Network Working Group Brian Bidulock
INTERNET-DRAFT Inet Technologies, Inc

Expires in six months April 2001

SS7 MTP3 Peer-to-Peer Adaption Layer
<draft-bidulock-sigtran-m3peer-00.txt>

Status of this Memo

Abstract

This Internet Draft defines a protocol supporting the transport of SS7 MTP User and Management messages over IP using the services of the Stream Control Transmission Protocol (SCTP). This protoocl would be used between SS7 Signalling Points employing the MTP Level 3 protocol. The SS7 Signalling Points may also employ standard SS7 MTP Level 3 route sets and routes over SS7 MTP Level 2 services.

1. Introduction

This document provides a specification for an SS7 MTP3 (Message Transfer Part - Level 3) Peer-to-Peer protocol and NIF (Network Interworking Function) for seamless and transparent interworking of the SS7 Network and IP network. The purpose for this protocol, in contrast to the M3UA protocol, is to provide the ability to given the SS7 Network the appearance of remote Signalling Point Codes within the IP network, while allowing a flexible ASP architecture within the IP network which does not necessarily conform to the SS7 view.

1.1. Scope

There is a need for Switched Circuit Network (SCN) signalling protocol delivery over an IP network. This includes delivery from a Signalling Gateway (SG) to a Media Gateway Controller (MGC) or IP Signalling Point, as described in the Framework Architecture for Signalling Transport [1]. This could allow for convergence of some signalling and data networks. SCN Signalling nodes would have access to databases and other devices in the IP network domain that do not employ SS7 signalling links or Message Transfer Parts. There may also be operational and cost advantages when traditional signalling route sets and routes are replaced by IP network "connections".

The delivery mechanism described in this document allows for full MTP3 message handling and network management capabilities interworking between the SS7 network and the IP network. An SS7 node equipped with an IP network connection is called an IP Signalling Point (IPSP). The ISPSs function as traditional SS7 nodes using the IP network instead of SS7 links.

The delivery mechanism should

- Support seamless interworking between MTP3 nodes within the SS7 network and IP hosts within the IP network.
- Support the MTP Level 3 / MTP User interface boundary.
- Support management of SCTP transport associations and traffic instead of MTP3 route sets and routes.
- Support asynchronous reporting of status changes to management.

1.2. Terminology

MTP The Message Transfer Part of the SS7 protocol [2].

MTP2
MTP Level 2, the MTP Signalling Link Level.

MTP3
MTP Level 3, the MTP Signalling Network Level.

MTP3-User
A protocol that normally uses the services of MTP Level 3.

Signalling End Point (SEP)
A node in an SS7 network that originates or terminates signalling messages. One example is a central office switch.

IP Signalling Point (IPSP)
An SS7 Signalling Point with an IP network connection used for SS7 over IP.

Signalling Gateway (SG)
A signalling agent that receives/sends SCN native signalling at the edge of the IP network [4]. In this context, an SG is an SS7 Signalling Point that has both an IP network connection used for SS7 over IP, and a traditional (non-IP) link to an SS7 network.

1.3. Abbreviations

1.4. Assumptions

The following simplifying assumptions are made:

(1) We assume that processor delays are short an much smaller that queueing or transport delays. This assumption derives from the understanding that processor power can always be increased with respect to load to satisfy this condition.

(2) We assume transmission delays and bit error rates for multi-mode fibre for long haul connections.

(3) For protocols which amalgamate multiple packets into one datagram, we assume that such amalgamation is not performed.

(4) For protocols which must segment messages over a certain size, we assume that messages are not segmented.

(5) Poisson distributed inter-arrival times between messages is assumed. This assumption is the basis for SS7 performance analysis and will also be used for SCTP and SSCOP MCE analysis for common comparison.

1.5. Architecture

The architectures considered are illstrated in the figures following. Each figure is referred to by the analysis sections and are labelled for reference.

```
+-------------------------------------------------------------+
|                                                             |
|  +----------+                                +----------+   |
|  |          |                                |          |   |
|  |     +----|                                |----+     |   |
|  |     | A  |--------------------------------|  C |     |   |
|  |     +----|                                |----+     |   |
|  |          |                                |          |   |
|  |  Node 1  |                                |  Node 2  |   |
|  |          |                                |          |   |
|  +----------+                                +----------+   |
|                                                             |
+-------------------------------------------------------------+
|                                                             |
|        Figure 1.1 - Simplex Point-to-Point Connection       |
|                                                             |
+-------------------------------------------------------------+

+-------------------------------------------------------------+
|                                                             |
|  +----------+                                +----------+   |
|  |          |                                |          |   |
|  |     +----|                                |----+     |   |
|  |     | A  |--------------------------------|  C |     |   |
|  |     +----|                                |----+     |   |
|  |          |                                |          |   |
|  |     +----|                                |----+     |   |
|  |     | B  |--------------------------------|  D |     |   |
|  |     +----|                                |----+     |   |
|  |          |                                |          |   |
|  |  Node 1  |                                |  Node 2  |   |
|  |          |                                |          |   |
|  +----------+                                +----------+   |
|                                                             |
+-------------------------------------------------------------+
|                                                             |
|        Figure 1.2 - Duplex Point-to-Point Connection        |
|                                                             |
+-------------------------------------------------------------+
+--------------------------------------------------------------+
|                                                              |
|              IEEE 802.3           IEEE 802.3                 |
|  +----------+ ==========           ========== +----------+   |
|  |          |  |   |               |   |  |          |  |
|  |    +----|  |  +---+           +---+  |  |----+    |  |
|  |    | A  |---+  | R |------------| R |  +---|  C |    |  |
|  |    +----|     +---+           +---+      |----+    |  |
|  |          |                           |          |  |
|  |  Node 1  |                           |  Node 2  |  |
|  |          |                           |          |  |
|  +----------+                           +----------+   |
|                                                              |
+--------------------------------------------------------------+
|                                                              |
|          Figure 1.3 - Simplex LAN/WAN Connection             |
|                                                              |
+--------------------------------------------------------------+
```

```
+--------------------------------------------------------------+
|                                                              |
|              IEEE 802.3           IEEE 802.3                 |
|  +----------+ ==========           ========== +----------+   |
|  |          |  |   |               |   |  |          |  |
|  |    +----|  |  +---+           +---+  |  |----+    |  |
|  |    | A  |---+  | R |------------| R |  +---|  C |    |  |
|  |    +----|     +---+           +---+      |----+    |  |
|  |          |                           |          |  |
|  |    +----|     +---+           +---+      |----+    |  |
|  |    | B  |---+  | R |------------| R |  +---|  D |    |  |
|  |    +----|  |  +---+           +---+  |  |----+    |  |
|  |          |  |   |               |   |  |          |  |
|  |  Node 1  | ==========           ========== |  Node 2  |  |
|  |          | IEEE 802.3           IEEE 802.3 |          |  |
|  +----------+                           +----------+   |
|                                                              |
+--------------------------------------------------------------+
|                                                              |
|          Figure 1.4 - Duplex LAN/WAN Connection              |
|                                                              |
+--------------------------------------------------------------+
```

Simplex Point-to-Point Connection

Figure 1.1 illustrates a simplex point-to-point connection. In this configuration it is assumed that the IP stack on each node queues packets to a driver which places them directly on the media connecting the two nodes. This media is considered to be a simple connection with no queueing, just transport delay. This simulates a point-to-point link connection directly attached to each node.

Duplex Point-to-Point Connection

Figure 1.2 illustrates a duplex point-to-point connection. This architecture could be extended to the point where the number of connections is N>=1. This is the situation where each node has ’N’ interfaces with ’N’ point-to-point connections.

Simplex LAN/WAN Connection

Figure 1.3 illustrates a simplex LAN/WAN configuration. Each interface on each node is connected to its own LAN interface. The example LAN used is IEEE 802.3 (Ethernet). Each router (R) is connected both to a LAN and to a Point-to-Point connection. Point-to-Point connections are extended to connect the routers together.

Several additional simplifying assumptions are made for this (and the following) configuration:

(1) It is assumed that ARP table entries are manually configured for each interface and router, and that no broadcasts or ARP queuries are required to establish the MAC address associated with the destination IP address of each interface in the system.

(2) Routing tables for each interface and router is assumed to be statically configured. No routing protocols are broadcast.

(3) Router to router connections are Point-to-Point links which are extended over a distance.

(4) Each IEEE 802.3 (Ethernet) LAN is assumed to be a 10baseT connection to an Ethernet HUB. All connection distances between the interface cards, hubs, and routers adjacent to each node are considered to be short distances.

(5) It is assumed that no other traffic (other than the traffic under consideration) is present on any LAN or any router. Thus, the ARP and RIP assumptions.

(6) Routers are assumed to be OSPF (Open Shortest Path First) FQ (Fair Queueing) routers with per-destination queues.

Duplex LAN/WAN Connection

Figure 1.4 illustrates a duplex LAN/WAN configuration. Each interface on each node is connected to its own LAN interface to an IEEE 802.3 Ethernet LAN over 10baseT running at 10Mbps. Each router is connected on the one side to a LAN and on the other side to a point-to-point link to one of the other routers.

All of the assumptions for the simplex LAN/WAN case are considered assumptions in this configuration as well.

2. Consideration and Background

2.1. Considerations

2.1.1. SS7

SS7 considerations (and the complete analysis) is provided in [4]. Similar considerations provided for links in SS7 will be assumed for PPP links in the configurations analyzed here for the sake of comparison.

2.1.2. PPP Link

Similar considerations as are applied for SS7 links in [4] are considered here for PPP links. Regardless of the validity of assumptions made for SS7 links in [4], the same assumptions will be made here for the sake of comparison.

2.1.3. Ethernet 802.3

Ethernet analysis is based on [5]. Assumptions and proofs will not be presented here, please refer to [5] for more details. The results of the Ethernet analysis presented in the cited works will be incorporated into this analysis without modification.

2.1.4. Router

Routers are assumed to be OSPF (Open Shortest Path First) FQ (Fair Queueing) routers. The analysis of the queueing characteristics of routers of this type are presented in [6] and will be recreated here. The results of the router analysis presented in the cited works will be incorporated into this analysis without modification. Nevetheless, the following are some of the simplifying assumptions normally taken for routers:

(1) It is assumed that outgoing interfaces are chosed on metrics which indicate the hop distance from the router to the destination. An available route with the lowest metric is chosen. If there are two or more routes which lead to the same destination and which have the same metric, one of the routes will be determinately chosen.

(2) Although it is available, it is assumed that no load sharing exists in the selection of routes with the same metric.

(3) The return address on the datagram is not taken into consideration when routing messages.

2.1.5. IP Stack

Considerations for IP stack queueing and work times are based on the 4.4BSD implementation. Although other implementations exist, this model is used to yeild representative results. Several simplifying assumptions are made concerning the 4.4BSD IP stack as follows:

(1) Processor work times are considered to be short with respect to emission and propagation times. To show that this assumption is workable, consider that a significant amount of processor power can be applied for low levels of traffic to allow the assumption to hold. This is not necessarily an cost-effective situation, but one which is easier to study.

(2) Delays from the indication of the ability to output information onto the interface to the actual output of data is assumed to be negligible. This assumes that processors have asynchronous interrupts and that interrupt service routines are sufficiently quick to permit the output of back-to-back frames on the interface without significantly loading the main processor. This is a reasonable assumption for modern computer architectures and operating system software.

(3) A buffer exists behind the input to the IP level. This buffer holds messages which are waiting for transmission to the interface. This buffer is limited in size.

(4) It is assumed that there are no network configuration or routing table changes which would require additional delays in the background.

2.1.6. SCTP

For M2UA SCTP behaviour, we are concerned with SS7 performance over SCTP. For this analysis it is sufficient that a stream within an SCTP association be analyized in comparison to a SS7 link (at the same loaing points) in the simplest case. For M3UA, additional analysis will provide the effects of mixing SS7 SLS (link) traffic into multiple streams within a single SCTP association. It will also be considered what the effects of "changeover" between streams within the SCTP association will be, and what the effects of changeover between streams in different associations will be.

2.1.7. SSCOP-MCE

2.2. Background Information

Some background information is in order before we begin the analysis of SCTP and SSCOP-MCE. This background information and baseline analysis is necessary to understand the effects of various network configurations.

2.2.1. Link

Links can be graded into a number of error rates as illustrated in Table 2.1.

Table 2.1 - Bit Error Rates

Type of Transport	Bit Error Rate	Speed	EM/8
Satellite Links	1 x 10ˆ-3		
PCM Channel	1 x 10ˆ-6	64k bps	125us
PCM Link	1 x 10ˆ-6	1.544M bps	
Ethernet LAN	1 x 10ˆ-9	10M bps	800ns
SONET span	1 x 10ˆ-10	675M bps	

Links are modeled as a queue where the service time is dependent upon the link’s bit rate and the length of the packet for transmission. This length of time is called the "emission time" for the packet. The emission time is simply the the packet length in bits divided by the transmission speed in bits per second.

2.2.2. Router

2.2.3. Ethernet

Ethernet links are specialized forms of links. They do not simply have the emission times associated with communications links: there is the possibility of CDMA collisions which adds further delay (and HOL blocking) to some packets.

Other LAN protocols (e.g., 802.4 Token Bus, or 802.5 Token Ring) or switched LAN technologies may also be considered.

2.2.4. IP Stack

IP stacks based on the BSD socket implementations (or even SVR4 streams) use a software interrupt procedure for moving message buffers from the upper layers down to the device driver and finally on to the media. This mechanism is not normally asynchronous or event driven, but works on the basis of setting a software interrupt which is examined on a quantum basis in operating systems which include pre-emtive scheduling. Non-pre-emptive scheduled operating systems may use a task window (for time-triggered real-time systems) or execution of a system event loop (for run-to-completion systems) to emulate an interrupt in the operating system.

These approaches to moving the messsage buffers down the IP protocol stack include some queueing associated with the software interrupt. These queueing effects should be considered in the analysis of all protocols using the IP stack.

The IP stack must also be considered to introduce queueing delays when broadcast based protocols are used for address resolution (e.g., ARP) and routing (e.g., RIP). In this document it is assumed that these effects can be avoided (e.g., through manual ARP entries and ICMP redirects).

Other considerations of the IP stack include the fact that the IP stack performs routing.

3. SCTP Analysis

This section proceeds with the SCTP analysis. To be able to simplify the task of the analysis, initially, the following assumptions are made:

- Only one association is considered.
- Only one stream is considered in the association.

Definitions

F - Frequency (bps) of the Transmission Medium
Pe - Probability of Error
Pl - Probability of Message Loss

3.1. Probability of Message Loss

Message loss has several contributing factors. Unfortunately, SCTP has no mechanism for distinguish between message loss due to any specific factor. There is no indication given by the lower layer to SCTP or by the SCTP peer as to what the cause of the loss might be. SCTP procedures do no distinguish between loss types.

Factors contributing to message loss are as follows:

- Loss due to bit errors.
- Loss within the IP stack.
- Loss within network routers.
- Loss due to excessive delay.

SCTP has two mechanisms which compete when a message is lost. Messages which are detected as lost through the receipt of four successive SACK reports, are retransmitted. Messages which have not been acknowledged within the message's time-to-live as determined by the ULP are considered lost. Messages cannot be retransmitted more than 'Path.retrans.max' to any given destination. Messages cannot be retransmitted more than 'retrans.max' on an association.

3.1.1. Loss due to Bit Errors

Bit errors have several effects on SCTP. SCTP has a 32-bit Adler Checksum which is used to protect SCTP datagrams against bit errors. The SCTP header is not separately protected, the entire datagram is protected by one checksum. The result of detection of a bit error with the Adler-32 Checksum is the discard of the datagram.

The actual number of packets discarded from the detection of bit errors is dependent on a number of factors:

- The probability of a bit error (Pe).
- The size of the packets transmitted (l).
- The packet load (L).
- The probability of detecting a bit error (Pd).

The response of SCTP to received datagrams with checksum errors (detected bit errors) will be to silently discard the datagram at the receiver. One of three things will occur in response to this silent discard:

(1) Another datagram will arrive at the receiver which will cause a SACK Gap Report to be issued for the lost datagram.

(2) The sender's RTO (receive time out) will expire and the chunks in the discarded datagram will be retransmitted.

(3) The sender's lifetime for the packet will expire and the association will be lost.

The silent discard behaviour of datagrams with detected bit errors means that bit error rates may be translated into datagram loss rates. The datagram loss rate contributed by the bit error rate can be added to the other datagram loss rates to arrive at a final loss rate. Analysis can then be calculated on the basis of this combined loss rate.

3.1.2. Loss within the IP stack.

Most losses associated with IP stack implementations include buffer overruns and ARP queries. These two factors are considered below:

Buffer Overruns

Buffer overruns result when the protocol layer above IP generates IP datagrams faster than the IP layer can post these datagrams to the device driver; or, the device driver receives datagrams faster that it can place these datagrams on the media. If these differences are significant and last for a sufficient period of time, input buffer chains to either the IP level or the device driver will result in the loss of a packet. Packets lost to the IP level due to lack of input buffer space, SCTP will be notified by the IP layer; however, if the overrun buffer is between the IP layer and the device driver, datagram loss will not be reported.

ARP Losses

Some IP stack implementations discard the packet which is queued for transmission which causes an ARP cache miss. This would result in packet loss at some ARP cache miss rate. ARP cache misses are due either to cache entry expiry, or initial cache entry establishment. Because this analysis assumes that ARP cache misses can be removed by manually or statically allocating IP to MAC address mappings, these losses are not considered here.

3.2. Queueing Effects of Retransmission

3.2.1. Retransmission with Multiple Paths

3.2.2. The Cost of Retransmission

3.3. Effects of the Path

3.3.1. The Effects of RTT on Delay

Backward Error Correction

Forward Error Correction

3.3.2. The Effects of Throughput Limitation on Delay

Path Queueing

Path queueing is an effect where the path has a limited throughput capability and attempts at feeding higher packet rates at the path will result in the internal queueing of packets along the path. This is a congestive condition which results in packets being delayed longer than they would under lesser load conditions.

Any given path can be modelled as a series of queues. The delay along a link is normally a determinate and scalar delay. The delay through a router or interface is a queueing delay.

A characteristic of the path which can be used to determine whether delays are due to congestion or whether delays are due to link delays is the RTT. When paths a not congested, they experience low RTT values. When paths are congested they experience higher RTTs. A mechanism which can be used to determine whether paths or congested or not is by comparing the minimum theoretical or experienced RTT to the current measured RTT.

A hypothetical situation is illustrated in Figure 1 below. Although the actual curve of round trip times to congestion levels is not known, there are several things which can be said:

- Because paths are modeled as a directed graph of queues connected by determinate scalar delay links, it can be said that the RTT to Congestion function is monotonically increasing.
- Because paths have queues with a finite number of servers, it can be said that the RTT to congestion function is asymptotically increasing at the maximum throughput of the path.
- When paths are experiencing no congestion, the RTT will be at a minimum. The RTT minimum is the y-axis intercept of the RTT/Congestion curve.
- When the current measured RTT is higher than the minimum RTT, then the path is experiencing some congestion.
- When the current measured RTT is higher than twice the acceptable transfer time, the path is experiencing excessive congestion.

Congestion on a path may be caused either by the traffic which is being applied to the path with RTT measured, or other traffic along a portion of the same path which is either being applied from the same endpoint or applied from other endpoints. Congestion may also occur from failure of component associated with queues along the path.

Path Loss

3.3.3. The Effects of Noise on Delay

Path Loss

3.4. Congestion

3.4.1. Traffic Offerred vs. Traffic Transferred

Regardless of the precise nature of an IP network configuration between two points, some things may be said about congestion. There are a limited number of types of systems which are used to make up an IP network, and a limited range of characteristics which these systems exhibit under load.

Figure 2 illustrates load curves for three systems as follows:

```
+-------------------------------------------------------------+
|                                                           |
|   |                         |       |            |
|   |                                 |            |
|   |                         |      / |            |
|   |                                |            |
|   | Twice              Acceptable |      | | Maximum      |
|   | Acceptable          Congestion       /    Throughput    |
|   | Transfer Timer                 |     |  |             |
|  R |    |                           /                |
|  T |    |                      |   _/    |            |
|  T |    v                        __/                  |
| - -|- - - - 2 x ATT - - - - - - _____/- - - - -|- - - - - -    |
|   |              _________/                        |
|   | _________________/             |        |             |
|   |/                                           |
| - -+- - - - RTT Mininmum  - - - - - -|- - - - -|- - - - - -    |
|   |                                          |
|   |                         |       |            |
|   |                                          |
|   |                         |       |            |
| ---+-------------------------------------------------------   |
|   |                  Congestion  |        |             |
|   |                                          |
|                                              |
+-------------------------------------------------------------+
|                                              |
|                Figure 1 - RTT vs. Congestion               |
|                                              |
+-------------------------------------------------------------+
```

(1) The first system corresponds to a wire. It has a fixed linear relationship between the traffic offered and the traffic transfered: all of the traffic offered is transferred.

(2) The second system corresponds to a

(3) The third system corresponds to a CDMA medium. The characteristic of CDMA is that collisions increase with traffic load and impede the ability of the system to transfer data. Therefore, this is a good system to model CMDA access media such as ethernet.

3.4.2. Traffic Offerred vs. Delay

3.5. A Model for Delay

3.6. Transmission and Retransmission Procedures

3.7. Congestion/Flow Control Procedures

4. Procotol Elements

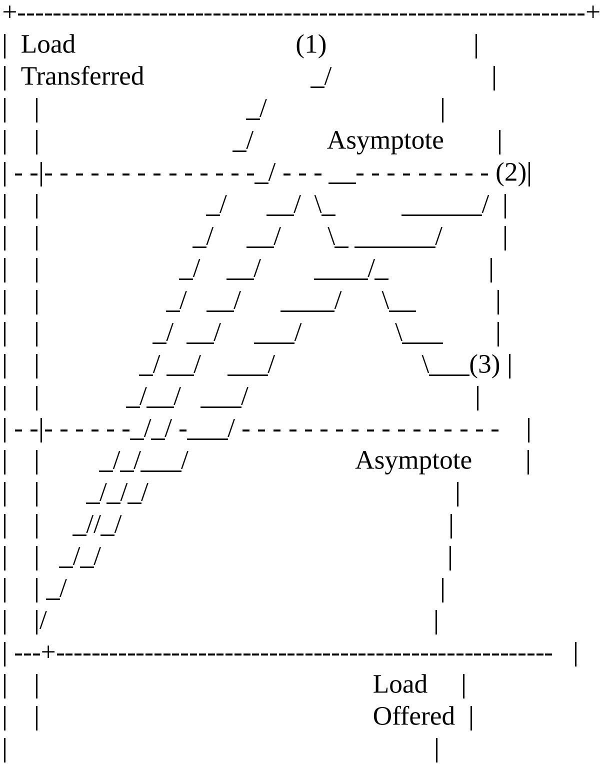

Figure 2 - Load Curves for Various Systems

5. Procedures

5.1. Transmission and Retransmission Procedures

5.2. Congestion/Flow Control Procedures

6. Examples

Security Considerations

This congestion procedures relies on the mechanisms of Simple Control Transport Protocol (SCTP) to provide security.

7. Acknowledgements

8. References

[1] Simple Control Transmission Protocol, draft-ietf-sigtran-sctp-07.txt, February 2000

[2] Distributed Systems, 2nd Edition, Sape Mullender Ed., Addison-Wesley, 1993, ISBN 0-201-62427-3

[3]

[4]

[5]

[6]

[7]

[8]

[9]

[10]

[11]

[12]

[13]

[14]

[15]

[16]

[17]

[18]

[19]

9. Author’s Addresses

Brian F. G. Bidulock Tel: +1-972-578-3959
Inet Technologies, Inc. EMail: brian.bidulock@inet.com
1255 W 15th Street
Plano, TX 7507
USA

This Internet Draft expires October 2000.