11010101.

- Problem Code: provides the reason why a component was rejected. It consists of a Problem Type followed by an associated Problem Specifier. There are 5 Problem Types, and in each Problem Type there are 2 to 6 Problem Specifiers.
- Parameter Sequence Identifier/Set Identifier: the same as those in Figure 4.3.
- Parameters: provide the information of the cause of the rejection of a received component.

The Abort (P-Abort)/Abort (User Abort) has the format

8 7 6 5 4 3 2 1
Package Type Identifier Total TCAP Message Length
Transaction ID Identifier Transaction ID Length Responding Transaction IDs
P-Abort Cause (or User Abort Information) Identifier P-Abort Cause (or User Abort Information) Length P-Abort Cause (or User Abort Information)

Figure 4.17: Abort Transaction Portion Format

where

- Package Type Identifier: is 11110110.
- P-Abort Cause Identifier: is 11010111.
- P-Abort Cause: there are 6 values for it. For example, the Unrecognized Package Type has the value 00000001.
- User Abort Information Identifier: is 11011000.

- User Abort Information: the ASE or the TCAP layer may provide any information desired as the contents of the User Abort Information.

4.9 Handling Failures at the SCCP Layer

If the SCCP at SCP A' cannot deliver the Unitdata message to the TCAP, it sends a Unitdata Service message to SSP A [Boss97]. The format of this message has been shown in Figure 4.12, where the Return Cause is:

00000011 - Failure of called subsystem; 00000100 - Subsystem not equipped;

00000101 - Network failure; 00000111 - Other failures.

4.10 Message Processing at a SSP

After the ASE q at SSP A receives the information requested, it will do further processing [Boss97]. For example, in 1-800 service, the regular telephone number translated at SCP A' is further used by SSP A to route the phone call to a company's service center which is the nearest to the customer who initiated the phone call.

Chapter 5. Management & Control Functions for the Interworking

5.1 Summary

This chapter addresses the reliability of the proposed TCAP over IP solution. Because a SG has played such a key role in the protocol conversion, it is very important, at all time, to keep it from failing. It is recommended that once a SG fails, the affected nodes in SS7 networks shall be immediately informed through the management and control functions, and the SG's redundant peer shall take over all its functions for the data transmission.

5.2 Redundancy of SG and NAS

To enhance the reliability of our TCAP/IP, a SG shall have a redundant peer, i.e., a primary SG has a stand-by SG, as shown in Figure 5.1. In the event of a SG's failure, the reliability can still be guaranteed by automatically switching from this SG to its stand-by SG.

In addition, to keep the TCAP/IP traffic load balanced, a SG shall be connected to two NAS. When the active SG detects that one NAS is busy, it automatically redirects the subsequent messages to another NAS with free capacity.

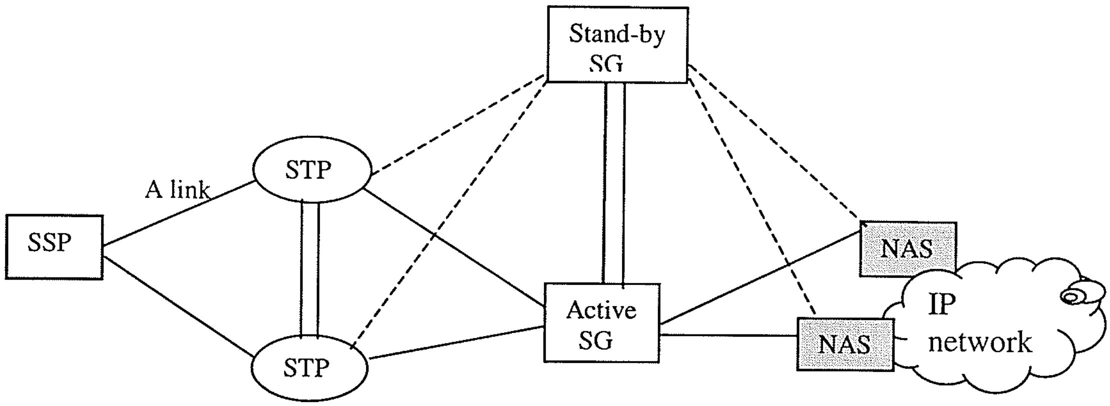


Figure 5.1: SG and NAS Redundancy

There are several alternatives to provide the redundancy for a SG:

- Cold Stand-by: the stand-by SG has no knowledge of current state of the alive SG, and has to retrieve it from the controlled devices or reset them before taking over.
- Warm Stand-by: the stand-by SG always has some knowledge of current state of the live SG. It may either retrieve additional state from the controlled devices or invoke a partial reset in specific cases where recovery is impossible.
- Hot Stand-by: the stand-by SG has a full knowledge of current state of the alive SG and can take over without loss of service. The stand-by SG constantly polls the active SG for new and changed status and updates itself on a regular basis. When the stand-by SG becomes active, its configuration mirrors that of the former active SG.

5.3 Management and Control of a Remote SG

The management and control of a remote SG is implemented at the Adaptation layer of a local SG by mapping the IP ICMP messages to the SCCP Unitdata Service messages. Once the local SG receives an ICMP message carrying the IP address and UDP port number of a remote SG, it is aware that the SCCP layer at the indicated SG might experience a failure, a congestion or a recovery. Then the Adaptation layer translates the IP address and UDP port number into SS7 point code and subsystem number, and possibly translates ICMP Type & Code value into SCCP Return Cause value. The Adaptation layer then forms a corresponding Adaptation Indication primitive carrying the translated point code, subsystem number and possibly SCCP Return Cause value, and passes it to SCCP layer at the local SG. This Adaptation Indication primitive could be an Adaptation Pause Indication, or an Adaptation Resume Indication, or an Adaptation Status Indication.

1) Adaptation-pause Indication

This primitive indicates to the SCCP layer at the local SG of the inability to send any TCAP/SCCP messages to the indicated remote SG. It contains the fields of Affected Point Code and Affected Subsystem Number. Thus, the SCCP layer at the local SG must not send any message to that remote SG.

The Adaptation layer generates this primitive as a result of a remote SG's error which was detected by Adaptation layer supervision mechanisms. This supervision is performed by employing the services of ICMP Echo and Echo Reply messages. The periodic polling of a remote SG is implemented by sending ICMP Echo messages.

2) Adaptation-resume Indication

This primitive notifies the SCCP layer at the local SG that it is available to send TCAP/SCCP messages again to the indicated remote SG. It contains the fields of Affected Point Code and Affected Subsystem Number. Thus, the SCCP layer at the local SG can resume sending messages to that remote SG.

The Adaptation layer generates this primitive when the SCCP layer at a remote SG starts to answer again to the ICMP Echo messages, as explained above.

3) Adaptation-status Indication

This primitive informs the SCCP layer at the local SG that the SCCP layer of the indicated remote SG is unavailable, or that a congestion is occurring at the SCCP layer at the indicated remote SG. It contains the fields of Affected Point Code, Affected Subsystem Number, and Return Cause parameter.

This Adaptation layer generates this primitive if the Adaptation layer at a remote SG receives a message for SCCP, but the SCCP layer is not working due to a failure or a congestion occurring on it. Then the remote SG generates an ICMP Destination Unreachable message and sends it back to the local SG.

Once the local SG receives this ICMP message, the Adaptation layer examines its fields of Type & Code and maps them to the corresponding Return Cause parameter. Then an Adaptation-status Indication primitive is passed from the Adaptation layer to the SCCP layer, carrying the Affected Point Code, Affected Subsystem Number and Return Cause parameter. Finally, a Unitdata Service message is formed at SCCP layer and returned to the origination node in SS7 network.

The following table gives the mapping of the values of the Type & Code field of ICMP messages and the values of the Return Cause parameter of the SCCP Unitdata Service messages [Post81]:

ICMP Messages Type & Code	Value	SCCP Return Cause	Value
Destination Unreachable Net Unreachable Code	3 0	Network Failure	5
Destination Unreachable Host Unreachable Code	3 1	Unequipped User	4
Destination Unreachable Protocol Unreachable Code	3 2	Unqualified	7
Destination Unreachable Port Unreachable Code	3 3	Subsystem Failure	3
Source Quench Code	4 0	Network Congestion	6

Table 5.1: Mapping of ICMP Type & Code to SCCP Return Cause

Chapter 6. Performances of the Interworking

6.1 Summary

In this chapter, an overview of the mandatory performance requirements of TCAP/SS7 for delays and packet loss rates defined in the ITU-T or Bellcore SS7 specification documents are provided. An overview of modeling SS7's performance and some approaches to measuring the performance parameters of a SG through simulations are also provided.

6.2 Performance Requirements & Modeling of SS7

The end-to-end AIN performance is a function of the performance of each AIN network element, such as SSP, STP, SCP, and their interfaces. Therefore, the performance objectives and requirements must be defined for each component. The SSP generally originates a query to the SCP and waits for its response. The performance of a TCAP application relies on the Query Response time. The performance at a SCP depends not only on its several components and interfaces but also on the application processes involved [Seth99].

6.2.1 Message Delay

MTP3 peer-to-peer procedures require response within 500ms to 1200ms, which includes round trip time and processing at the remote end [Seth99].

6.2.2 Query Response Time

Query Response time is the time it takes for a querying entity (e.g. SSP) to send a query to a database host (e.g. SCP) and for the database to process the query and return data to the querying entity. The actual PSTN working data shows that the mean values of most simple TCAP application query-response time are in the range from 250ms to 500ms [Seth99].

6.2.3 SSP Response Time

The SSP Response Time is the period that starts when a stimulus occurs at the SSP and ends when the SSP completes its response to the stimulus. It involves sending a message on the outgoing signaling link, so it consists of SSP processing time, and the link output delay. Its probabilistic average and confidence interval for TCAP services are given by Table 6.1 [Seth99]:

	SSP Response Time (TCAP Message)
Mean	210 - 222ms
95% probability	≤ 342 - 354ms

Table 6.1: SSP Response Time

6.2.4 Message Transfer Time at a STP

This is the period which starts when the last bit of the signal unit leaves the incoming signaling data link and ends when the last bit of the signal unit enters the outgoing signaling data link for the first time. It consists of STP processor handling time and outgoing link delay. Its probabilistic averages and confidence intervals corresponding to three different levels of traffic load are given by Table 6.2 [Seth99]:

	Normal traffic load	Normal traffic load +15%	Normal traffic load +30%
Mean	20ms	40ms	100ms
95% probability	40ms	80ms	200ms

Table 6.2: Message Transfer Time at a STP

6.2.5 SCP Response Time

The SCP response time is defined as the interval that begins when the last bit of a call-related message enters the SCP, and ends when the last bit of this call-related message leaves the SCP. It is a sum of the SCP Handling Time, Disk Lookup (table lookup) Time and Link Output Delay. The total SCP response time for a simple TCAP transaction (one query, one response) is summarized in Table 6.3 [Seth99]:

	Daily Peak	Yearly Peak
Mean value	$\leq$250ms	$\leq$400ms
0.95 Probability	300ms	600ms

Table 6.3: SCP Response Time

6.2.6 Message Loss Rate

No more than 10^{-7} message loss rate due to a transport failure is allowed [Itub89].

6.2.7 Performance Modeling of SS7

Performance modeling or evaluation is an important step in estimating the required quality of service, particularly when new services are to be introduced into an existing network, for example, the proposed interworking of TCAP/SS7 and UDP/IP.

The protocol architecture of SS7 has been evaluated by modeling individually specific parts of the architecture such as MTP, SCCP and TCAP. In this approach, the major functional parts of the layered architecture are modeled by queuing network elements. The SS7 overall performance is obtained by a combined decomposition and aggregation method. Hence the complex system is decomposed into the simpler single-input, single-processor M/G/1 priority queuing models, which are analyzed in isolation [Will90]. The overall response times and end-to-end transfer times are computed by summing the individual transfer times and response times along the corresponding paths of the signaling network. It should be observed that the decomposition-aggregation approach ignores mutual relationships between the individual components of the model.

Simplifying the approach to performance modeling of the signaling links of SS7 is described in ITU-T Recommendation Q.706 [Itub89]. In this document, the performance model is based on the input traffic aggregation and the single-input, single-processor M/G/1 priority assignment approximation. A collection of formulas for mean total queuing delay, variance of queuing delay and the proportion of messages delayed more than a given time is provided.

6.3 Performance Measurement of a SG

The aforementioned performance modeling studies of SS7 are concerned with the classical situation of handling high priority traffic. Consequently the inter-arrival time distribution can be assumed to be exponential (equivalently the number of messages can be described by Poisson process). On the TCP/IP side of a SG, the IP traffic does not meet this condition [Lela94] [Paxo95], because the self-similar and long range dependence nature of the aggregate Ethernet traffic make it quite different from the Poisson assumption [Lela94]. However, since the inter-arrival time of the packets towards a SG on the SS7 side follows an exponential distribution [Will90], it makes the simulation of the traffic through a SG feasible to implement.

Our recommendation is that to find the best possible performance model for TCAP/SS7 with UDP/IP configuration, a data driven approach shall be used, namely, having a record of observations of the input/output traffic, and letting data decide about the adequate model. The required data collection plan is illustrated in Figure 6.1:

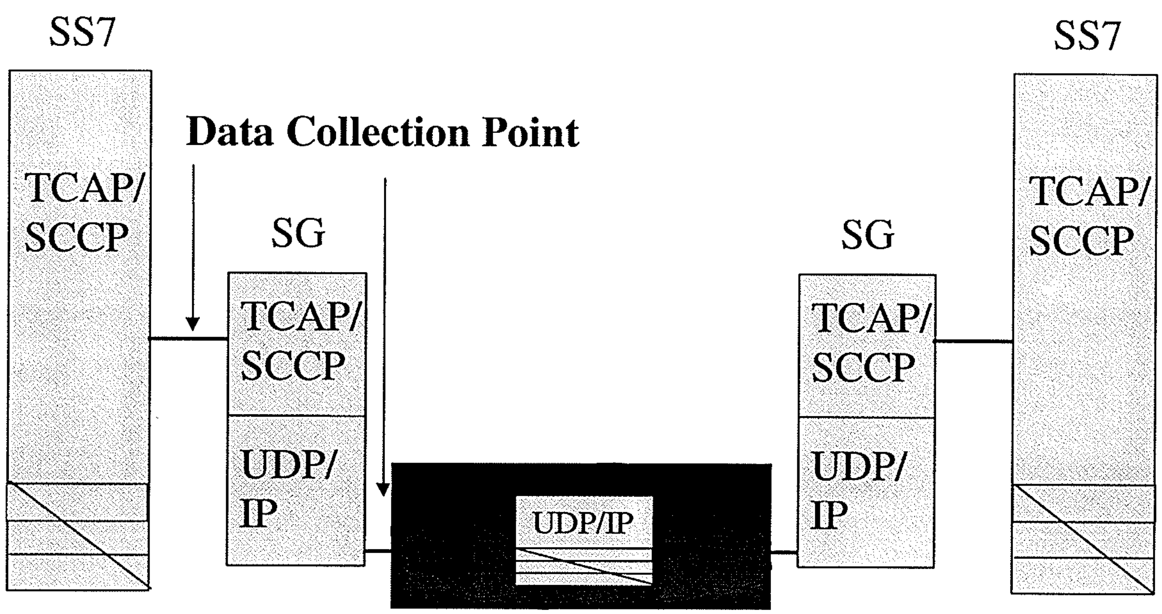


Figure 6.1: Data Collection Point

It should be noted that the proposed interworking system cannot be observed completely, i.e., one can only observe the input/output traffic through a SG while one desires to recover the hidden information about the probability distribution of its service time, delay time as well as some other performance parameters.

6.3.1 Parameters to be Measured in Simulations

The performance of a SG are identified through simulations in OPNET. Some performance parameters are measured and compared with the performance requirements given in the sections 6.2.1-6.2.7. These parameters includes:

- Service Delay (each layer)
- Queuing Delay (each buffer)
- SG Total Delay
- Packet Loss Rate (each buffer)

In addition, the throughput of a SG will also be identified in simulation, since it is a measurement of the rate at which a system can process information. It is measured as the number of bits that can be processed per second by a SG.

Chapter 7. The Performance Simulation and Analysis

7.1 Summary

In this chapter, the simulation results of the performance parameters with OPNET are provided and analyzed. They are compared with the performance requirements given in the previous chapter.

7.2 Simulations with OPNET

In the simulations, we will use OPNET, a popular computer network simulation software. Before the simulation is started, the packet arrival rate must be given. We assume that the inter-arrival time of the packets towards a SG follow an exponential distribution [Will90]. To identify how these performance parameters will change under different traffic load, the simulations will be run with three different arrival rates (denoted by G): G=10 packets/sec, G=100 packets/sec, and G=1000 packets/sec, which represent low, intermediate and high traffic loads, respectively. Each packet has around 640 bits. The buffer size of a queue is 4 packets, or around 2400 bits.

On Graph 1, a traffic generator generates the SS7 message traffic, which first goes to the buffer Q1, then goes through the MTP3 layer, the buffer Q2, the SCCP layer, the Adaptation layer, the UDP layer, and the buffer Q3, finally arrives at the IP layer.

7.3 Graphs of the Simulation Results

The simulation results of the performance parameters are given in the graphs of the CDF (cumulative distribution function), where the letter "G" denotes the packet arrival rate. See attached graph 2 - graph 40.

7.3.1 SG's Total Delay & Delay of an Individual Layer/Buffer

Graphs 2, 15 and 28 give CDFs of a SG's total delays for the packet arrival rates of 10 packets/sec, 100 packets/sec and 1000 packets/sec, respectively.

Graphs 3, 16, and 29 give CDFs of the delays at the buffer Q1 for the packet arrival rates of 10 packets/sec, 100 packets/sec and 1000 packets/sec, respectively.

Graphs 5, 18, and 31 give CDFs of the delays at the buffer Q2 for the packet arrival rates of 10 packets/sec, 100 packets/sec and 1000 packets /sec, respectively.

Graphs 9, 22, and 35 give CDFs of the delays at the buffer Q3 for the packet arrival rates of 10 packets/sec, 100 packets/sec and 1000 packets /sec, respectively.

Graphs 4, 17, and 30 give CDFs of the delays at the MTP3 layer for the packet arrival rates of 10 packets/sec, 100 packets/sec and 1000 packets /sec, respectively.

Graphs 6, 19, and 32 give CDFs of the delays at the SCCP layer for the packet arrival rates of 10 packets/sec, 100 packets/sec and 1000 packets /sec, respectively.

Graphs 7, 20, and 33 give CDFs of the delays at the Adaptation layer for the packet arrival rates of 10 packets/sec, 100 packets/sec and 1000 packets /sec, respectively.

Graphs 8, 21, and 34 give CDFs of the delays at the UDP layer for the packet arrival rates of 10 packets/sec, 100 packets/sec and 1000 packets /sec, respectively.

Graphs 10, 23, and 36 give CDFs of the delays at the IP layer for the packet arrival rates of 10 packets/sec, 100 packets/sec and 1000 packets /sec, respectively.

In each of the CDF graphs for delays, the *X*-axis represents the values of a delay in milliseconds, and *Y*-axis represents the cumulative probability of these values. A SG's total delay is composed of the delays at the buffers Q2 and Q3, the MTP3 layer, the SCCP layer, the Adaptation layer, the UDP layer and the IP layer. From the graphs of their CDFs, it can be seen that the delay at the Adaptation layer contributes the most to a SG's total delay. This is because the Adaptation layer performs the addressing translation, a key function of our TCAP over IP among the many functions performed at this layer. The least contributions to a SG's total delay are the three buffers' delays. Since a buffer's size is limited, the packet loss is inevitable when the traffic load is high. All the arriving packets to a buffer are blocked and discarded when the buffer overflows. The successfully queued packets in the buffer wait for a very short time before being extracted from the buffer, and only their queuing delays are counted during the simulation (i.e., the discarded packets do not contribute to the queuing delay). This explains why the queuing delay is very minimal compared with that of other parts.

The graphs for SG's total delays and the delays of all the layers (not including the buffer delays) show that each of their CDFs (cumulative distribution functions) can be approximated by an exponential probability distribution given by:

$$F(x) = 1 - e^{-\lambda x}, \quad x \geq 0.$$

For example, the CDF of a SG's total delay for a traffic load of 10 packets/sec (as shown in Figure 7.1) can be approximated by the exponential probability distribution of λ = 0.205 (as shown in Figure 7.2). More illustrative graphs are provided in the appendix. The value of λ is the average of the simulation data for a delay, i.e. λ can be calculated using the formula $\lambda = \frac{1}{n}\sum_{i=1}^{n} x_i$, where x_i (i=1,2,...,n) is the simulation data for a delay.

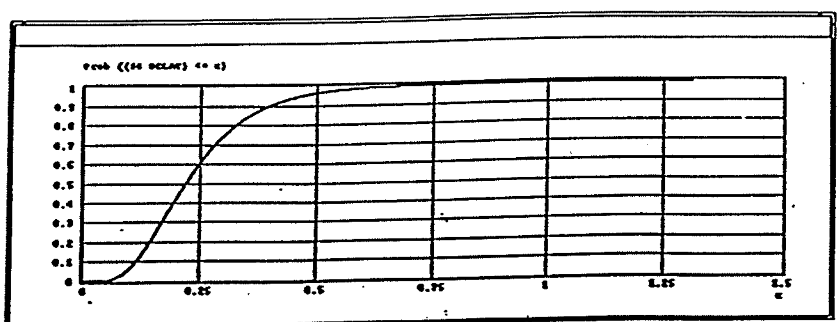


Figure 7.1: CDF of Total Delay at SG, G =10 packets/sec

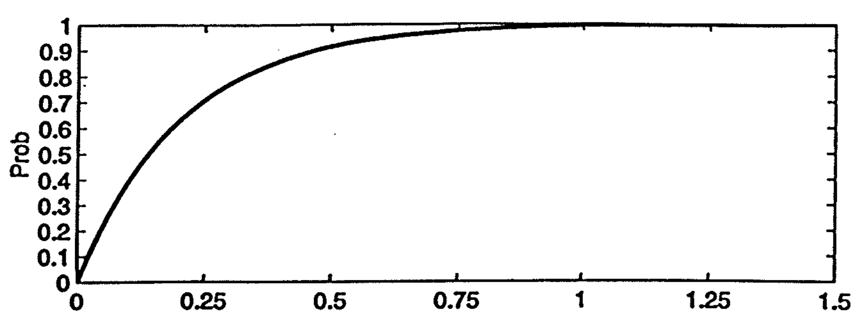


Figure 7.2: Exponential Probability Distribution of $\lambda = 0.205$

Using the formula $\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \lambda)^2}$, where x_i (i=1,2,…,n) is the simulation data for a delay, the statistical errors of the exponential distribution approximations of all non-buffer related delays can be calculated and they are usually less than 15%. The statistical error of the exponential distribution approximation of a SG's total delay is bigger than that of all the individual MTP, SCCP, Adaptation, UDP and IP layers, because it consists of all the individual statistical errors of these layers.

When the arrival rate is 10 packets/sec of low traffic load, the SG total delay is less than 15 ms with a 90% probability. After the rate increases to 100 packets/sec of intermediate traffic load, the SG total delay is less than 30 ms with a 90% probability; and less than 50 ms with 90% probability for the rate of 1000 packets/sec of high traffic load.

7.3.2 Packet Loss Rate

Graphs 11, 24 and 37 give the packet loss rates at the buffer Q1 for the packet arrival rates of 10 packets/sec, 100 packets/sec and 1000 packets /sec, respectively.

Graphs 12, 25 and 38 give the packet loss rates at the buffer Q2 for the packet arrival rates of 10 packets/sec, 100 packets/sec and 1000 packets /sec, respectively.

Graphs 13, 26 and 39 give the packet loss rates at the buffer Q3 for the packet arrival rates of 10 packets/sec, 100 packets/sec and 1000 packets /sec, respectively.

In each of the graphs of packet loss rate, the *X*-axis represent the time elapse in seconds after a simulation is started, and *Y*-axis represents packet loss rate in a unit of 10^{-3}. It is shown in the graphs that the packet loss rate at a buffer increments with the increase of the arrival rate. For the buffer Q1, The average packet loss rate is around 0.009×10^{-3} for the packet arrival rate of 10 packets/sec, is around 0.018×10^{-3} for the packet arrival rate of 100 packets/sec, and is around 0.048×10^{-3} for the packet arrival rate of 1000 packets/sec. For all the three packet arrival rates, the packet loss rates at the buffers Q2 and Q3 remain at nearly 0% all the time after a simulation starts. This suggests that after some packets are blocked and discarded at the buffer Q1, the traffic load towards buffers Q2 and Q3 be significantly reduced so that almost all the incoming packets towards them can be successfully queued and processed. It also suggests that the packet loss rate of the whole SG mainly be determined by the buffer Q1.

7.3.3 Throughput

Graphs 14, 27 and 40 give a SG's total throughputs for the packet arrival rates of 10 packets/sec, 100 packets/sec and 1000 packets /sec, respectively.

In a throughput's CDF graph, the *X*-axis represents the values of a throughput in 1000 bits/sec, and *Y*-axis represents the cumulative probability of these values.

According to the throughput formula:

$$Throughput = \frac{Total\_Number\_of\_Bits\_Transferred}{SG\_Total\_Delay} \quad ,$$

the throughput decrements with the increase of the total delay if the total number of bits transferred is fixed. This has been proved by the Graphs 14, 27 and 40, keeping in mind that a packet's size in the simulation is fixed to be around 640 bits.

Chapter 8. Conclusions and Further Work

For the interworking of TCAP/SS7 and TCP/IP, this thesis has studied the protocol conversion and the management and control functions, and has proposed the solutions against it. The simulation results with OPNET mostly meet the performance requirements for TCAP in SS7 network. However, there are still some other issues to be studied in the future, for example, if a TCAP application needs the support of SCCP connectionless protocol class 1, in which the messages must be delivered in sequence, we may need the support of TCP in stead of UDP. But whether the interworking with TCP is able to meet the real-time requirement of SS7 becomes a question, since TCP is much more complicated than UDP and it usually takes much more time for transmission than UDP?

A new protocol called SCTP (simple control transport protocol) has been issued against the interworking of SS7 and TCP/IP [Stew00] [Side00]. SCTP is an application-level datagram transfer protocol designed to transport SS7 signaling messages over IP networks, but capable of broader application. It can provide in-sequence message delivery, and can operate on the top of an unreliable datagram service such as UDP. In summary, SCTP offers the following services to its user:

- Acknowledged error-free and duplicate-free transfer of user data.
- Application-level segmentation to conform to discovered MTU (maximum transfer unit) size.
- Sequenced delivery of user datagrams within multiple streams, with an option for in-sequence delivery of individual datagrams.

- Optional multiplexing of user datagrams into SCTP datagrams, subject to MTU size restrictions.
- Enhanced reliability through support of multi-homing at either or both ends of the association.

The basic service provided by SCTP is the reliable transfer of user datagrams between peer SCTP users. It is a layer between the SCTP user application and an unreliable end-to-end datagram service.

How to employ SCTP to provide a reliable data transmission across SS7 and TCP/IP networks are still being studied in the telecom industry. Some companies have been developing the API (application programming interface) designs and source codes for it. Combining SCTP and the TCAP/IP mechanism proposed in the thesis needs further study.

References

[Ansm92] ANSI T1.111-1992, Signaling System No. 7 (SS7) Message Transfer Part (MTP), 1992.

[Anss92] ANSI T1.112-1992, Signaling System No. 7 (SS7) Signaling Connection Control Part (SCCP), 1992.

[Anst92] ANSI T1.114-1992, Signaling System No. 7 (SS7) Transaction Capabilities Application Part (TCAP), 1992.

[Atai97] Atai, A. and Gordon, J. "Architectural Solutions to Internet Congestion Based on SS7 and Intelligent Network Capabilities", a White Paper, *Telcordia Technologies (www.telcordia.com)*, 1997.

[Bhat97] Bhat, U. N., Miller, G. K., and Rao, S. S. "Statistical Analysis of Queueing Systems", *in Frontiers in Queuing: Models and Applications In Science and Engineering*, ed. J. H. Dshalalow, CRC Press, pp.351-394, 1997.

[Blac97] Black, U. *ISDN and* SS7. New Jersey: Prentice Hall, 1997.

[Boss97] Bosse, J. G. van. *Signaling in Telecommunications Network.* New York: Wiley-Interscience, 1997.

[Burk97] Burk, R. , Bligh, M. and Lee, Thomas. *TCP/IP Blueprints*, 1st Ed. Indianapolis: Sams Publishing, 1997.

[Cisc98] "Cisco SS7 Dial Access Solution System Integration Guidelines", *Cisco Systems (www.cisco.com)*, 1998.

[Higg97] Higginbottom, Gary N. *Performance Evaluation of Communication Networks,* Boston: Artech House, 1997.

[Itub89] ITU-T Recommendation Q706, "Signaling System No. 7 - Message Transfer Part Signaling Performance", *ITU-T*, Geneva, 1993.

[Lela94] Leland, W. E., Taqqu, M. S., Willinger, W. and Wilson, D. V. "On the Self-Similar Nature of Ethernet Traffic", *IEEE/ACM Transactions on Networking*, vol.2, pp. 1-15, 1994.

[Luce98] "Internet Call Diversion", Technical Backgrounder, *Lucent Technologies (www.lucent.com)*, 1998.

[Mato98] Matousek, J. and Ong, L. "Bay Network SS7-Internet Access Signaling Protocol", Internet-Draft: draft-long-ss7-signal-00.txt, *IETF*, January 1999.

[Mill94] Miller, Mark. *Analzing Broadband Networks, New York: M&T Books, 1994.*

[Nort98] "CVX SS7 Gateway", Product Brief, *Nortel Networks (www.nortelnetworks.com)*, 1998

[Pawl98] Pawlak, M. "Nonparametric Estimation of the G/G/1 Queueing System", *Advances in Matrix Analytic Methods for Stochastic Models*, ed. A. Alfa and S. Chakravarthy, Notable Publications, pp. 279-294, 1998.

[Paxo95] Paxon, V. and Floyd, S. "Wide-Area Traffic: The Failure of Poisson Modeling", *IEEE/ACM Transactions on Networking*, vol. 3, pp. 226-244, 1995.

[Post81] Postel, J. "Internet Control Message Protocol (ICMP)", RFC 792, *IETF* 1981.

[Seth99] Seth, T., Broscius, A., Huitema, C. and Lin, H. P., "Performance Requirements for TCAP Signaling Internet Telephony", Internet-Draft, Bellcore, *IETF*, Feb 1999.

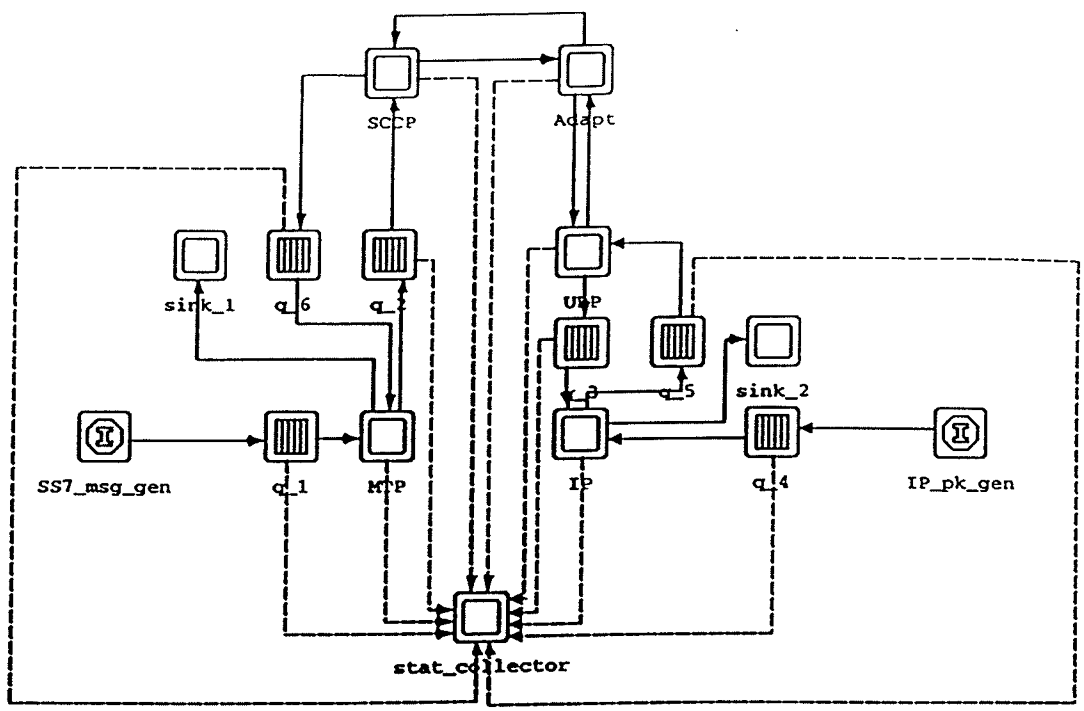
[Side00] Sidebottom, G., Ong, L., Mousseau, Guy and Rytina, Ian. "SS7 MTP3-User Adaptation Layer (M3UA)", Internet Draft: draft-ietf-sigtran-m3ua-02.txt, *IETF*, March 2000.

[Stew00] Stewar, R. R., Xi, Q., Morneaul, K., Shar, C. and Schwarzbaue, H. J. "Simple Control Transmission Protocol (SCTP)", Internt Draft: draft-ietf-sigtran-sctp-10.txt, *IETF*, June 2000.

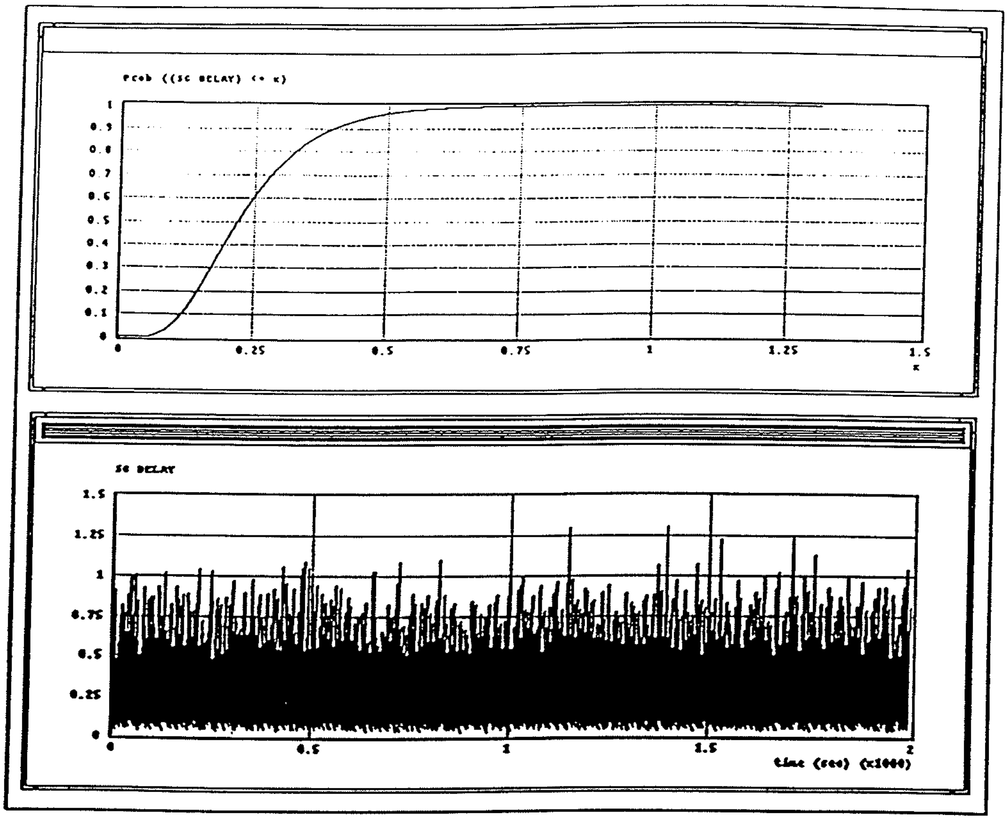
[Tane96] Tanenbaum, Andrew S. *Computer Networks*, 3rd Ed. New Jersey: Prentice Hall, 1996.

[Will90] Willmann, G. and Kuhn, P. J. "Performance Modeling of Signaling System No.7", *IEEE Communication Magazine*, pp. 44-56, July 1990.

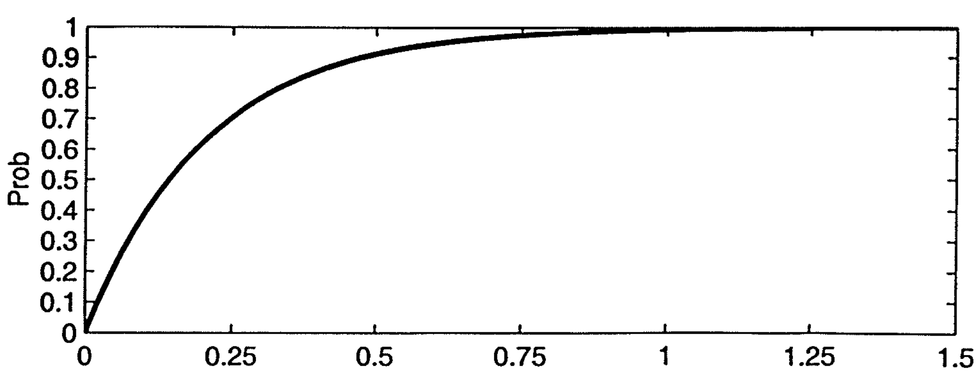
[3com98] "Leveraging the Intelligence of SS7 to improve IP-based Remote Access and other IP Services", a Technical Paper, *3Com (www.3com.com)*, 1998.



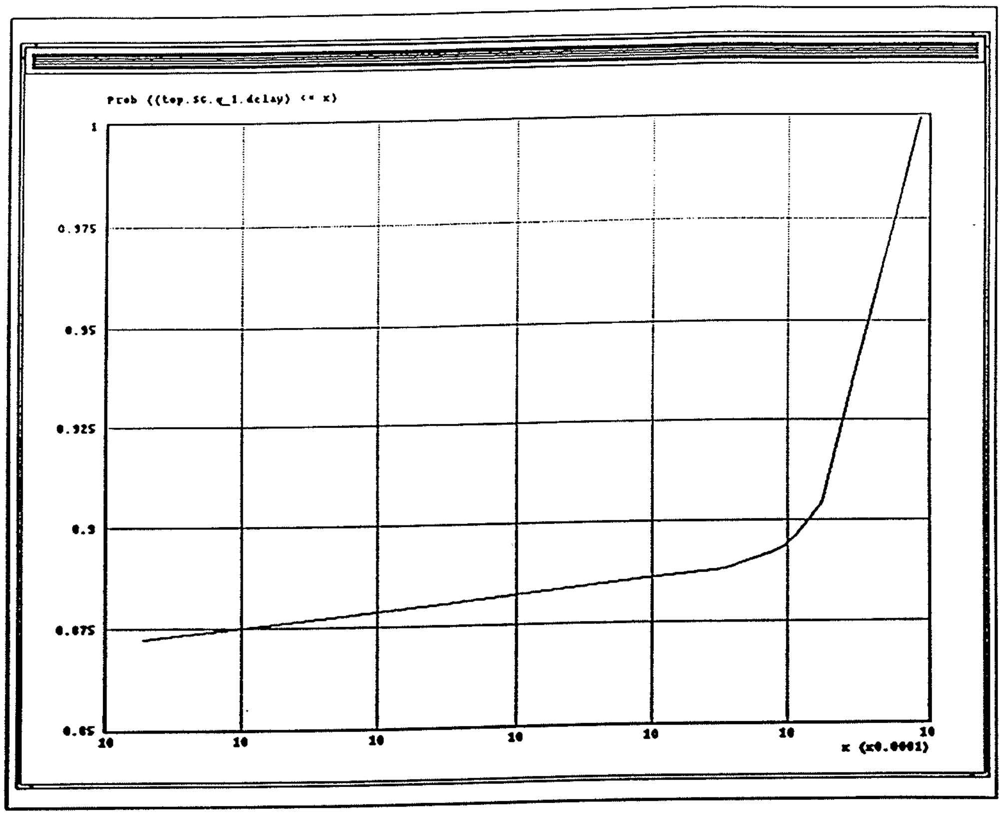
Graph 1: Opnet Signaling Gateway Architecture



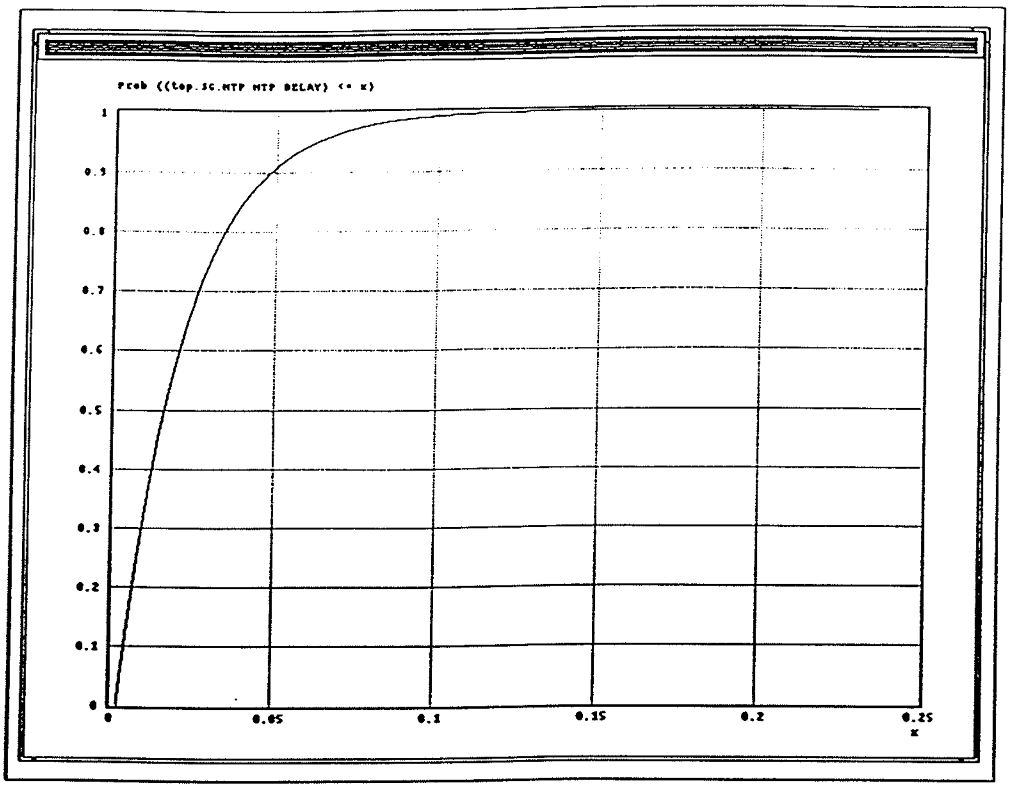
Graph 2: CDF of Total Delay at SG, G = 10 packets/sec



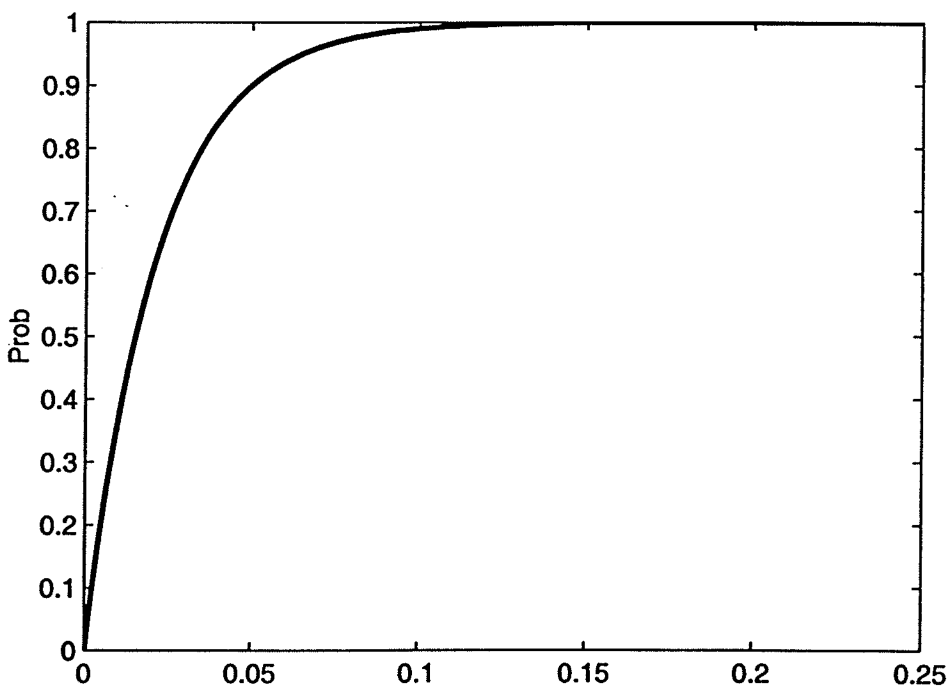
Graph 2a: Exponential Distribution of $\lambda = 0.205$



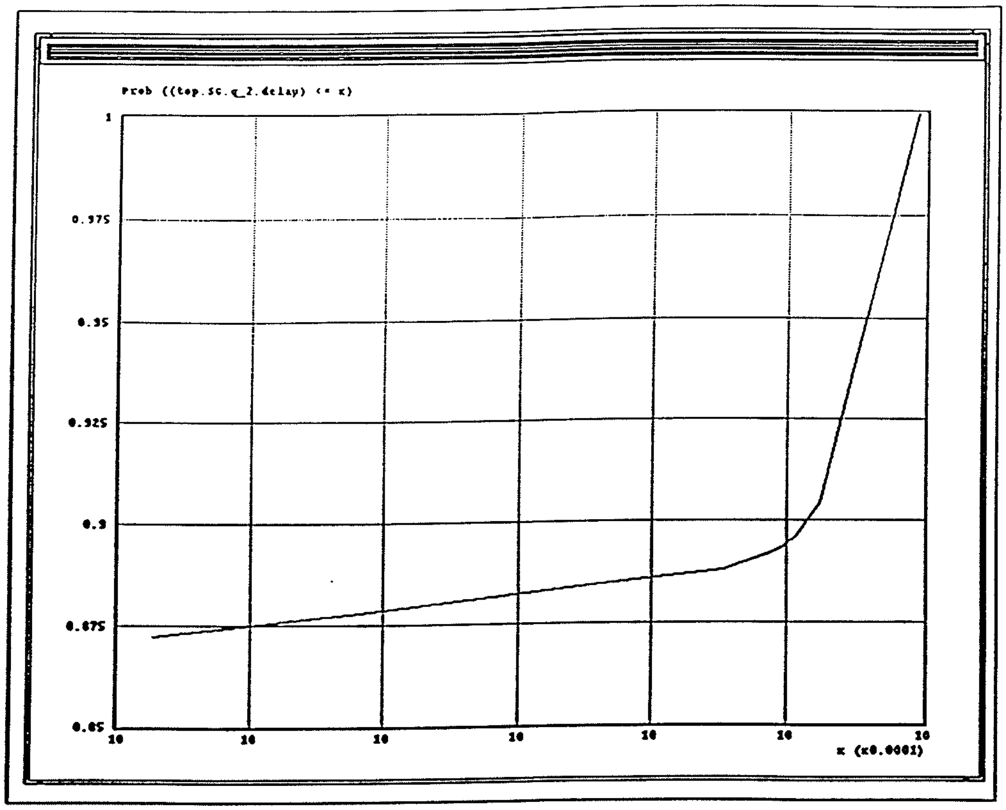
Graph 3: CDF of Delay at Buffer Q1, G = 10 packets/sec



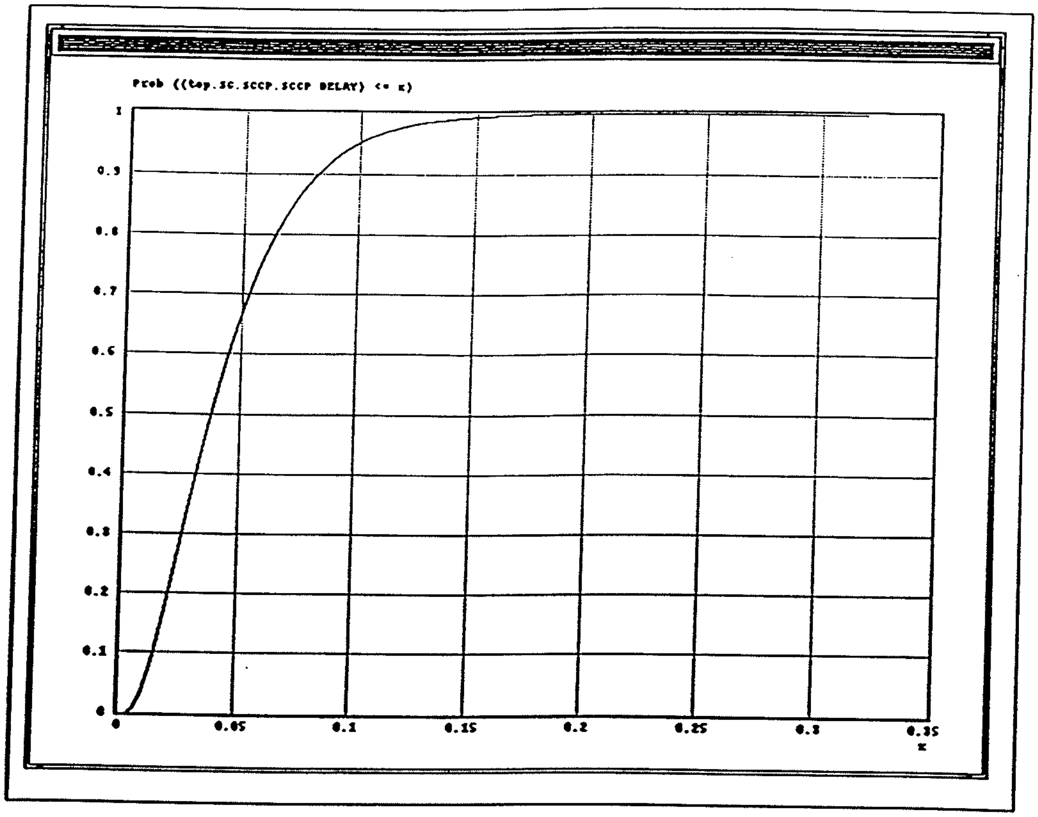
Graph 4: CDF of Delay at MTP Layers, G = 10 packets/sec



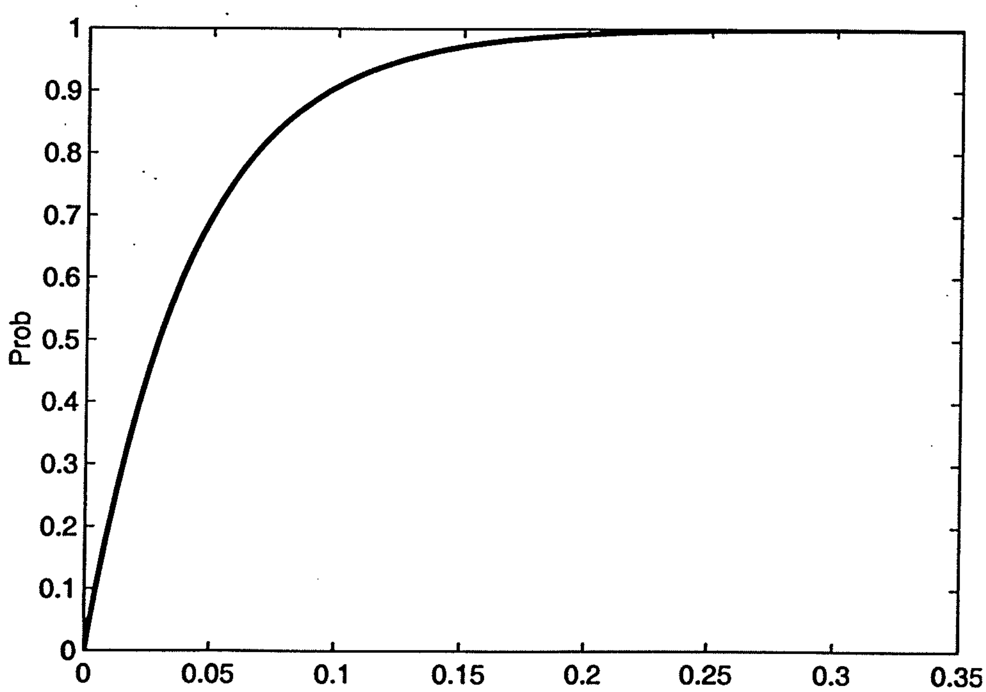
Graph 4a: Exponential Distribution of $\lambda = 0.0217$



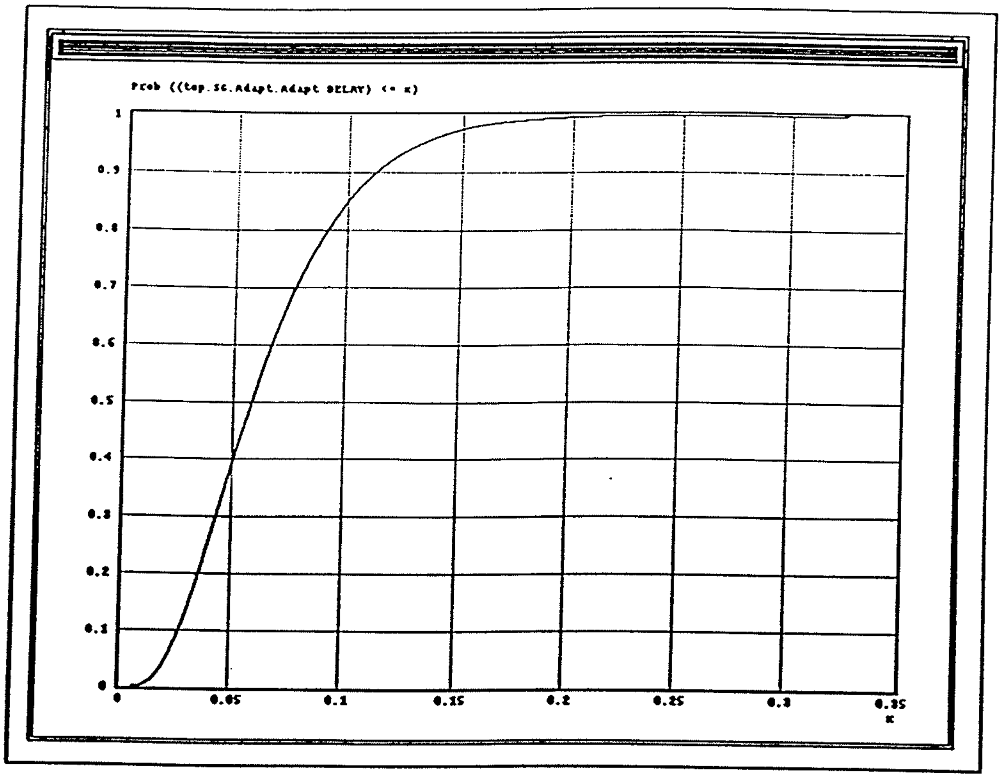
Graph 5: CDF of Delay at Buffer Q2, G = 10 packets/sec



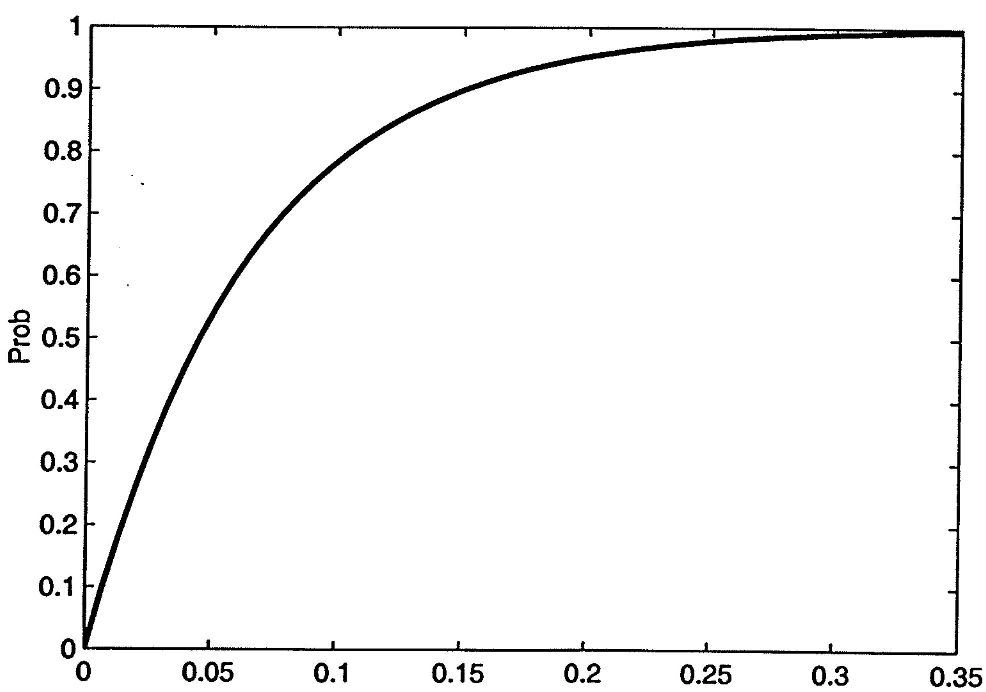
Graph 6: CDF of Delay at SCCP Layer, G = 10 packets/sec



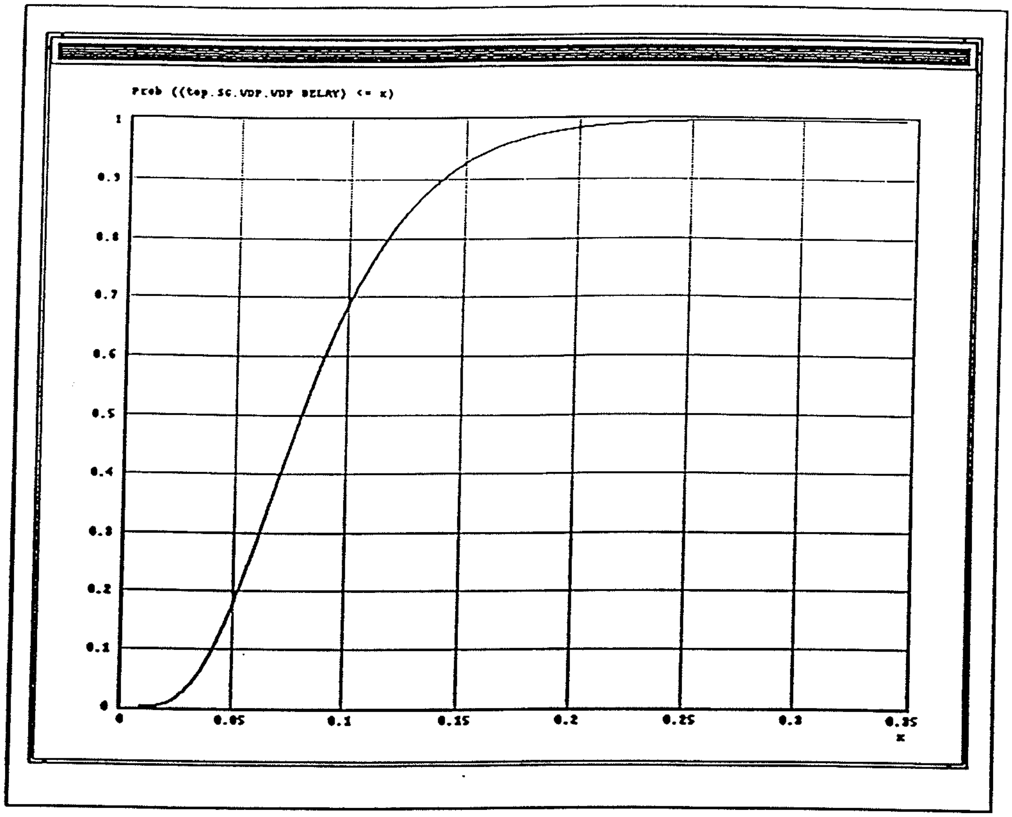
Graph 6a: Exponential Distribution of $\lambda = 0.042$



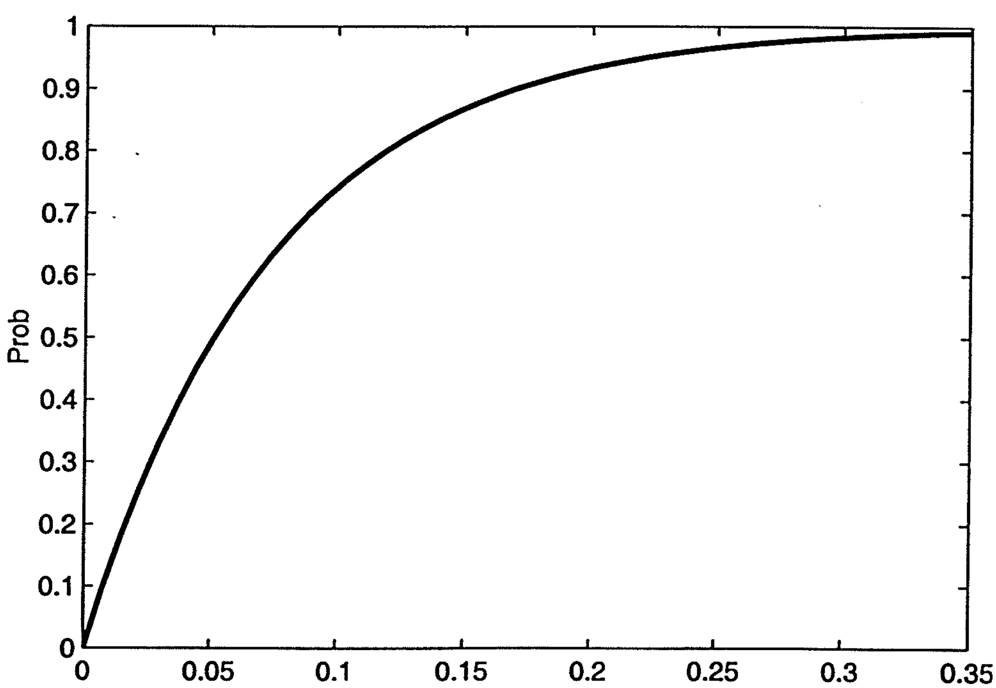
Graph 7: CDF of Delay at Adaptation Layer, G = 10 packets/sec



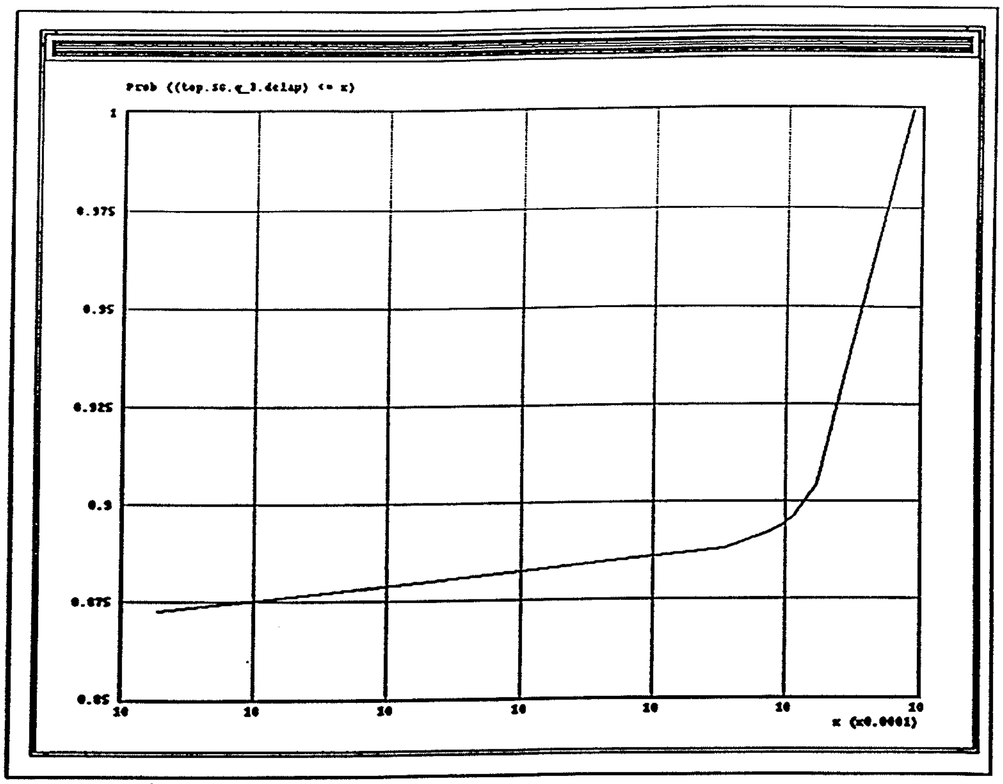
Graph 7a: Exponential Distribution of $\lambda = 0.0652$



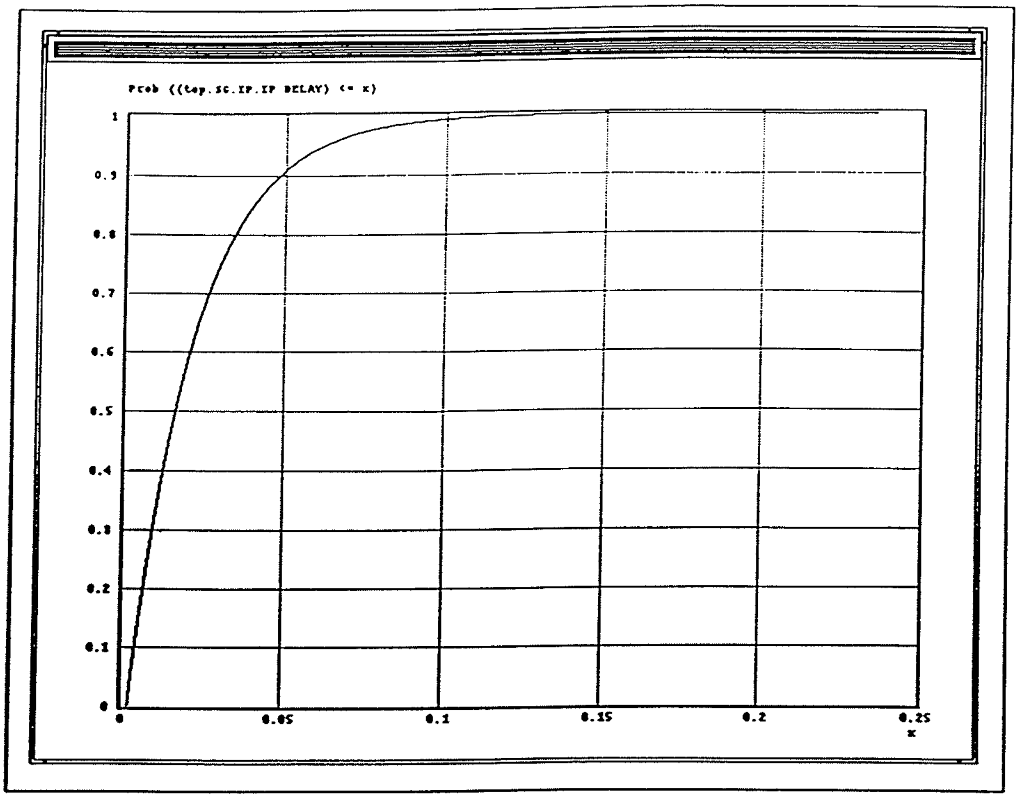
Graph 8: CDF of Delay at UDP Layer, G = 10 packets/sec



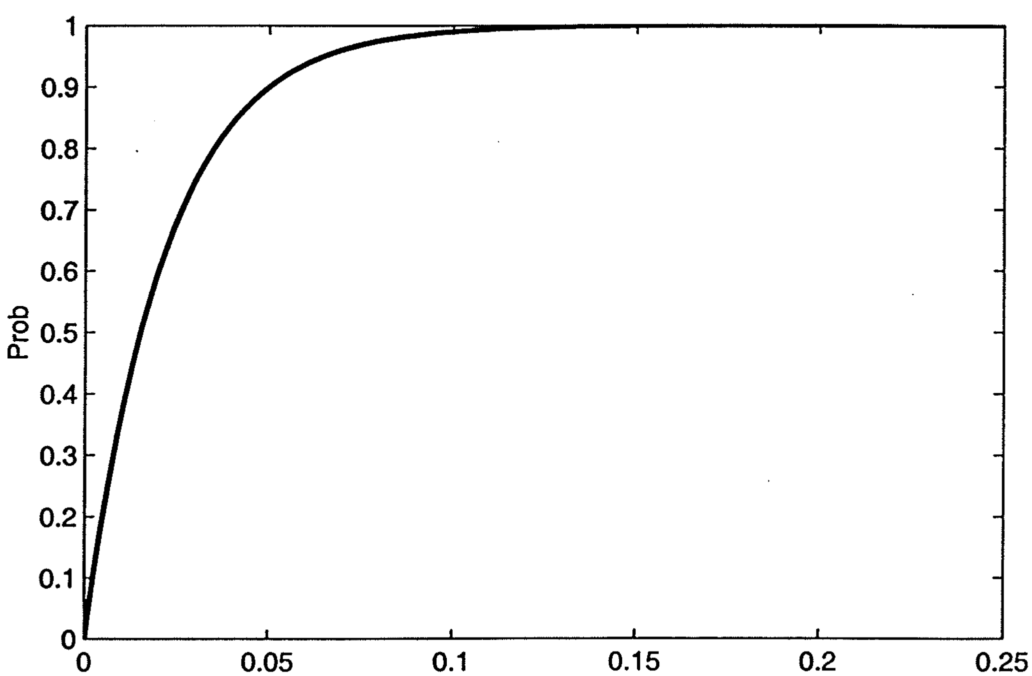
Graph 8a: Exponential Distribution of $\lambda = 0.0739$



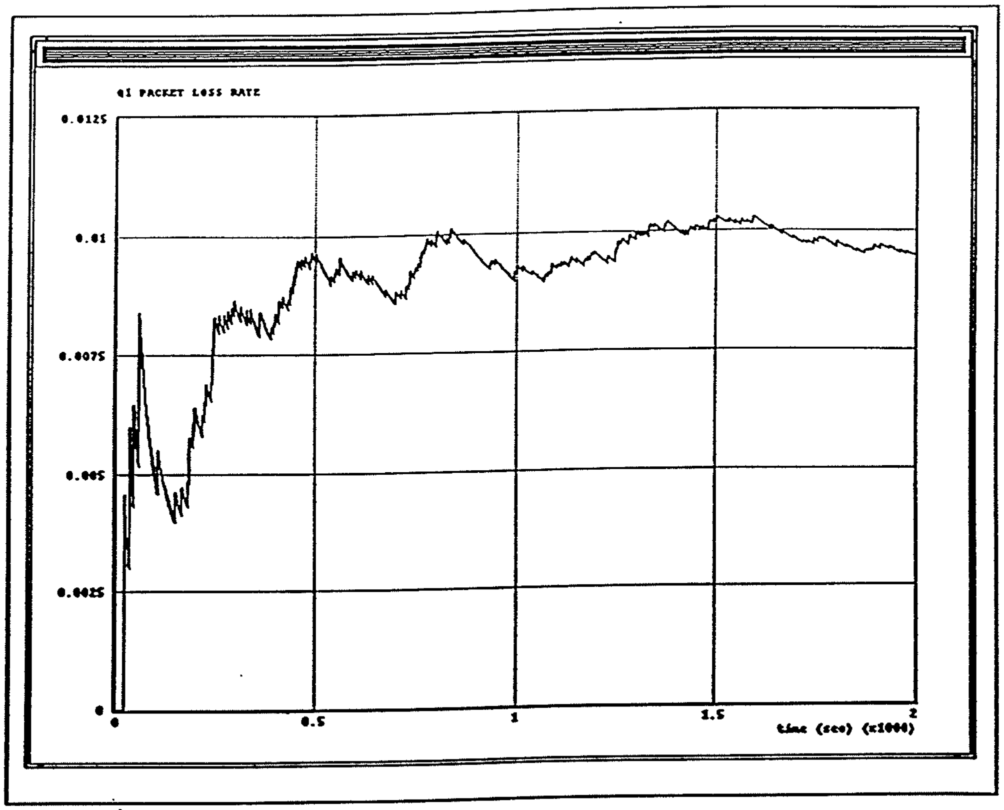
Graph 9: CDF of Delay at Buffer Q3, G = 10 packets/sec



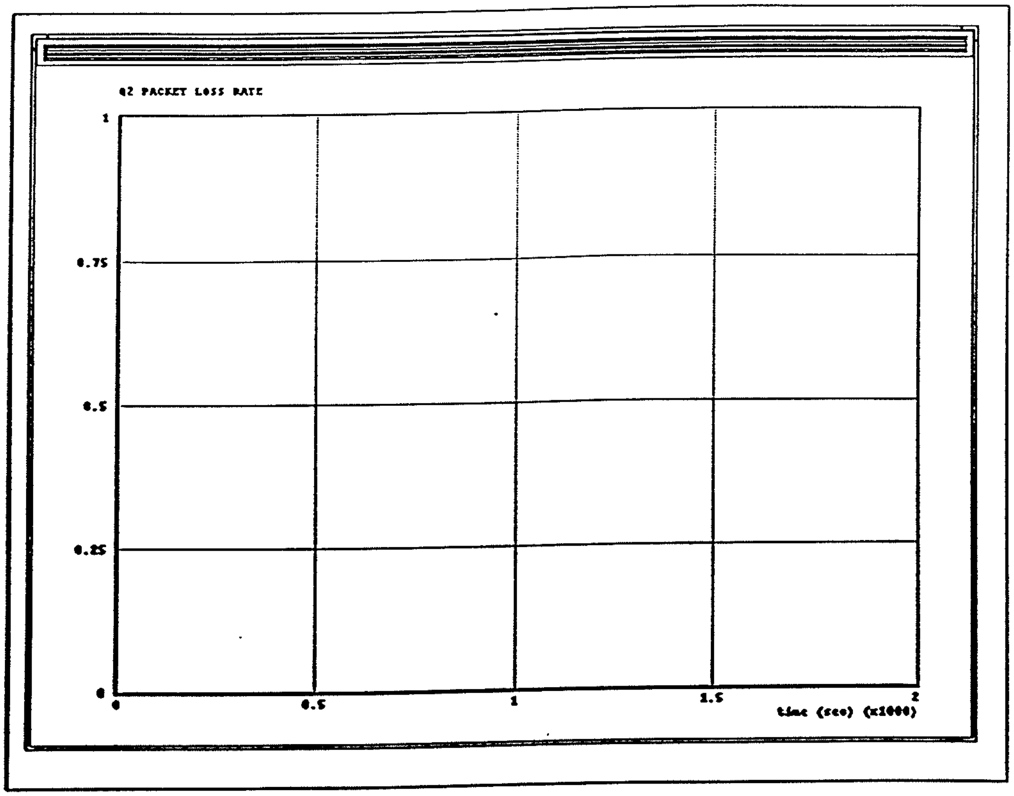
Graph 10: CDF of Delay at IP Layer, G = 10 packets/sec



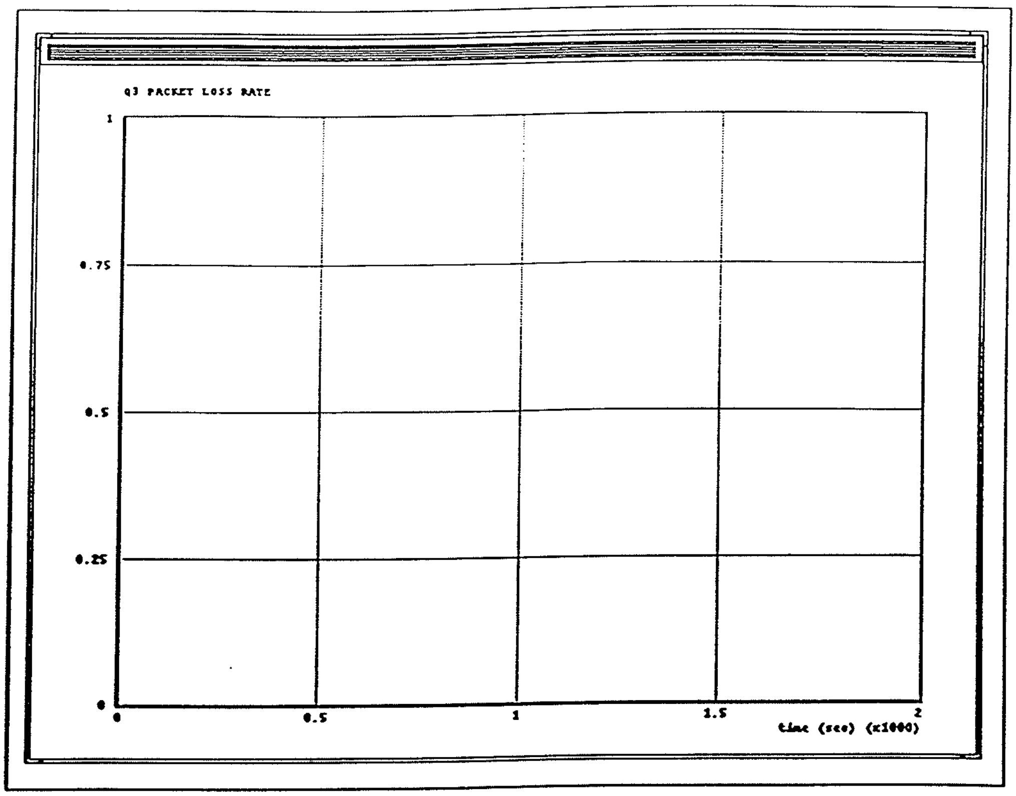
Graph 10a: Exponential Distribution of $\lambda = 0.0217$



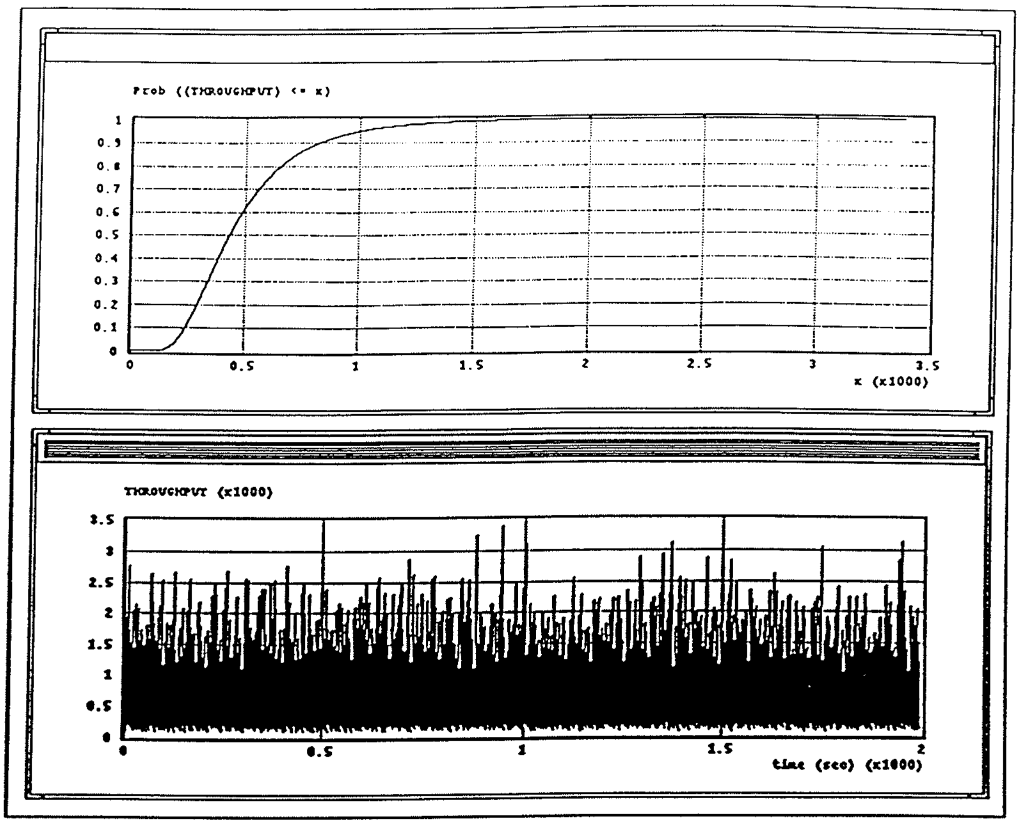
Graph 11: Packet Loss Rate at Buffer Q1, G = 10 packets/sec



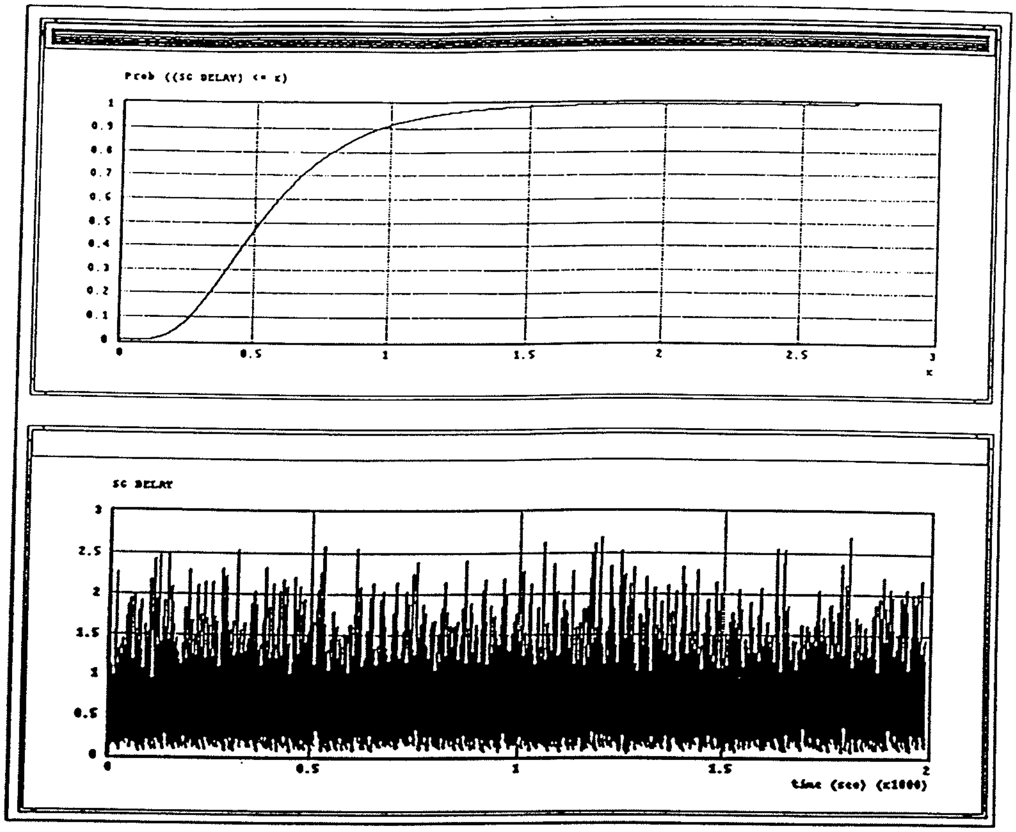
Graph 12: Packet Loss Rate at Buffer Q2, G = 10 packets/sec (it remains zero throughout the simulation)



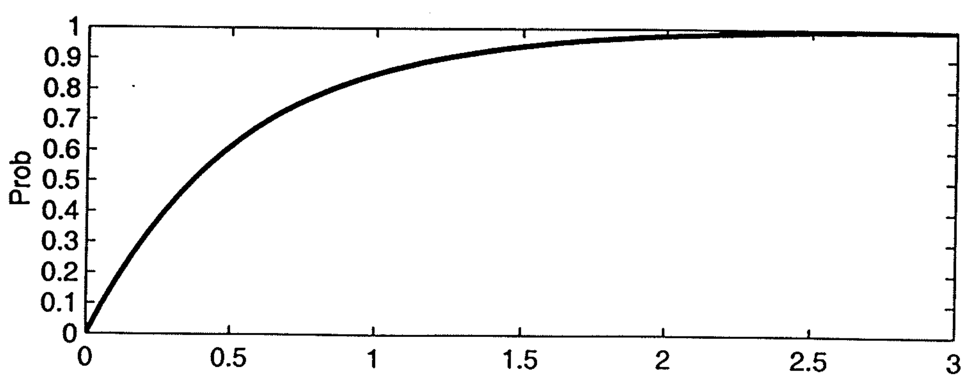
Graph 13: Packet Loss Rate at Buffer Q3, G = 10 packets/sec (it remains zero throughout the simulation)



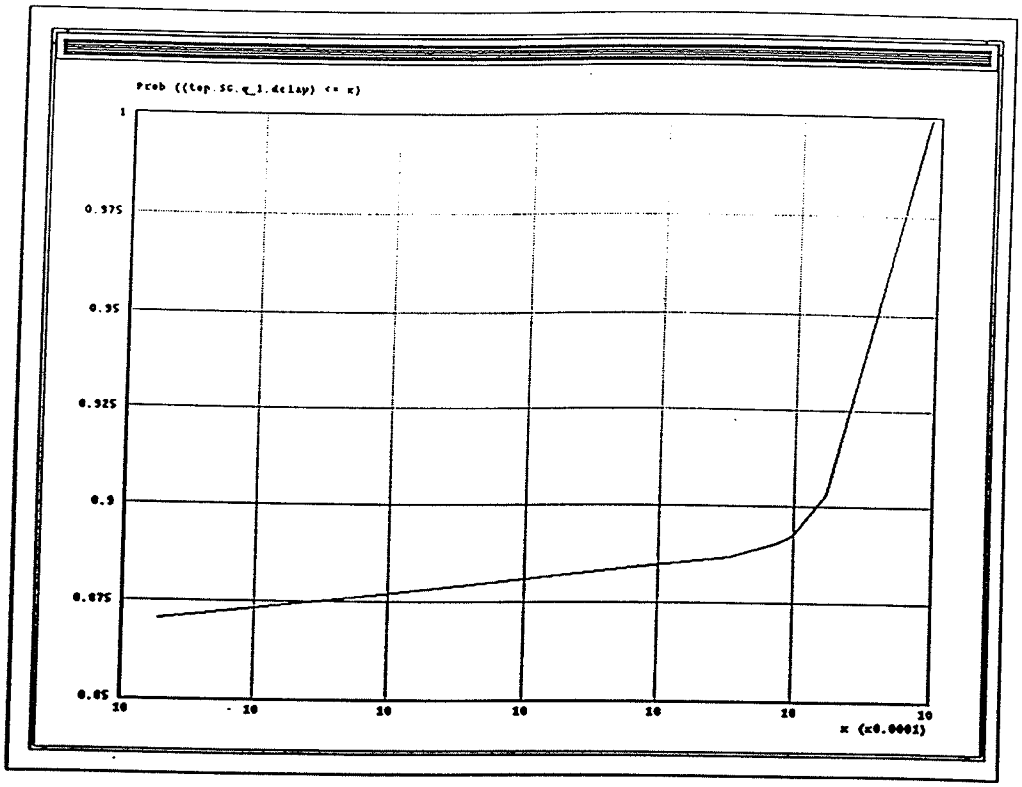
Graph 14: CDF of SG Throughput, G = 10 packets/sec



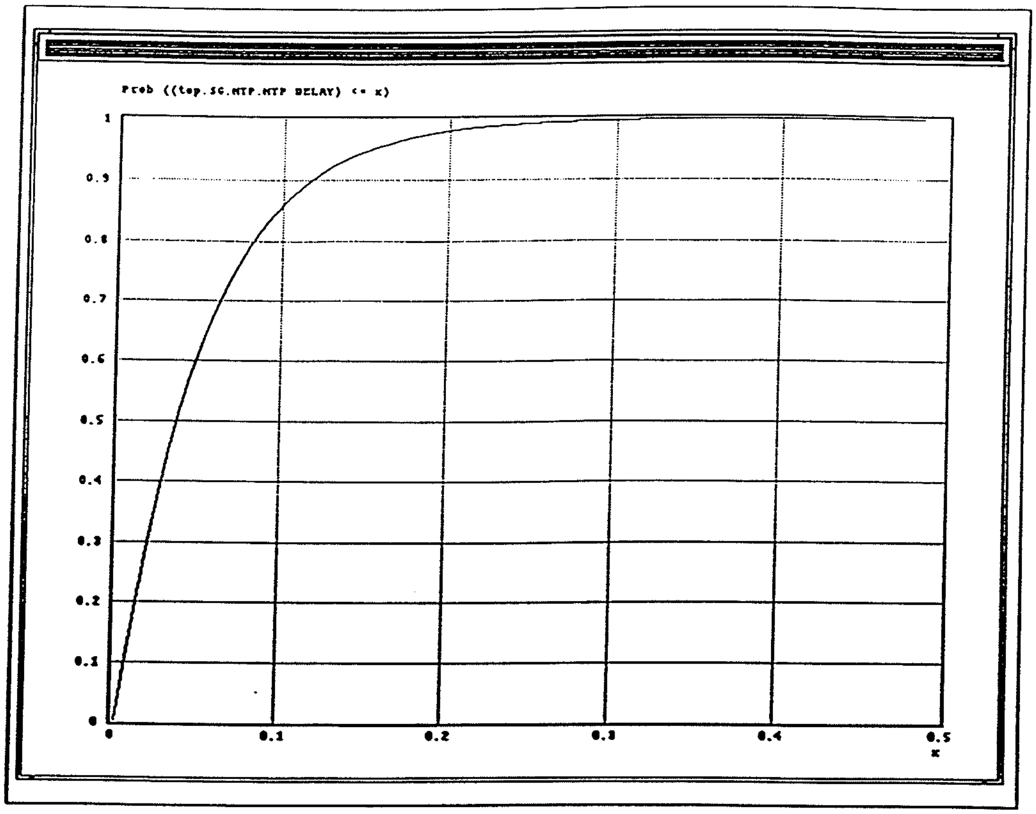
Graph 15: CDF of Total Delay at SG, G = 100 packets/sec



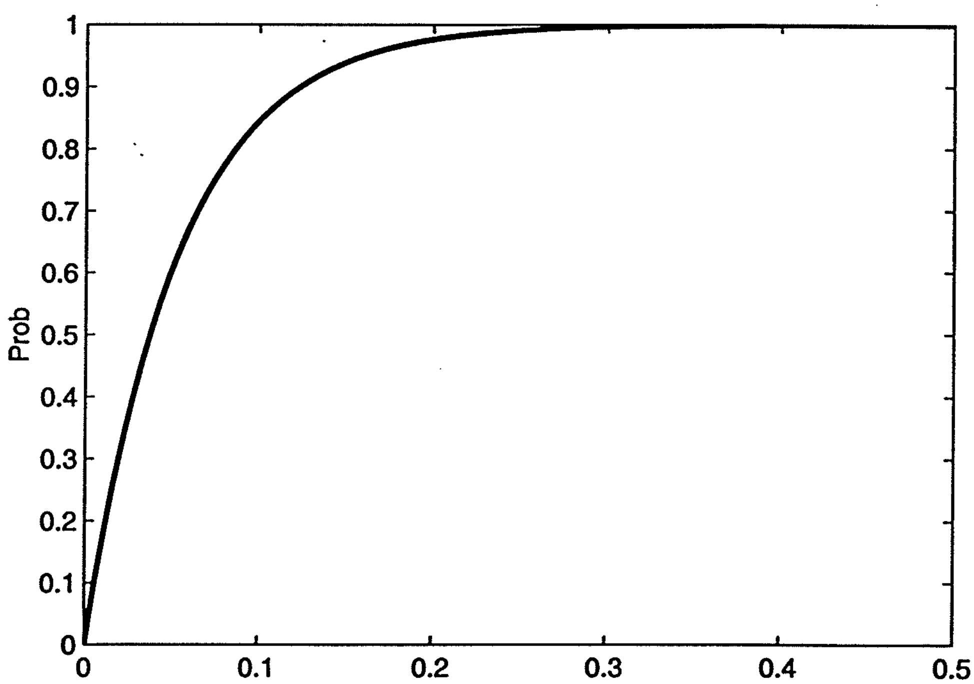
Graph 15a: Exponential Distribution of $\lambda = 0.527$



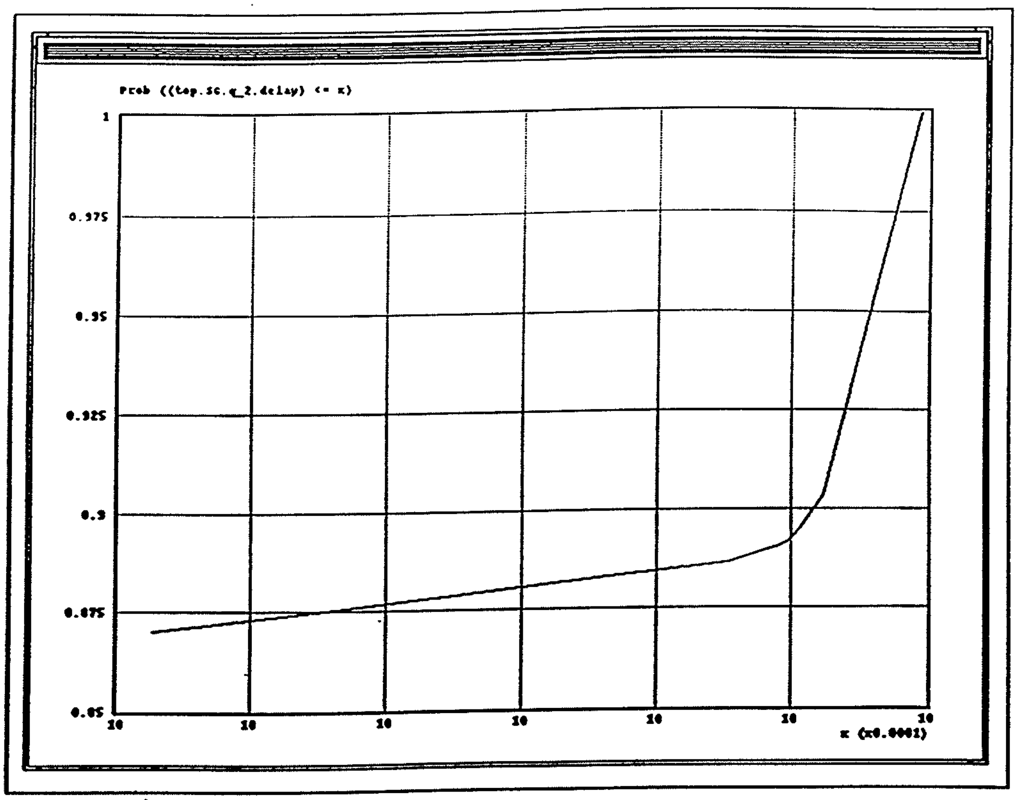
Graph 16: CDF of Delay at Buffer Q1, G = 100 packets/sec



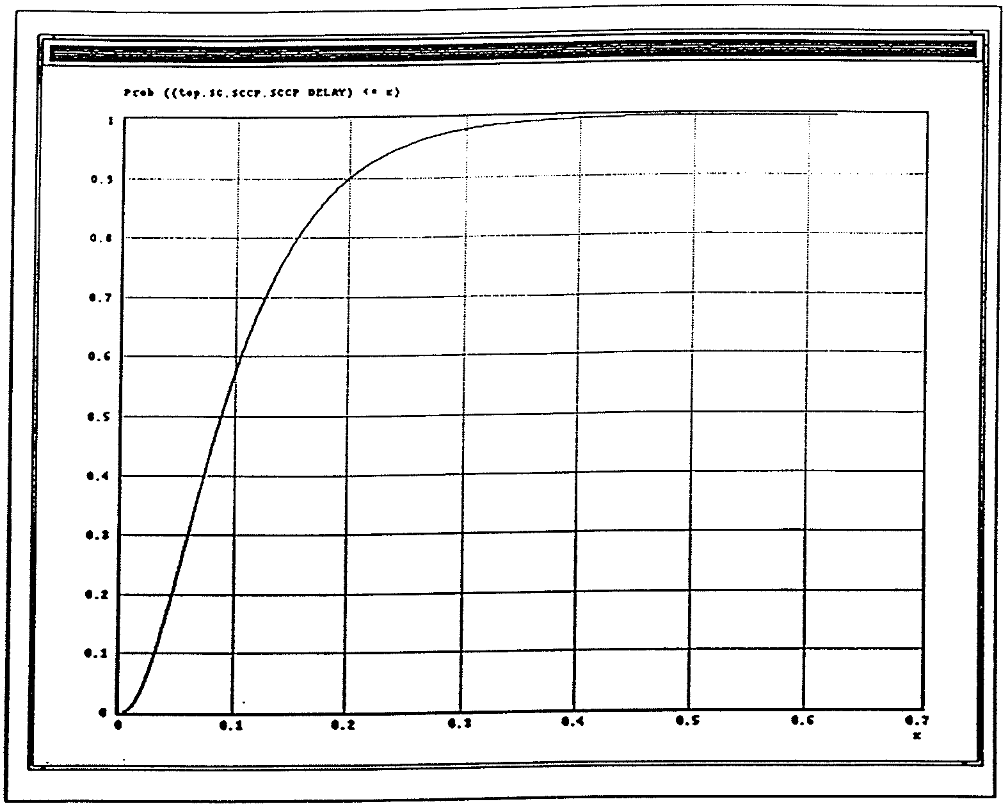
Graph 17: CDF of Delay at MTP Layers, G = 100 packets/sec



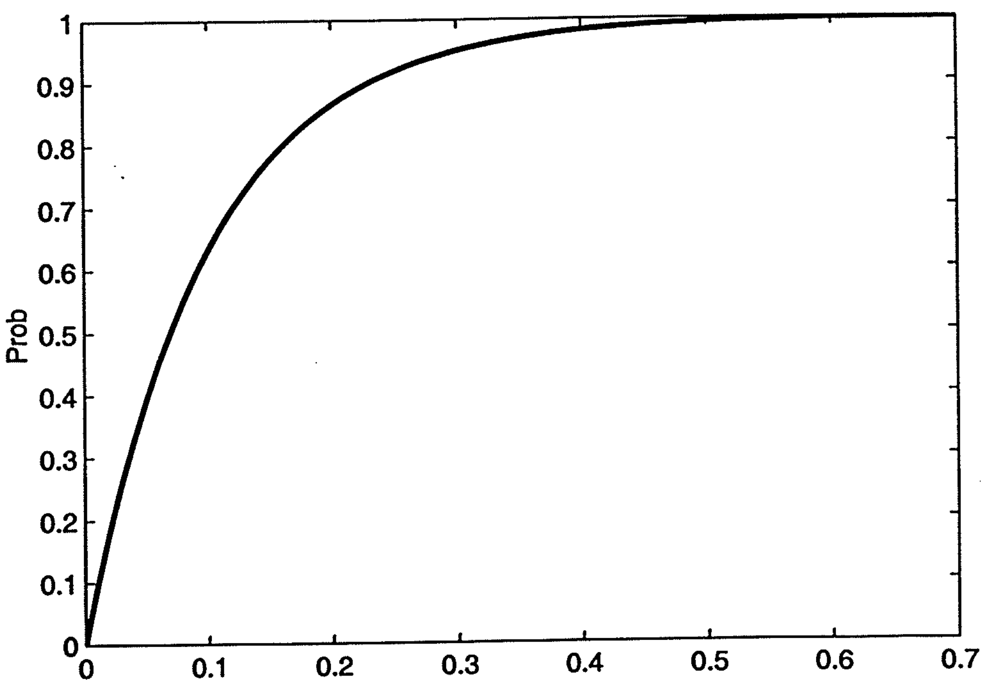
Graph 17a: Exponential Distribution of $\lambda = 0.0536$



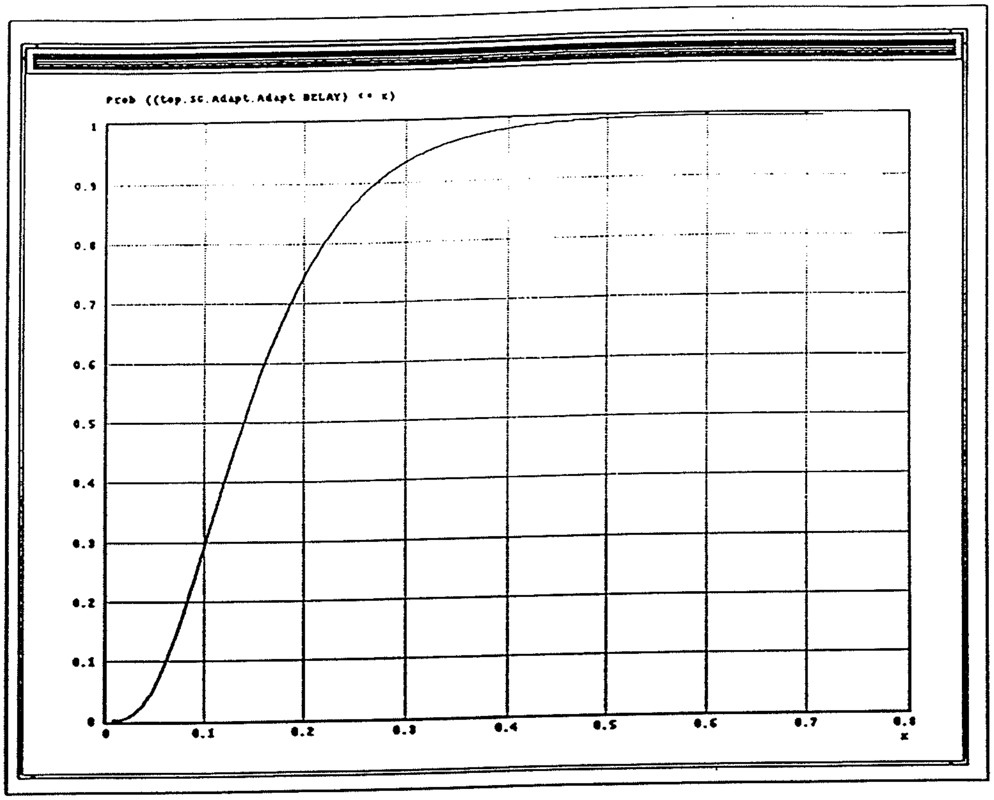
Graph 18: CDF of Delay at Buffer Q2, G = 100 packets/sec



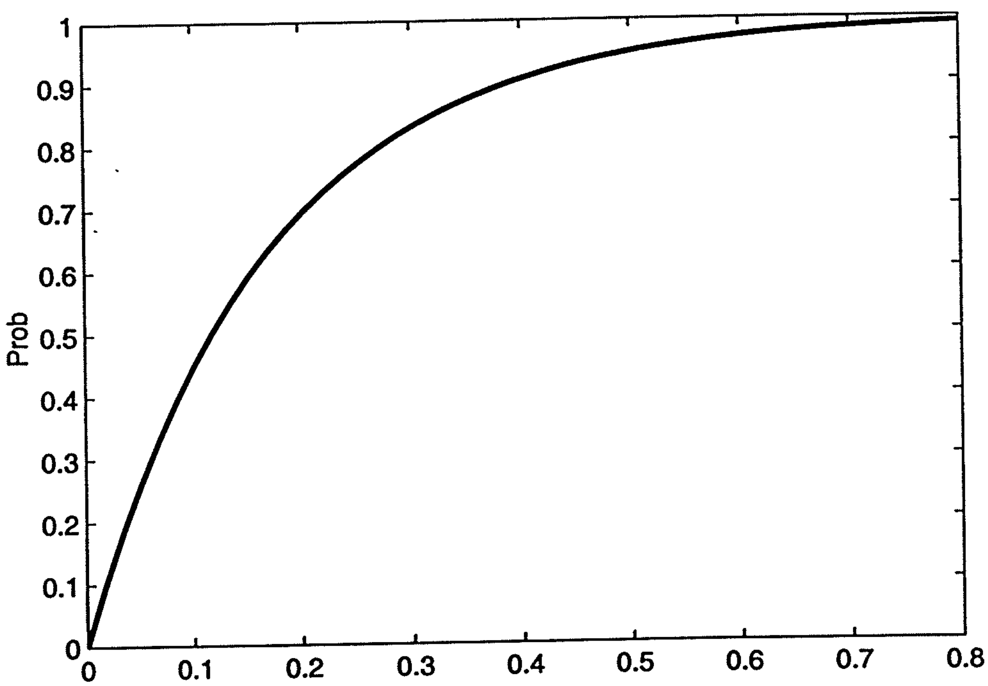
Graph 19: CDF of Delay at SCCP Layer, G = 100 packets/sec



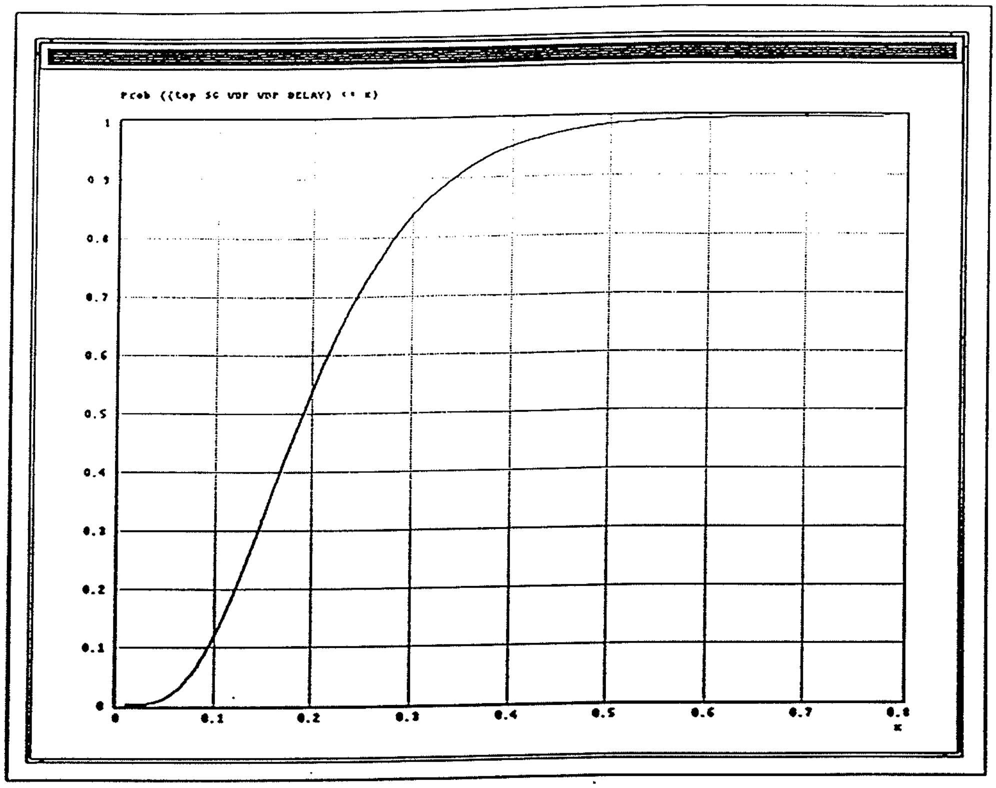
Graph 19a: Exponential Distribution of $\lambda = 0.1$



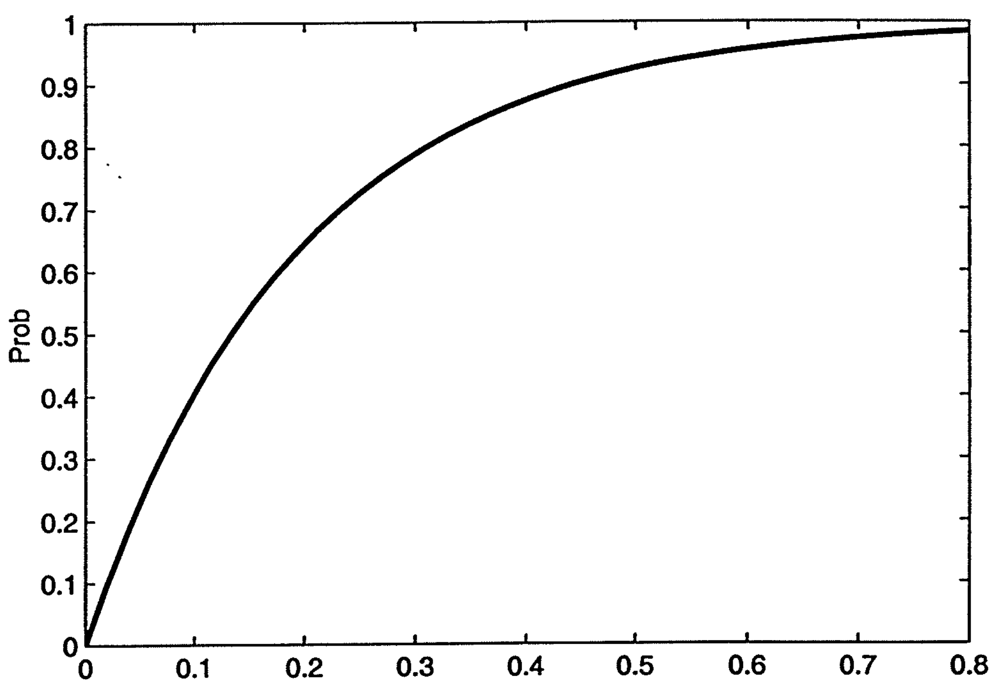
Graph 20: CDF of Delay at Adaptation Layer, G = 100 packets/sec



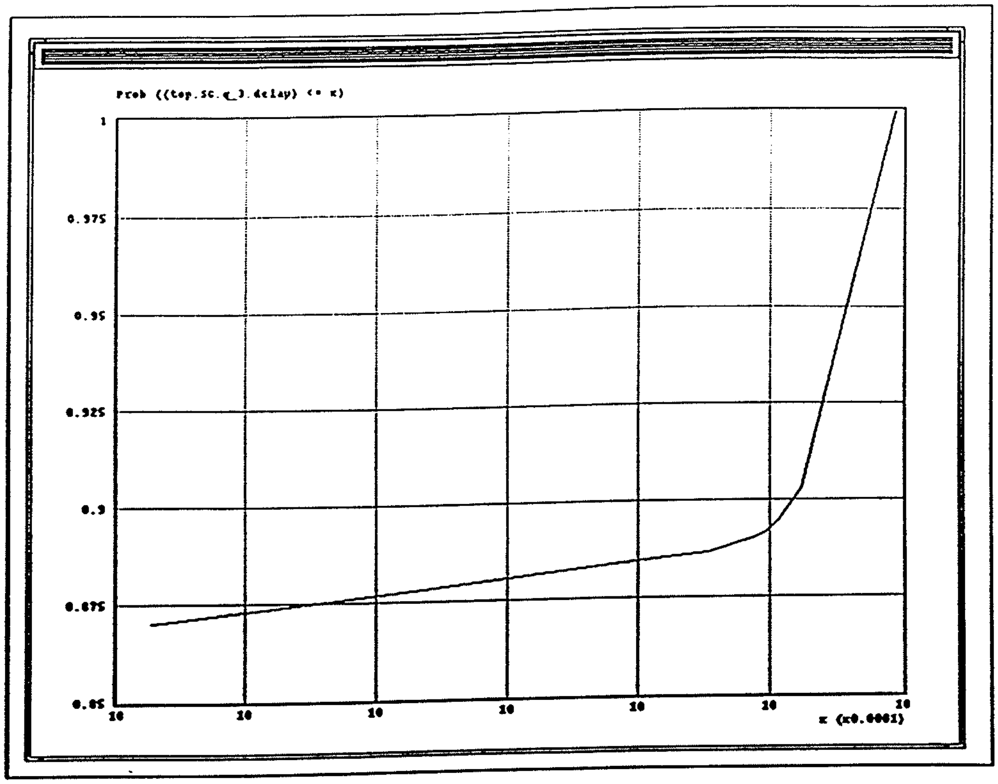
Graph 20a: Exponential Distribution of $\lambda = 0.1696$



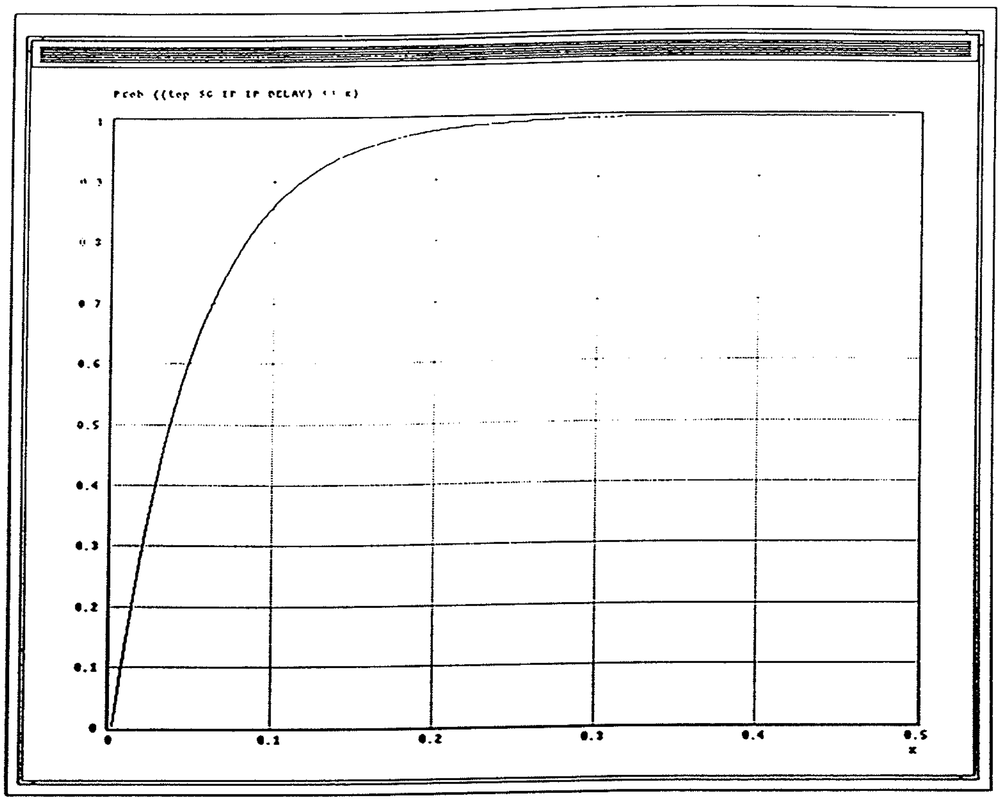
Graph 21: CDF of Delay at UDP Layer, G = 100 packets/sec



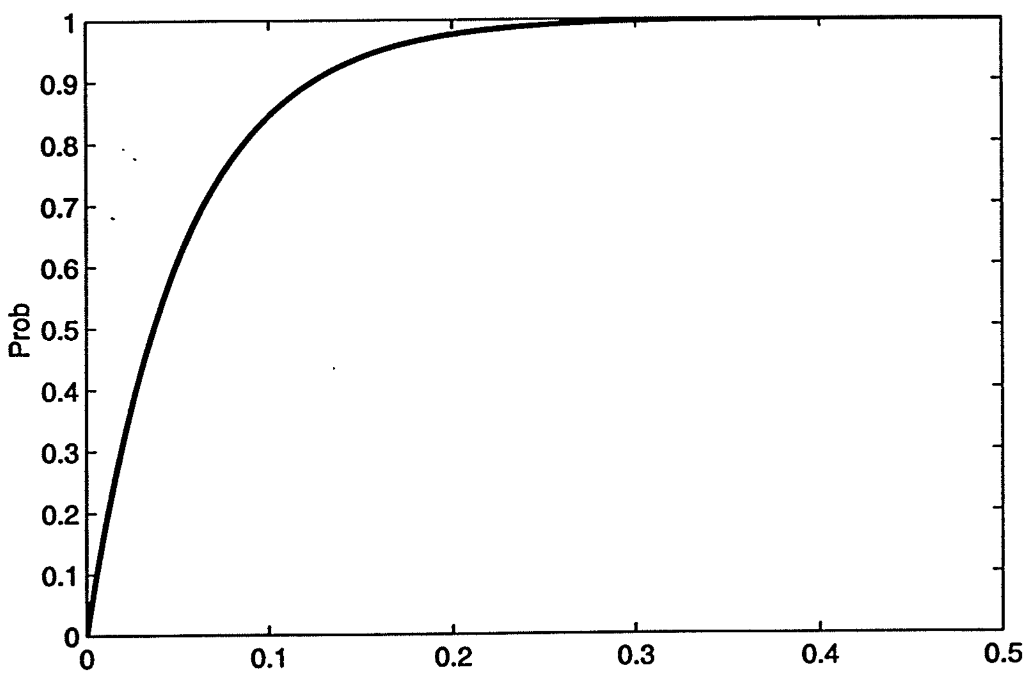
Graph 21a: Exponential Distribution of $\lambda = 0.1928$



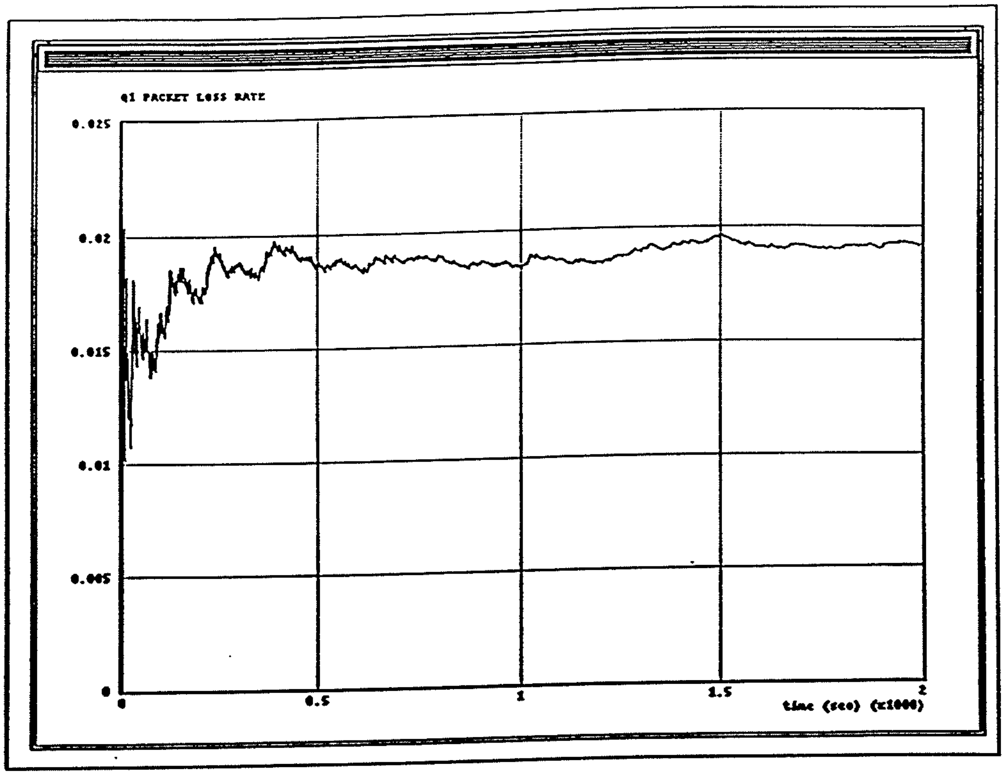
Graph 22: CDF of Delay at Buffer Q3, G = 100 packets/sec



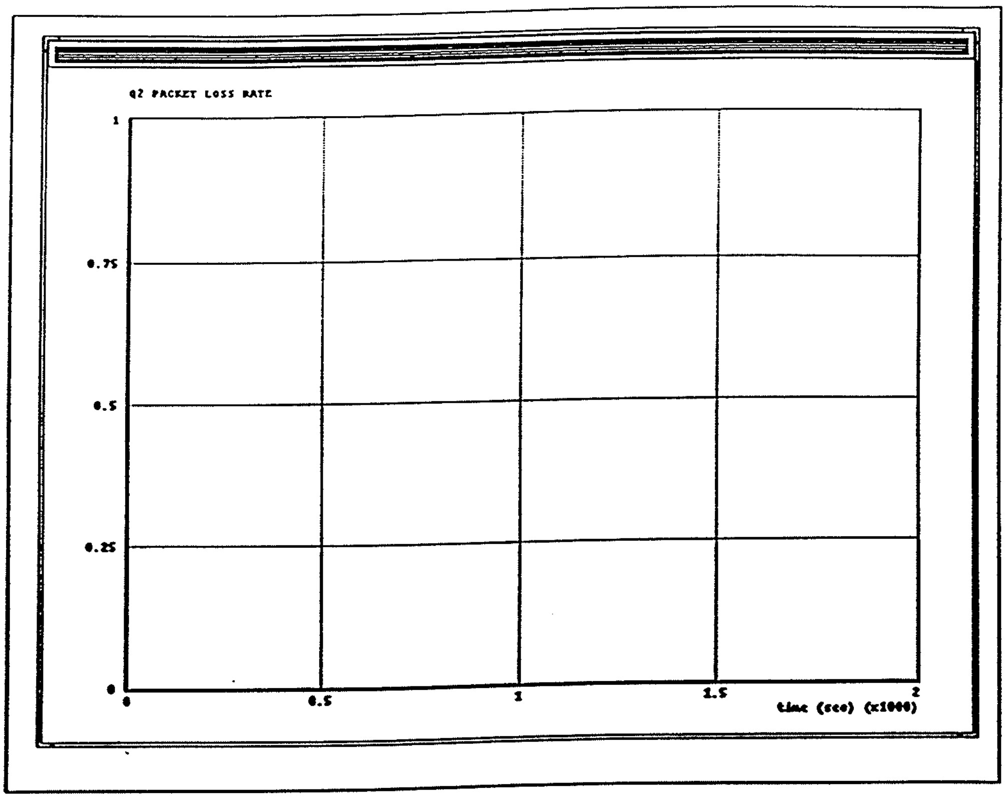
Graph 23: CDF of Delay at IP Layer, G = 100 packets/sec



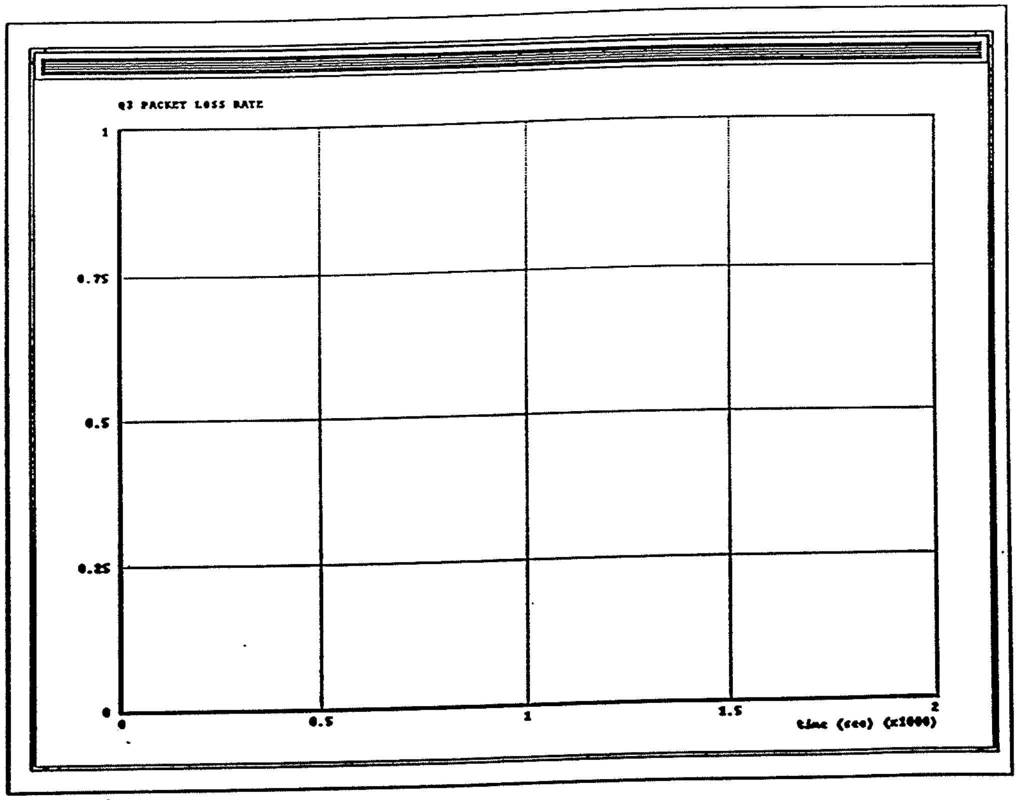
Graph 23a: Exponential Distribution of $\lambda = 0.0536$



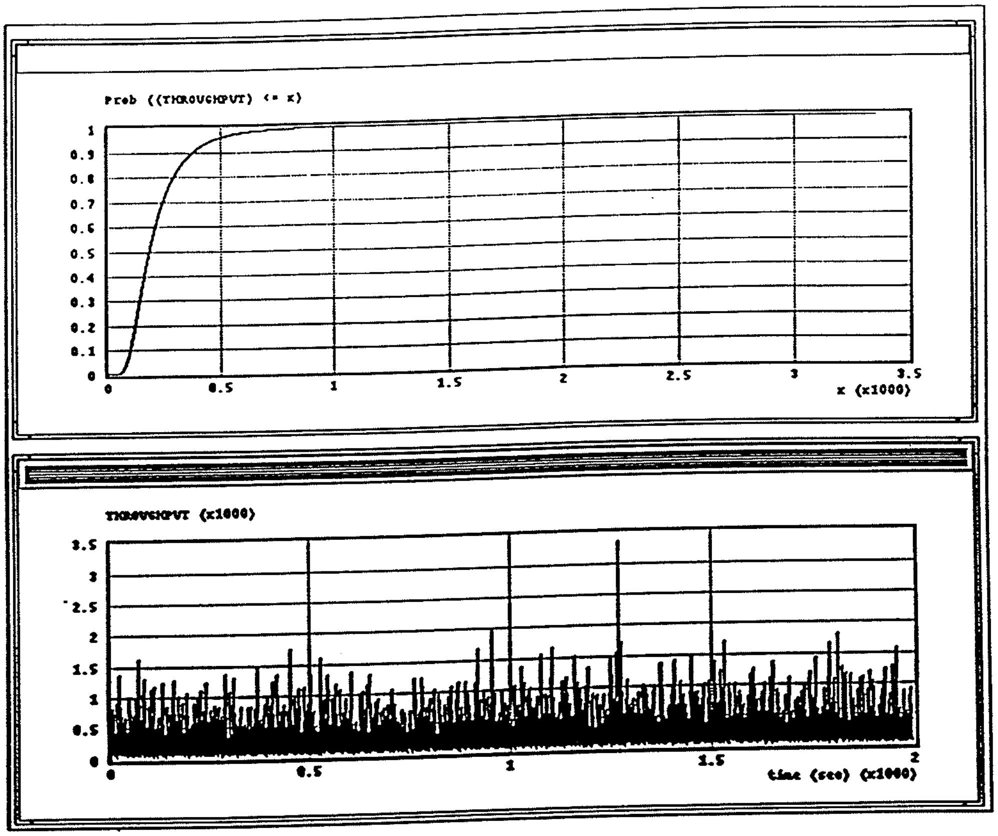
Graph 24: Packet Loss Rate at Buffer Q1, G = 100 packets/sec



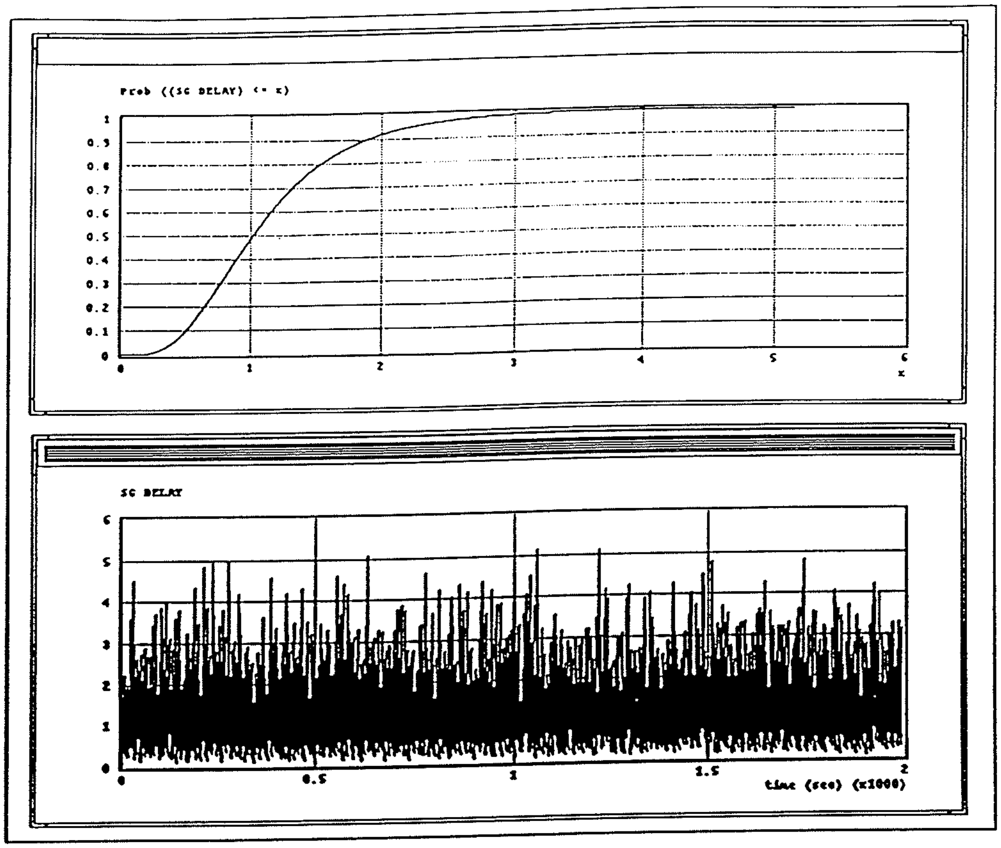
Graph 25: Packet Loss Rate at Buffer Q2, G = 100 packets/sec (it remains zero throughout the simulation)



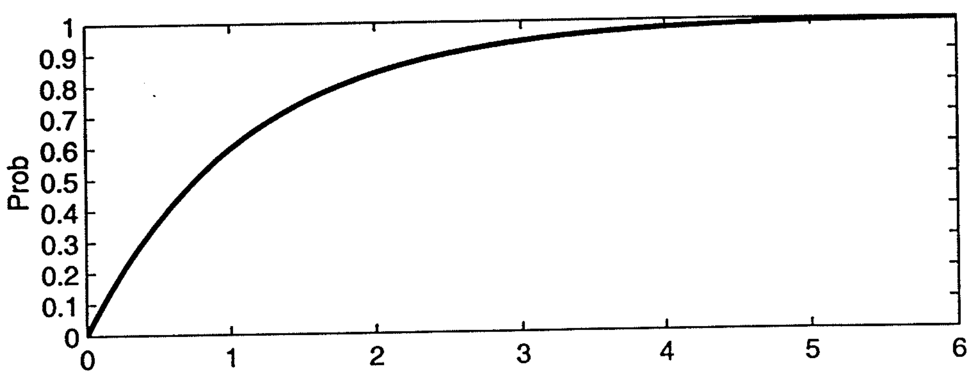
Graph 26: Packet Loss Rate at Buffer Q3, G = 100 packets/sec (it remains zero throughout the simulation)



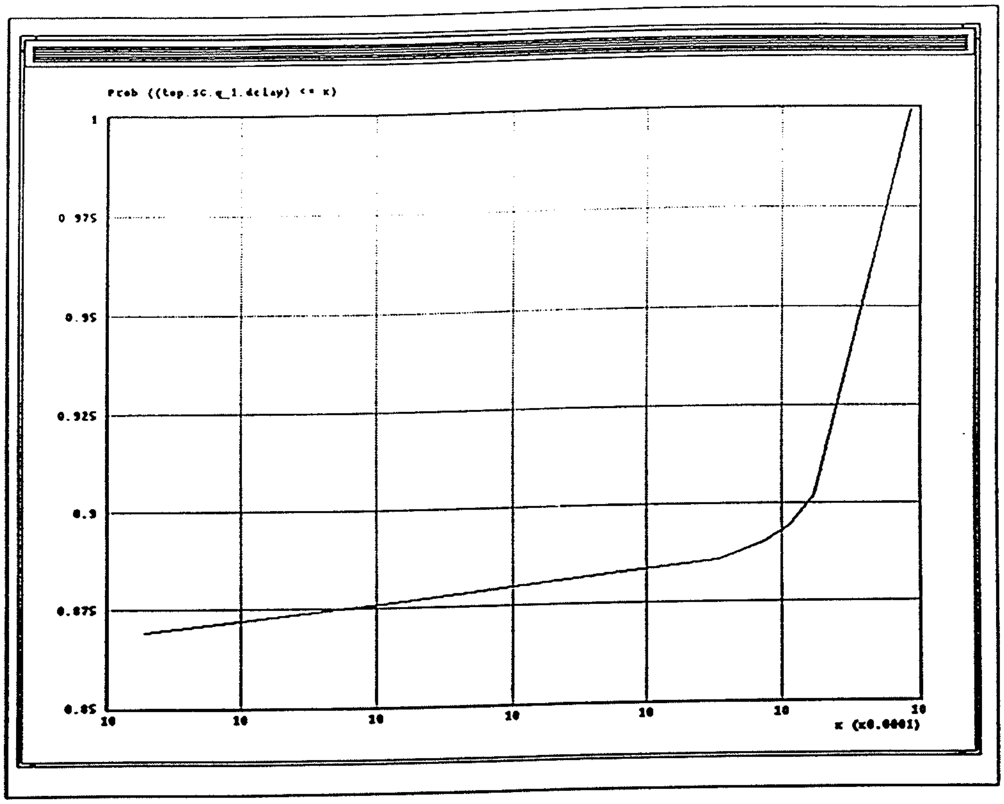
Graph 27: CDF of SG Throughput, G = 100 packets/sec



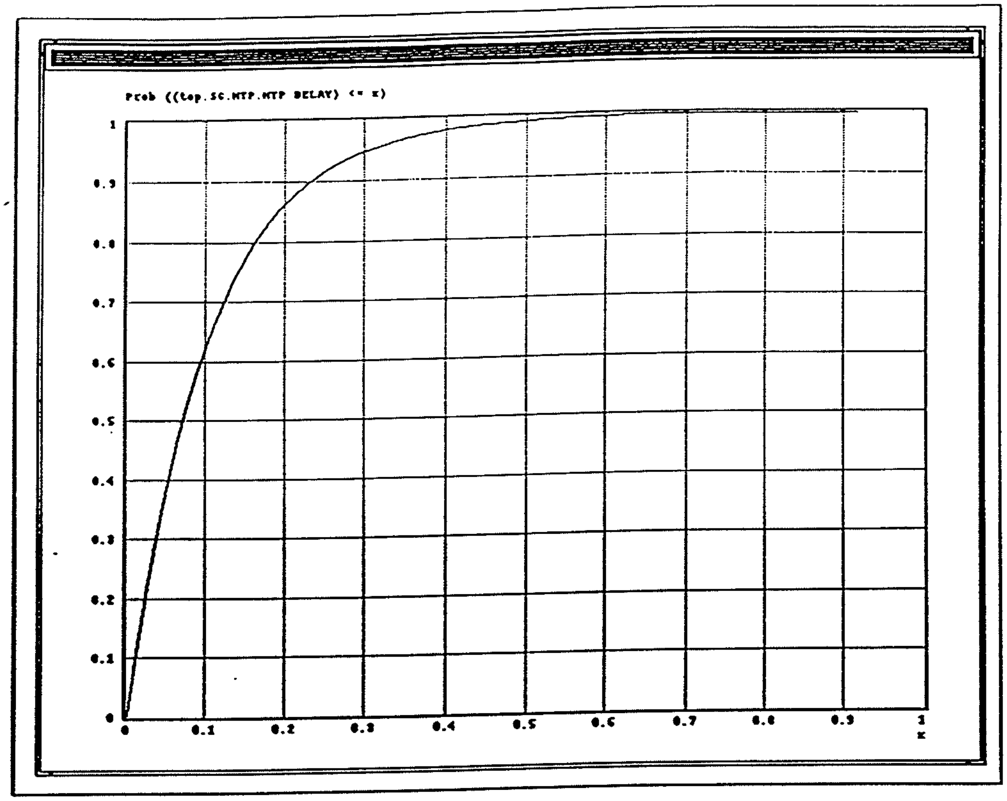
Graph 28: CDF of Total Delay at SG, G = 1000 packets/sec



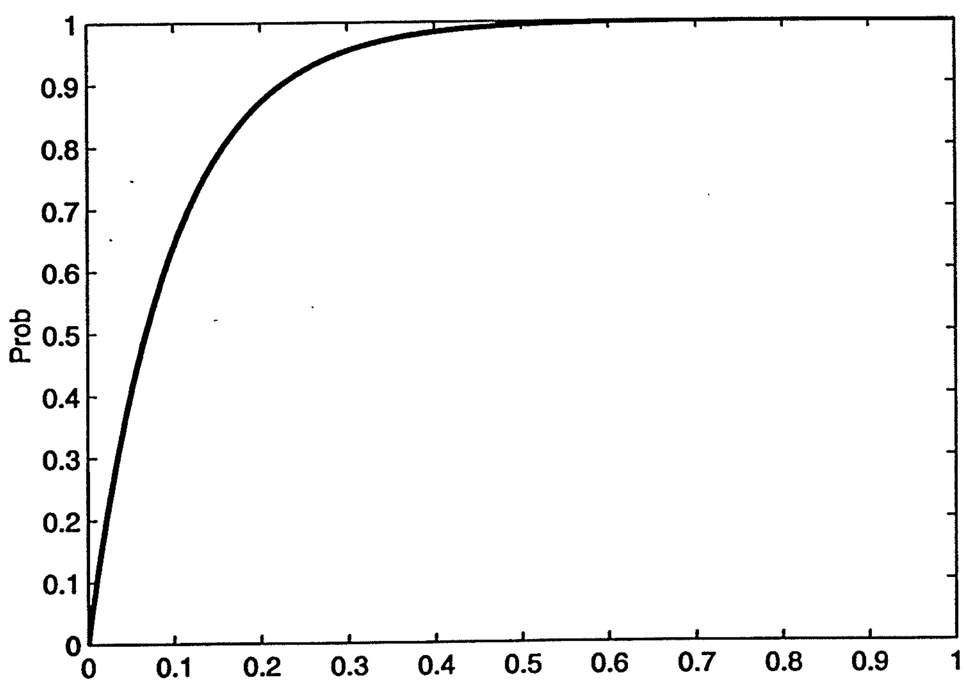
Graph 28a: Exponential Distribution of $\lambda = 1.099$



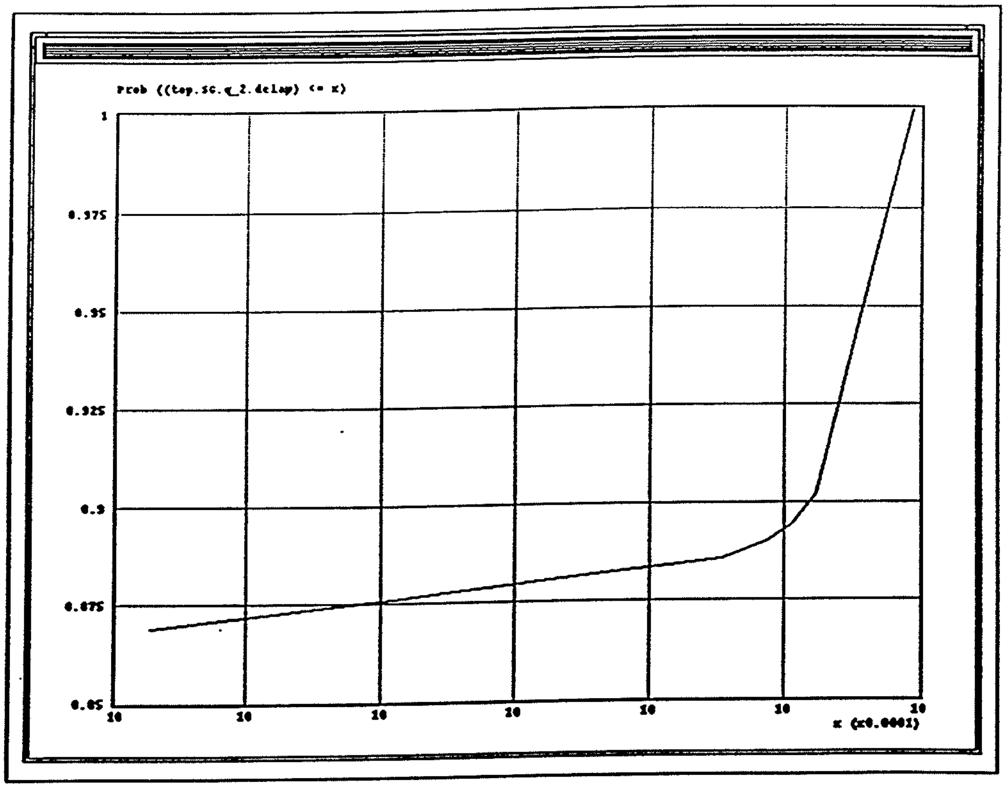
Graph 29: CDF of Delay at Buffer Q1, G = 1000 packets/sec



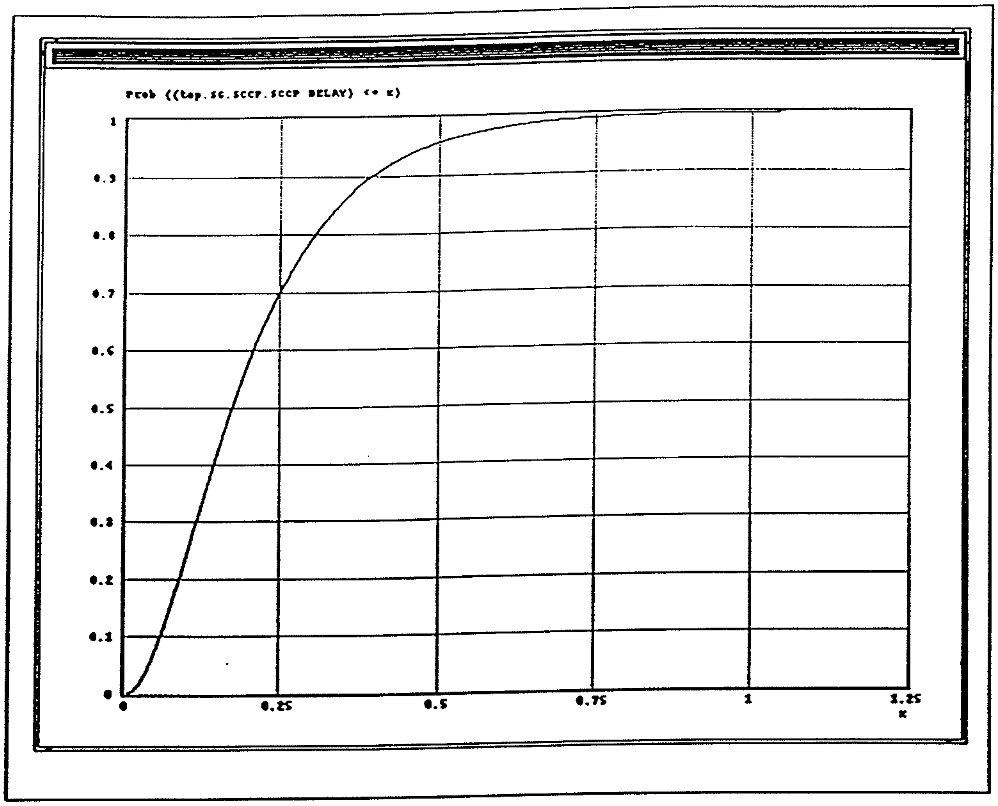
Graph 30: CDF of Delay at MTP Layers, G = 1000 packets/sec



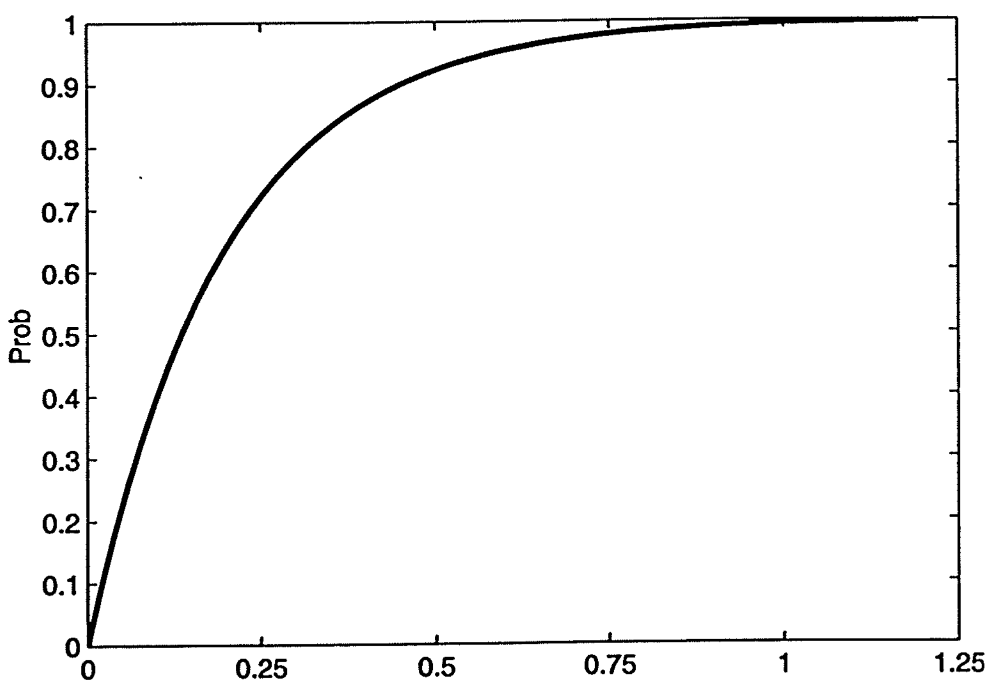
Graph 30a: Exponential Distribution of $\lambda = 0.0971$



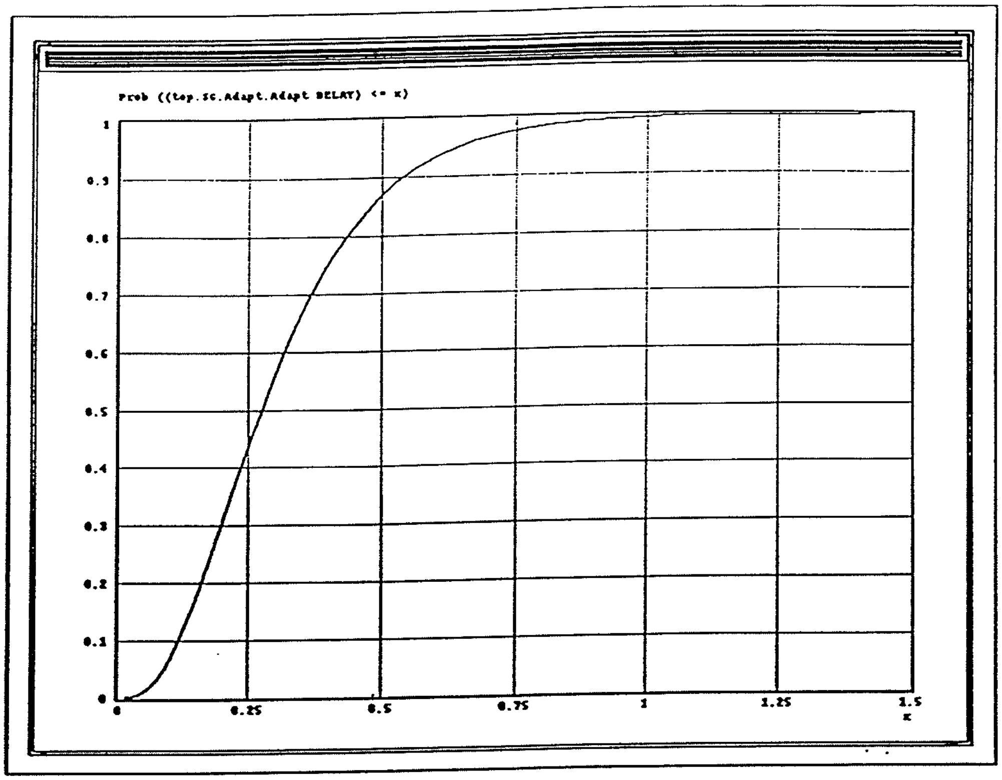
Graph 31: CDF of Delay at Buffer Q2, G = 1000 packets/sec



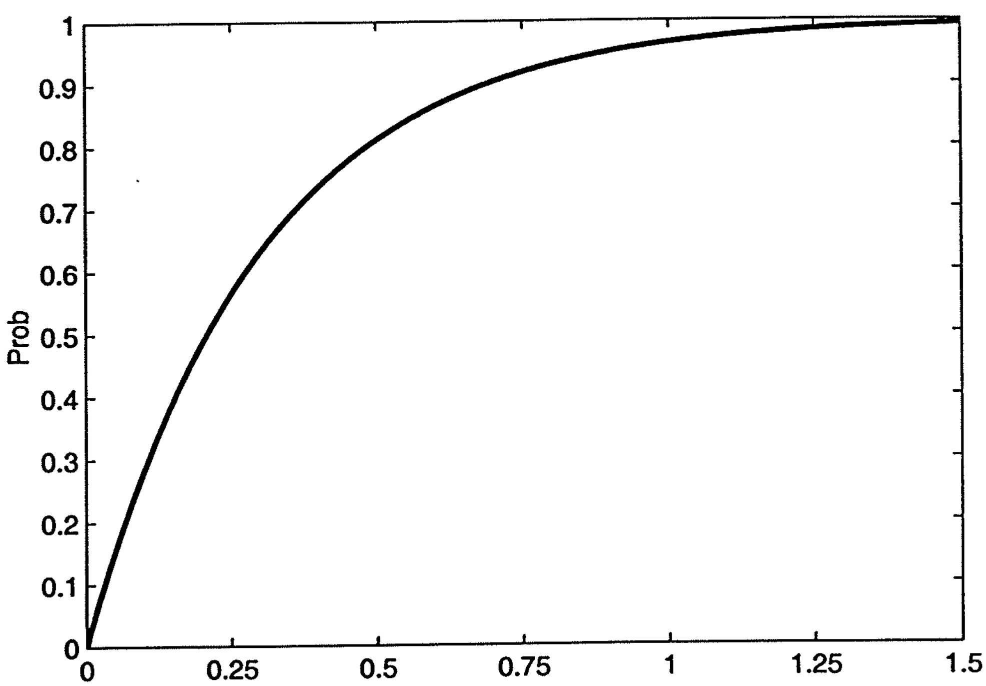
Graph 32: CDF of Delay at SCCP Layer, G = 1000 packets/sec



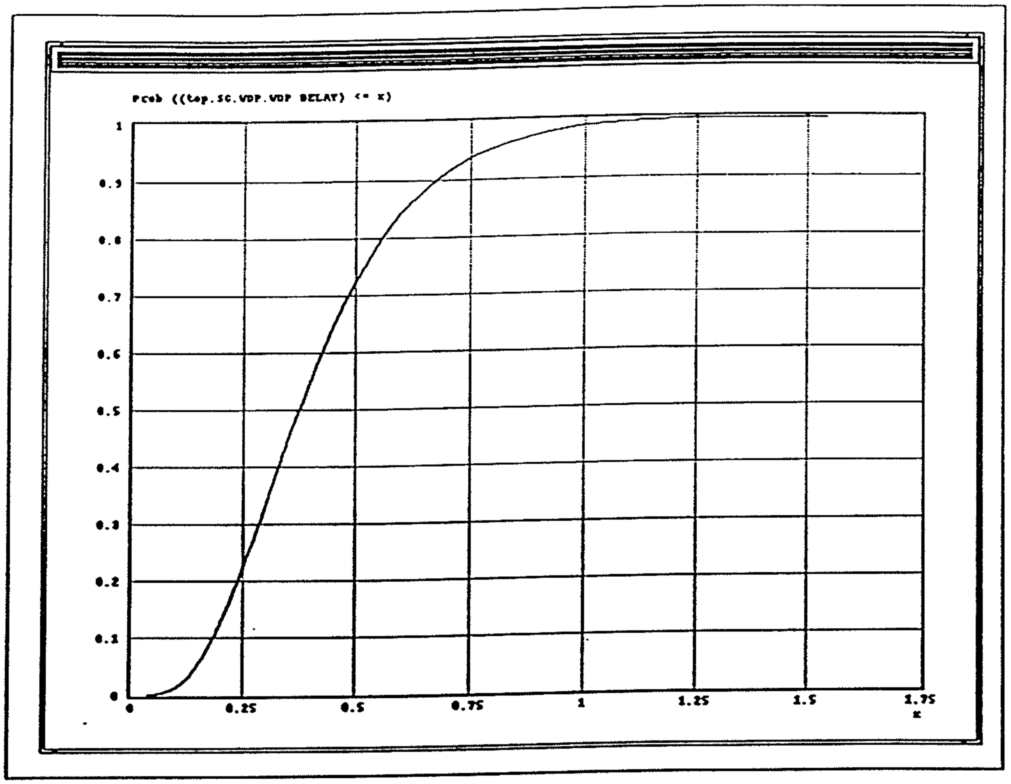
Graph 32a: Exponential Distribution of $\lambda = 0.1957$



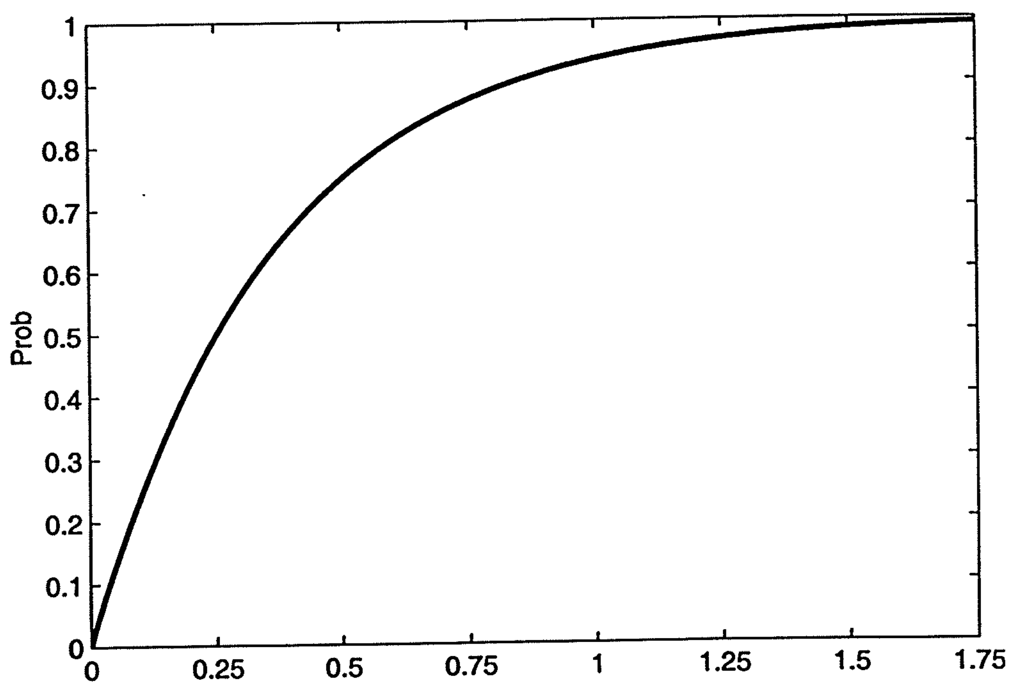
Graph 33: CDF of Delay at Adaptation Layer, G = 1000 packets/sec



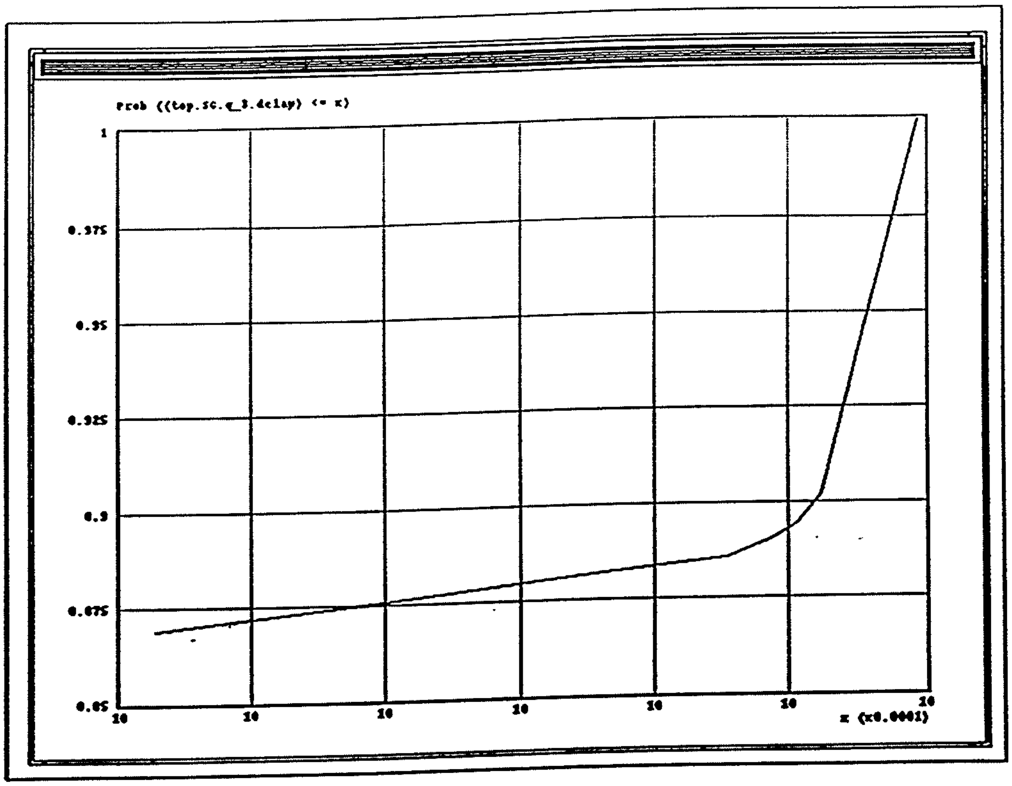
Graph 33a: Exponential Distribution of $\lambda = 0.3$



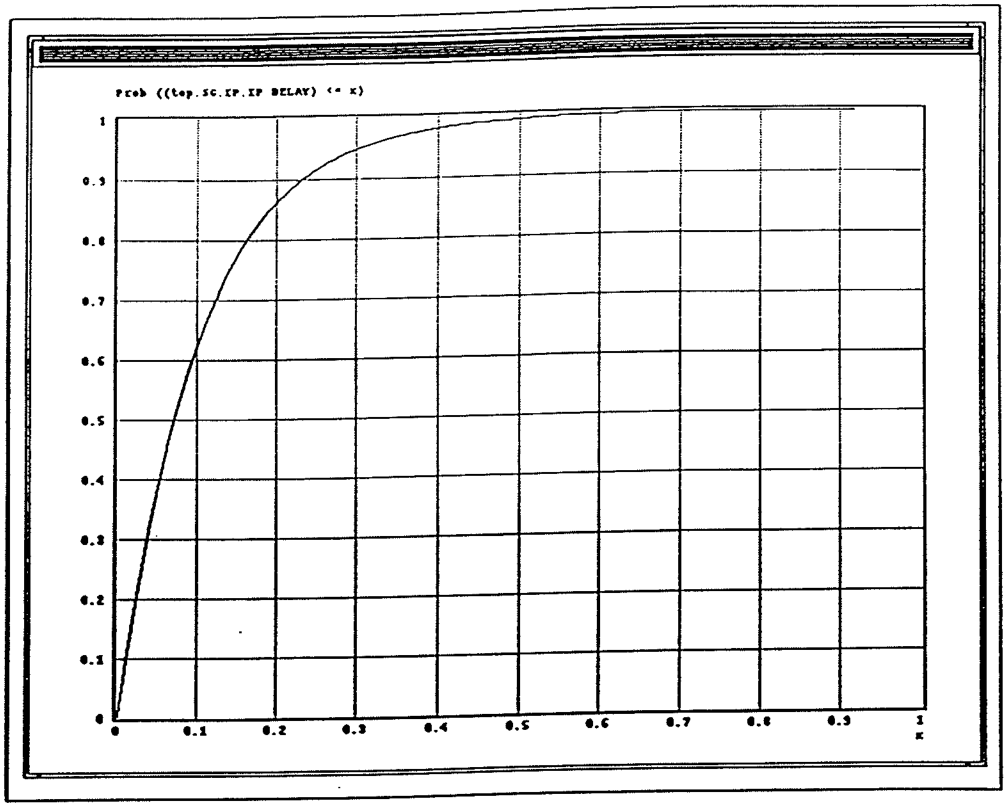
Graph 34: CDF of Delay at UDP Layer, G = 1000 packets/sec



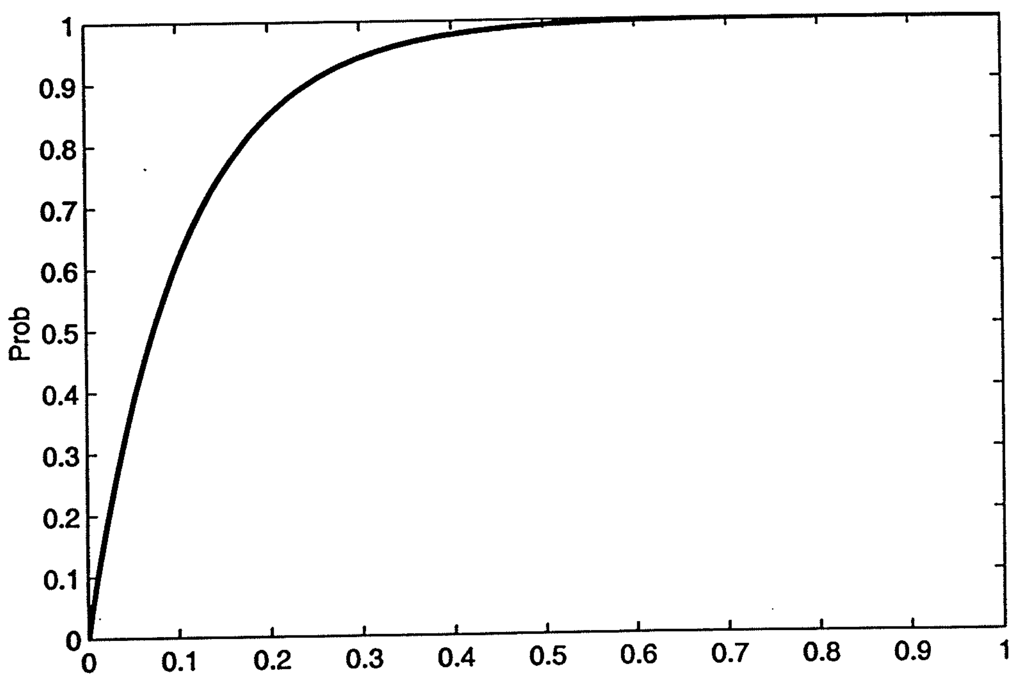
Graph 34a: Exponential Distribution of $\lambda = 0.362$



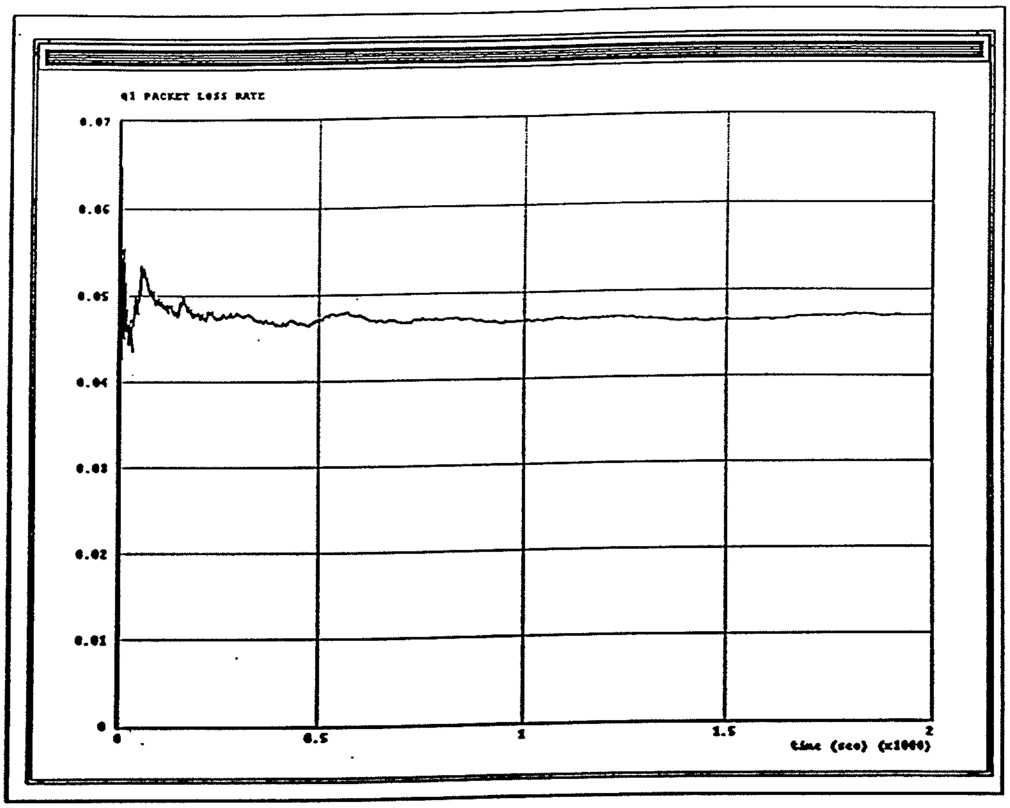
Graph 35: CDF of Delay at Buffer Q3, G = 1000 packets/sec



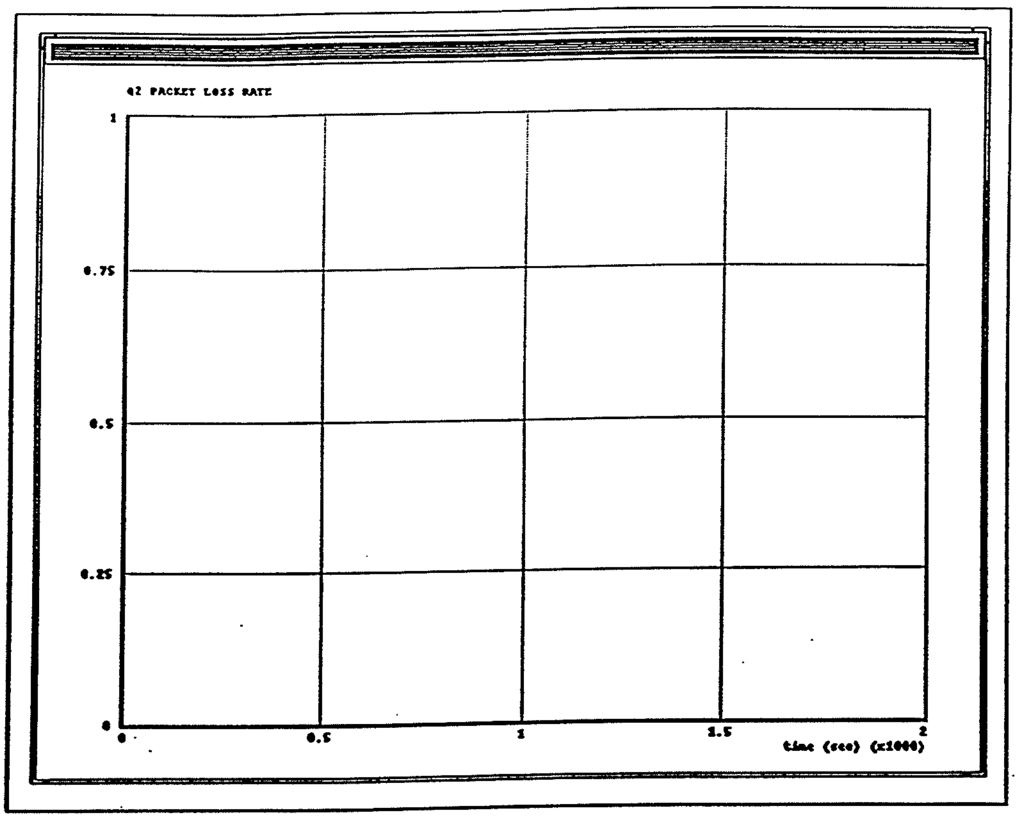
Graph 36: CDF of Delay at IP Layer, G = 1000 packets/sec



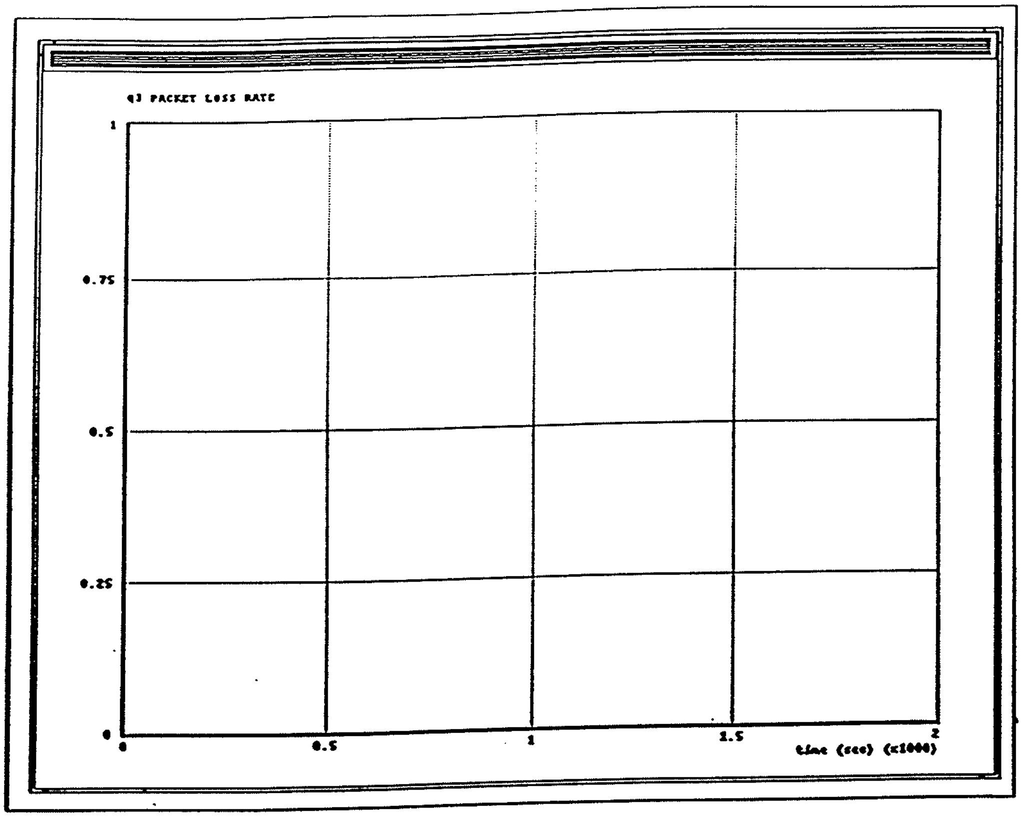
Graph 36a: Exponential Distribution of $\lambda = 0.1058$



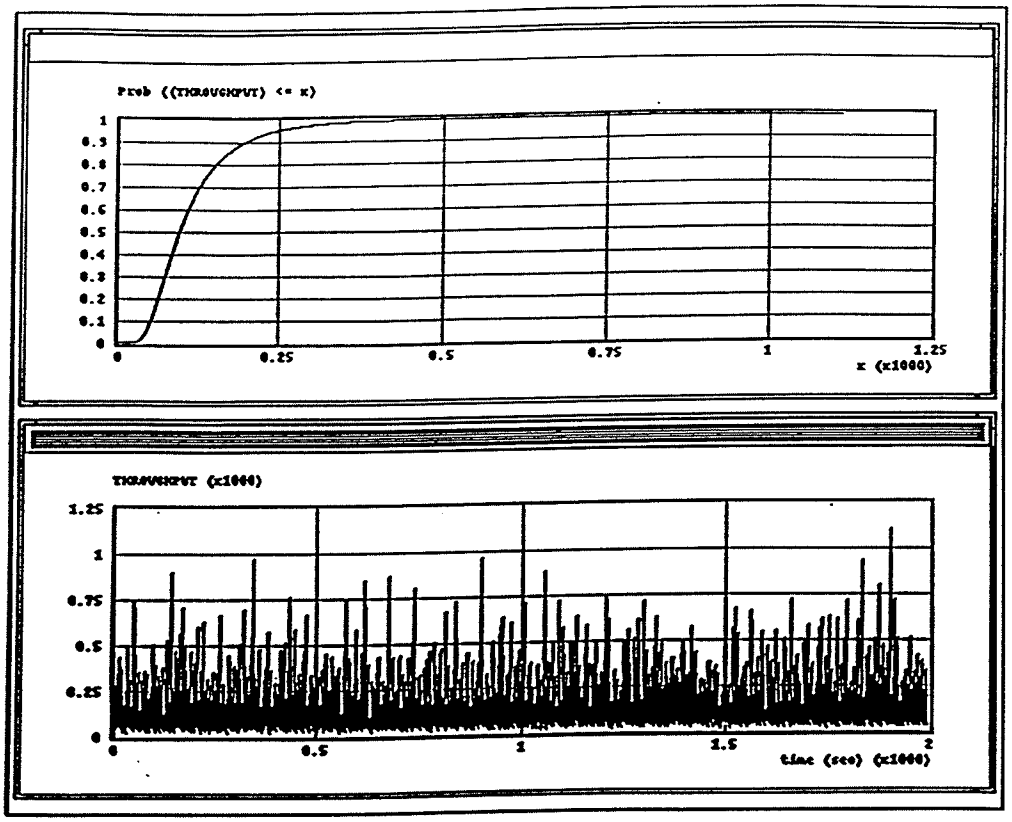
Graph 37: Packet Loss Rate at Buffer Q1, G = 1000 packets/sec



Graph 38: Packet Loss Rate at Buffer Q2, G = 1000 packets/sec (it remains zero throughout the simulation)



Graph 39: Packet Loss Rate at Buffer Q3, G = 1000 packets/sec (it remains zero throughout the simulation)



Graph 40: CDF of SG Throughput, G = 1000 packets/sec