                    The Cisco SRP MAC Layer Protocol

Status of this Memo

Copyright Notice

Abstract

   This document specifies the MAC layer protocol, "Spatial Reuse
   Protocol" (SRP) for use with ring based media. This is a second
   version of the protocol (V2).

   The primary requirements for SRP are as follows:

   - Efficient use of bandwidth using:
           spatial reuse of bandwidth
           local reuse of bandwidth
           minimal protocol overhead
   - Support for priority traffic
   - Scalability across a large number of nodes or stations attached to
      a ring
   - "Plug and play" design without a software based station management
      transfer (SMT) protocol or ring master negotiation as seen in
      other ring based MAC protocols [1][2]
   - Fairness among nodes using the ring
   - Support for ring based redundancy (error detection, ring wrap,
      etc.) similar to that found in SONET BLSR specifications.
   - Independence of physical layer (layer 1) media type.

   This document defines the terminology used with SRP, packet formats,
   the protocol format, protocol operation and associated protocol
   finite state machines.

Table of Contents

1. Differences between SRP V1 and V2

    This document pertains to SRP V2. SRP V1 was a previously published
    draft specification. The following lists V2 feature differences from
    V1:

    - Reduction of the header format from 4 bytes to 2 bytes.

    - Replacement of the keepalive packet with a new control packet that
       carries usage information in addition to providing a keepalive
       function.

    - Change bit value of inner ring to be 1 and outer to be 0.

    - Reduction in the number of TTL bits from 11 to 8.

    - Removal of the DS bit.

   - Change ordering of CRC transmission to be most significant octet
      first (was least significant octet in V1). The SRP CRC is now the
      same as in [5].

   - Addition of the SRP cell mode to carry ATM cells over SRP.

   - Changes to the SRP-fa to increase the usage field width and to
      remove the necessity of adding a fixed constant when propagating
      usage messages.

2. Terms and Taxonomy

2.1. Ring Terminology

   SRP uses a bidirectional ring. This can be seen as two symmetric
   counter-rotating rings. Most of the protocol finite state machines
   (FSMs) are duplicated for the two rings.

   The bidirectional ring allows for ring-wrapping in case of media or
   station failure, as in FDDI [1] or SONET/SDH [3]. The wrapping is
   controlled by the Intelligent Protection Switching (IPS) protocol.

   To distinguish between the two rings, one is referred to as the
   "inner" ring, the other the "outer" ring. The SRP protocol operates
   by sending data traffic in one direction (known as "downstream") and
   it's corresponding control information in the opposite direction
   (known as "upstream") on the opposite ring. Figure 1 highlights this
   graphically.

FIGURE 1. Ring Terminology

```
                                {outer_data
                          -----  inner_ctl}
          ---------------->| N |----------------
         | --------------| 1 |<-------------   |
         | | {inner_data  -----             | |
         | |  outer_ctl}                     | |
       -----                               -----
      | N |                               | N |
      | 6 |                               | 2 |
       -----                               -----
        ^ |                                 ^ |
      o | |                               i | |
      u | |                               n | |
      t | |                               n | |
      e | |                               e | |
      r | |                               r | |
        | v                                 | v
       -----                               -----
      | N |                               | N |
      | 5 |                               | 3 |
       -----                               -----
        | |                                 | |
        | |               -----             | |
        | --------------->| N |---------------  |
         ----------------| 4 |<----------------
                          -----
```
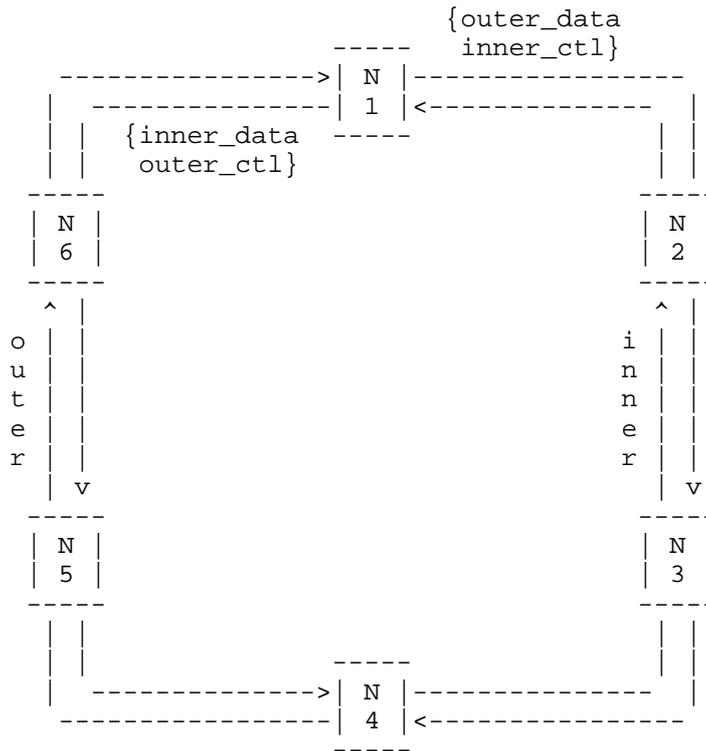
2.2. Spatial Reuse

   Spatial Reuse is a concept used in rings to increase the overall
   aggregate bandwidth of the ring. This is possible because unicast
   traffic is only passed along ring spans between source and
   destination nodes rather than the whole ring as in earlier ring based
   protocols such as token ring and FDDI.

   Figure 2 below outlines how spatial reuse works. In this example,
   node 1 is sending traffic to node 4, node 2 to node 3 and node 5 to
   node 6. Having the destination node strip unicast data from the ring
   allows other nodes on the ring who are downstream to have full access
   to the ring bandwidth. In the example given this means node 5 has
   full bandwidth access to node 6 while other traffic is being
   simultaneously transmitted on other parts of the ring.

2.3. Fairness

   Since the ring is a shared media, some sort of access control is
   necessary to ensure fairness and to bound latency. Access control can
   be broken into two types which can operate in tandem:

      Global access control - controls access so that everyone gets a
      fair share of the global bandwidth of the ring.

      Local access control - grants additional access beyond that
      allocated globally to take advantage of segments of the ring that
      are less than fully utilized.

   As an example of a case where both global and local access are
   required, refer again to Figure 2. Nodes 1, 2, and 5 will get 1/2 of
   the bandwidth on a global allocation basis. But from a local
   perspective, node 5 should be able to get all of the bandwidth since
   its bandwidth does not interfere with the fair shares of nodes 1 and
   2.

FIGURE 2. Global and Local Re-Use

```
                             . . . . . . . . . . . . . . . . . .
                             .                               .  .
                           -----                                .
            --------------->| N |----------------                .
            | --------------| 1 |<-------------  |               .
            | |             -----             |  |               .
            | |                               |  |               .
          -----                             -----                .
        .>| N |                             | N |. . .           .
        . | 6 |                             | 2 |   .            .
        . -----                             -----   .            .
        .   ^ |                               ^ |   .    .       .
        . o | |                             i | |   .    .       .
        . u | |                             n | |   .    .       .
        . t | |                             n | |   .    .       .
        . e | |                             e | |   .    .       .
        . r | |                             r | |   .    .       .
        .   | v                               | v   .    .       .
        .  -----                             -----  .    .       .
        . .| N |                             | N |<. .           .
           | 5 |                             | 3 |               .
           -----                             -----               .
            | |                               | |               .
            | |             -----             | |               .
            | --------------->| N |---------------  |           .
            ----------------| 4 |<----------------              .
                             -----                              .
                              ^                                 .
                              .                                 .
                             . . . . .<. . . . . . . . . . . . .
```
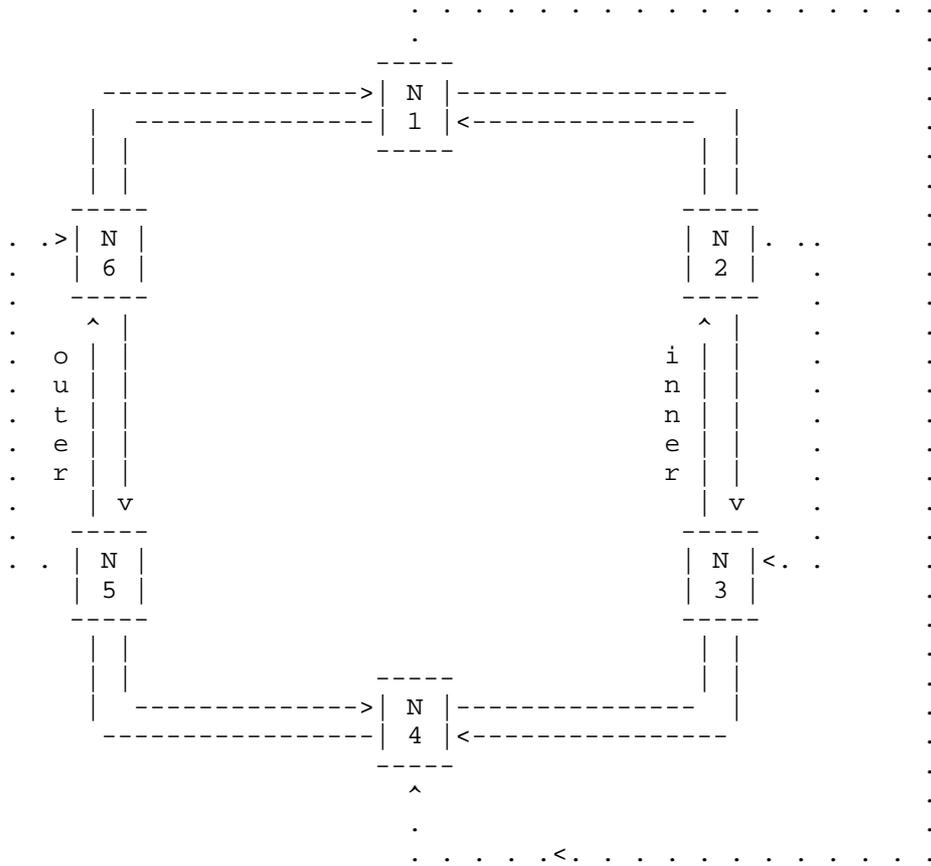
2.4. Transit Buffer

   To be able to detect when to transmit and receive packets from the
   ring, SRP makes use of a transit (sometimes referred as insertion)
   buffer as shown in Figure 3 below. High priority packets and low
   priority packets can be placed into separate fifo queues.

FIGURE 3. Transit buffer

```
              ^^                   ||
              ||                   vv
           |----|              |----|
           |    |              |    |
           |----|Rx            |----|Tx
           |    |Buffer        |    |Buffer
           |----|              |----|
           |    |              |    |
           |----|              |----|
           |    |              |    |
           |----|              |----|
           |    |              |    |
           |----|              |----|
            ^^     Transit      ||
            ||     Buffer       ||
            ||     |------|      vv
            ||     |  H   |
  ===========>|------|==========>
                   |  L   |
                   |------|
```
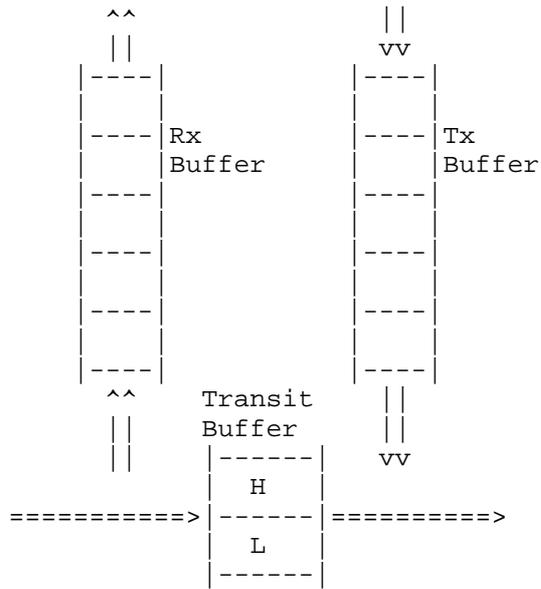
3. SRP Overview

3.1. Receive Operation Overview

   Receive Packets entering a node are copied to the receive buffer if a
   Destination Address (DA) match is made. If a DA matched packet is
   also a unicast, then the packet will be stripped. If a packet does
   not DA match or is a multicast and the packet does not Source Address
   (SA) match, then the packet is placed into the Transit Buffer (TB)
   for forwarding to the next node if the packet passes Time To Live and
   Cyclic Redundancy Check (CRC) tests.

3.2. Transmit Operation Overview

   Data sent from the node is either forwarded data from the TB or
   transmit data originating from the node via the Tx Buffer. High
   priority forwarded data always gets sent first. High priority
   transmit data may be sent as long as the Low Priority Transit Buffer
   (LPTB) is not full.

   A set of usage counters monitor the rate at which low priority
   transmit data and forwarded data are sent. Low priority data may be
   sent as long as the usage counter does not exceed an allowed usage
   governed by the SRP-fa rules and the LPTB has not exceeded the low
   priority threshold.

3.3. SRP Fairness Algorithm (SRP-fa) Overview

   If a node experiences congestion, then it will advertise to upstream
   nodes via the opposite ring the value of its transmit usage counter.
   The usage counter is run through a low pass filter function to
   stabilize the feedback. Upstream nodes will adjust their transmit
   rates so as not to exceed the advertised values. Nodes also
   propagate the advertised value received to their immediate upstream
   neighbor. Nodes receiving advertised values who are also congested
   propagate the minimum of their transmit usage and the advertised
   usage.

   Congestion is detected when the depth of the low priority transit
   buffer reaches a congestion threshold.

   Usage messages are generated periodically and also act as keepalives
   informing the upstream station that a valid data link exists.

3.4. Intelligent Protection Switching (IPS) Protocol Overview

   An SRP Ring is composed of two counter-rotating, single fiber rings.
   If an equipment or fiber facility failure is detected, traffic going
   towards and from the failure direction is wrapped (looped) back to go
   in the opposite direction on the other ring (subject to the
   protection hierarchy). The wrap around takes place on the nodes
   adjacent to the failure, under control of the IPS protocol. The wrap
   re-routes the traffic away from the failed span.

   An example of the data paths taken before and after a wrap are shown
   in Figures 4 and 5. Before the fiber cut, N4 sends to N1 via the
   path N4->N5->N6->N1.

   If there is a fiber cut between N5 and N6, N5 and N6 will wrap the
   inner ring to the outer ring. After the wraps have been set up,
   traffic from N4 to N1 initially goes through the non-optimal path
   N4->N5->N4->N3->N2->N1->N6->N1.

   Subsequently a new ring topology is discovered and a new optimal path
   is used N4->N3->N2-N1 as shown in Figure 6. Note that the topology
   discovery and the subsequent optimal path selection are not part of
   the IPS protocol.

FIGURE 4. Data path before wrap, N4 -> N1

```
                                    -----
          ###############>| N |----------------
          #  --------------| 1 |<-------------   |
          # |              -----              |  |
          # |                                 |  |
        -----                               -----
        | N |                               | N |
        | 6 |                               | 2 |
        -----                               -----
         ^  |                                ^  |
         #  |                                |  |
         #  |                                |  |
         #  |                                |  |
         #  |                                |  |
         #  |                                |  |
         #  |                                |  |
         #  v                                |  v
        -----                               -----
        | N |                               | N |
        | 5 |                               | 3 |
        -----                               -----
         #  |                                |  |
         #  |              -----             |  |
         #  -------------->| N |--------------  |
          ###############| 4 |<---------------
                          -----
```
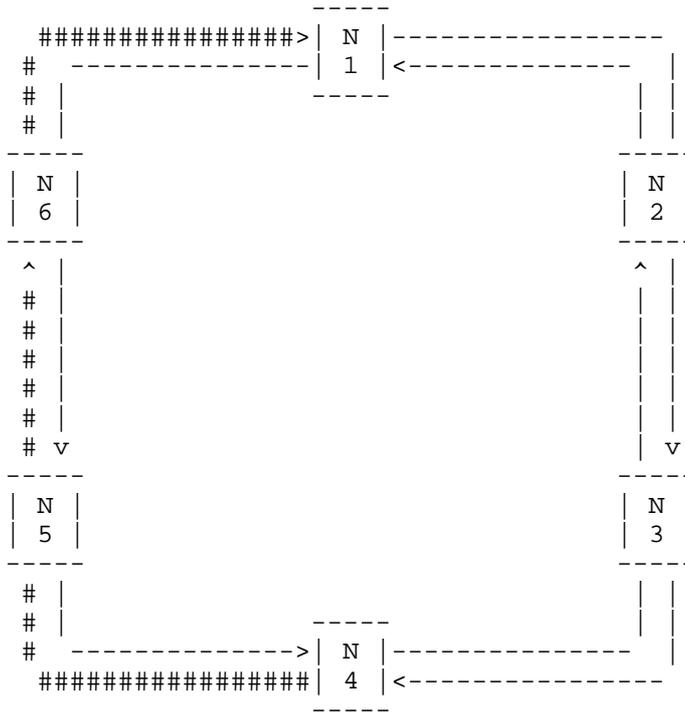
The ring wrap is controlled through SONET BLSR [3][4] style IPS
signaling. It is an objective to perform the wrapping as fast as in
the SONET equipment or faster.

The IPS protocol processes the following request types (in the order
of priority, from highest to lowest):

   1. Forced Switch (FS): operator originated, performs a protection
      switch on a requested span (wraps at both ends of the span)

   2. Signal Fail (SF): automatic, caused by a media Signal Failure
      or SRP keep-alive failure - performs a protection switch on a
      requested span

FIGURE 5. Data path after the wrap, N4 -> N1

```
                              -----
              ################>| N |----------------
              #  #############| 1 |<#############   |
              # #              -----             #  |
              # v                                #  |
            -----                              -----
            | N |                              | N |
            | 6 |                              | 2 |
            -----                              -----
             ^ # wrap                           ^  |
             ###                                #  |
          _____                            #  |
          fiber cut                             #  |
          ---------                             #  |
             ###                                #  |
             # v wrap                           # v
            -----                              -----
            | N |                              | N |
            | 5 |                              | 3 |
            -----                              -----
            # #                                 #  |
            # #              -----              #  |
            #  ##############>| N |##############   |
             ################| 4 |<---------------
```
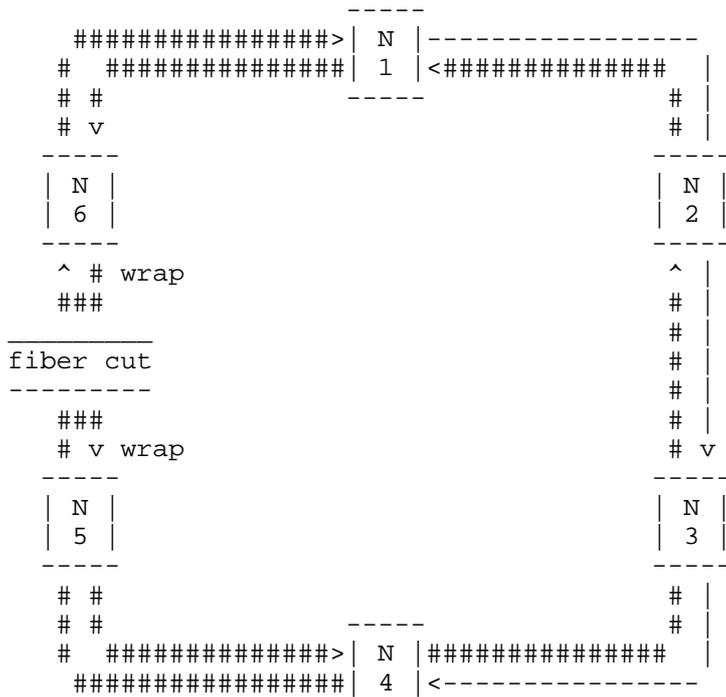
   3. Signal Degrade (SD): automatic, caused by a media Signal
      Degrade (e.g. excessive Bit Error Rate) - performs a protection
      switch on a requested span

   4. Manual Switch (MS): operator originated, like Forced Switched
      but of a lower priority

   5. Wait to Restore (WTR): automatic, entered after the working
      channel meets the restoration criteria after SF or SD condition
      disappears. IPS waits WTR period before restoring traffic in
      order to prevent protection switch oscillations

   If a protection (either automatic or operator originated) is
   requested for a given span, the node on which the protection has been
   requested issues a protection request to the node on the other end of
   the span using both the short path (over the failed span, as the
   failure may be unidirectional) and the long path (around the ring).

   FIGURE 6. Data path after the new topology is discovered

```
                                 -----
                ----------------| N |----------------
               |  --------------| 1 |<#############   |
               | |               -----             #  |
               | |                                 #  |
               | v                                 #  |
              -----                              -----
             | N |                              | N |
             | 6 |                              | 2 |
              -----                              -----
               ^ | wrap                          ^  |
               --                                #  |
                                                 #  |
             _____                           #  |
             fiber cut                           #  |
             ---------                           #  |
               --                                #  |
               | v wrap                          # v
              -----                              -----
             | N |                              | N |
             | 5 |                              | 3 |
              -----                              -----
               | |                               #  |
               | |               -----           #  |
               |  -------------->| N |##############   |
                ----------------| 4 |<----------------
                                 -----
```
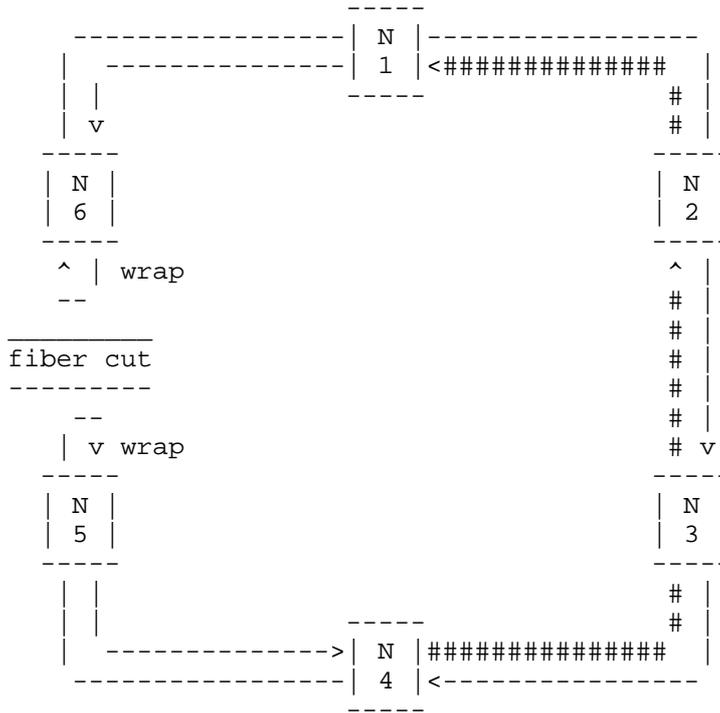
   As the protection requests travel around the ring, the protection
   hierarchy is applied. If the requested protection switch is of the
   highest priority e.g. Signal Fail request is of higher priority than
   the Signal Degrade than this protection switch takes place and the
   lower priority switches elsewhere in the ring are taken down, as
   appropriate. If a lower priority request is requested, it is not
   allowed if a higher priority request is present in the ring. The only
   exception is multiple SF and FS switches, which can coexist in the
   ring.

   All protection switches are performed bidirectionally (wraps at both
   ends of a span for both transmit and receive directions, even if a
   failure is only unidirectional).

4. Packet Formats

   This section describes the packet formats used by SRP. Packets can be
   sent over any point to point link layer (e.g. SONET/SDH, ATM, point
   to point ETHERNET connections). The maximum transfer unit (MTU) is
   9216 octets. The minimum transfer unit for data packets is 55
   octets. The maximum limit was designed to accommodate the large IP
   MTUs of IP over AAL5. SRP also supports ATM cells. ATM cells over
   SRP are 55 octets. The minimum limit corresponds to ATM cells
   transported over SRP. The minimum limit does not apply to control
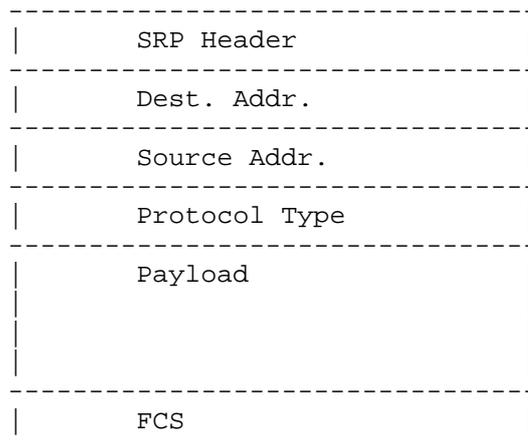   packets which may be smaller.

   These limits include everything listed in Figure 7: but are exclusive
   of the frame delineation (e.g. for SRP over SONET/SDH, the flags used
   for frame delineation are not included in the size limits).

   The following packet and cell formats do not include any layer 1
   frame delineation. For SRP over POS, there will be an additional
   flag that delineates start and end of frame.

4.1. Overall Packet Format

   The overall packet format is show below in Figure 7:

   FIGURE 7. Overall Packet Format

```
                    ---------------------------------
                    |         SRP Header            |
                    ---------------------------------
                    |         Dest. Addr.           |
                    ---------------------------------
                    |         Source Addr.          |
                    ---------------------------------
                    |         Protocol Type         |
                    ---------------------------------
                    |         Payload               |
                    |                               |
                    |                               |
                    |                               |
                    ---------------------------------
                    |         FCS                   |
                    ---------------------------------
```

   The frame check sequence (FCS) is a 32-bit cyclic redundancy check
   (CRC) as specified in RFC-1662 and is the same CRC as used in Packet
   Over SONET (POS - specified in RFC-2615). The generator polynomial
   is:

   CRC-32:

   x32 + x26 + x23 + x22 + x16 + x12 + x11 + x10 + x8 + x7 + x5 + x4 +
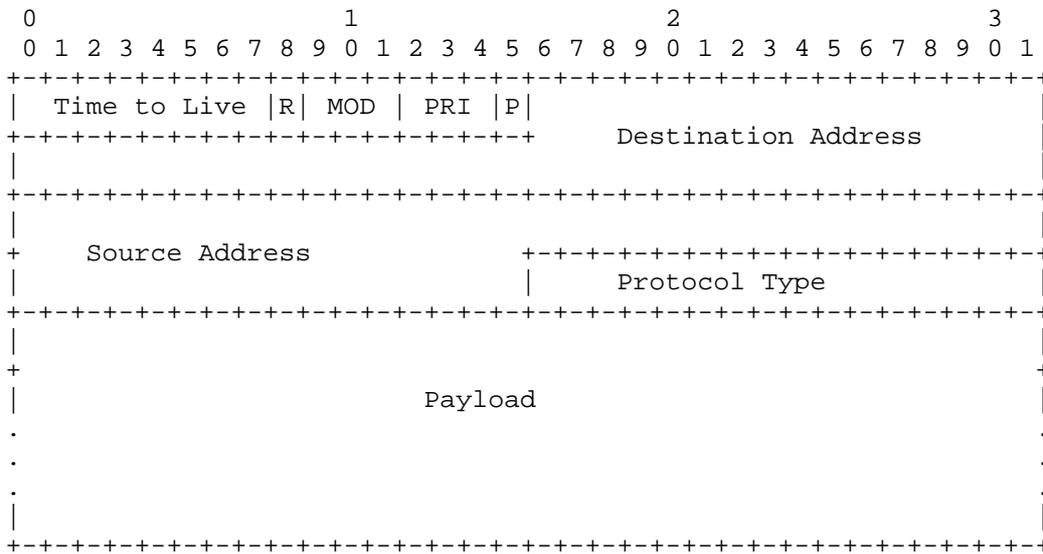   x2 + x + 1

   The FCS is computed over the destination address, source address,
   protocol type and payload. It does not include the SRP header.

   Note that the packet format after the SRP header is identical to
   Ethernet Version 2.

4.2. Generic Packet Header Format

   Each packet has a fixed-sized header. The packet header format is
   shown in Figure 8.

   FIGURE 8. Detailed Packet Header Format

```
       0                   1                   2                   3
       0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      | Time to Live |R| MOD | PRI |P|                               |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+    Destination Address          |
      |                                                               |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                                                               |
      +    Source Address             +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                               |       Protocol Type           |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                                                               |
      +                                                               +
      |                          Payload                              |
      .                                                               .
      .                                                               .
      .                                                               .
      |                                                               |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

   The fields are described below.

4.2.1. Time To Live (TTL)

   This 8 bit field is a hop-count that must be decremented every time a
   node forwards a packet. If the TTL reaches zero it is stripped off
   the ring. This allows for a total node space of 256 nodes on a ring.
   However, due to certain failure conditions (e.g. when the ring is

wrapped) the total number of nodes that are supported by SRP is 128.
When a packet is first sent onto the ring the TTL should be set to at
least twice the total number of nodes on the ring.

4.2.2. Ring Identifier (R)

   This single bit field is used to identify which ring this packet is
   designated for. The designation is as follows:

        TABLE 1. Ring Indicator Values

        Outer Ring   0
        Inner Ring   1

4.2.3. Priority Field (PRI)

   This three bit field indicates the priority level of the SRP packet
   (0 through 7). The higher the value the higher the priority. Since
   there are only two queues in the transit buffer (HPTB and LPTB) a
   packet is treated as either low or high priority once it is on the
   ring. Each node determines the threshold value for determining what
   is considered a high priority packet and what is considered a low
   priority packet. However, the full 8 levels of priority in the SRP
   header can be used prior to transmission onto the ring (transmit
   queues) as well as after reception from the ring (receive queues).

4.2.4. MODE

   This three bit field is used to identify the mode of the packet. The
   following modes are defined in Table 2 below.

        TABLE 2. MODE Values

        Value   Description

        000   Reserved
        001   Reserved
        010   Reserved
        011   ATM cell
        100   Control Message (Pass to host)
        101   Control Message (Locally Buffered for host)
        110   Usage Message
        111   Packet Data

   These modes will be further explained in later sections.

4.2.5. Parity Bit (P-bit)

   The parity bit is used to indicate the parity value over the 15 bits
   of the SRP header to provide additional data integrity over the
   header. Odd parity is used (i.e. the number of ones including the
   parity bit shall be an odd number).

4.2.6. Destination Address

   The destination address is a globally unique 48 bit address assigned
   by the IEEE.

4.2.7. Source Address

   The source address is a globally unique 48 bit address assigned by
   the IEEE.

4.2.8. Protocol Type

   The protocol type is a two octet field like that used in EtherType
   representation. Current defined values relevant to SRP are defined in
   Table 3 below.

        TABLE 3. Defined Protocol Types

        Value   Protocol Type

        0x2007   SRP Control
        0x0800   IP version 4
        0x0806   ARP

4.3. SRP Cell Format

   SRP also supports the sending of ATM cells. The detailed cell format
   is shown below:

FIGURE 9. SRP Cell Format

```
        0                   1                   2                   3
        0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       | Time to Live |R| MOD | PRI |P|            VPI/VCI            |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |      VCI          | PTI |C|    HEC       |                   |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+                   +
       |                                                             |
       .                                                             .
       .                        ATM    Payload                       .
       .                       ( 48 Bytes )       +-+-+-+-+-+-+-+-+
       |                                          |                   |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

   Packet nodes would typically ignore (never receive or strip) and
   always forward ATM-cells. The idea is that ATM switches and routers
   could coexist in a ring. Note that SRP cells do not contain an FCS.
   Data integrity is handled at the AAL layer.

4.4. SRP Usage Packet Format

   SRP usage packets are sent out periodically to propagate allowed
   usage information to upstream nodes. SRP usage packets also perform
   a keepalive function. SRP usage packets should be sent approximately
   every 106 usec.

   If a receive interface has not seen a usage packet within the
   keepalive timeout interval it will trigger an L2 keepalive timeout
   interrupt/event. The IPS software will subsequently mark that
   interface as faulty and initiate a protection switch around that
   interface. The keepalive timeout interval should be set to 16 times
   the SRP usage packet transmission interval.

FIGURE 10. Usage Packet Format

```
        0                   1                   2                   3
        0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       | Time to Live |R| MOD | PRI |P|                               |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+ Originator MAC Address       +
       |                                                             |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       | Reserved                      |     Usage                   |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

   A USAGE of all ones indicates a value of NULL.
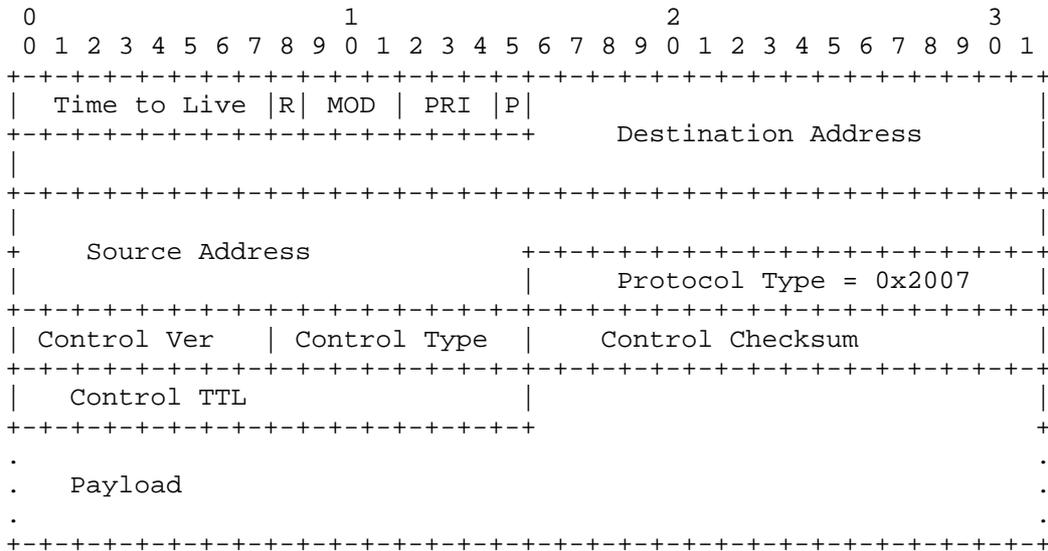
4.5. SRP Control Packet Format

   If the MODE bits are set to 10X (SRP control) then this indicates a
   control message. Control messages are always received and stripped by
   the adjacent node. They are by definition unicast, and do not need
   any addressing information. The destination address field for
   control packets should be set to 0's. The source address field for a
   control packet should be set to the source address of the
   transmitting node.

   Two types of controls messages are defined : Pass to host and Locally
   buffered. Pass to host messages can be passed to the host software by
   whatever means is convenient. This is most often the same path used
   to transfer data packets to the host. Locally buffered control
   messages are usually reserved for protection messages. These are
   normally buffered locally in order to not contend for resources with
   data packets. The actual method of handling these messages is up to
   the implementor.

   The control packet format is shown in Figure 11.

   FIGURE 11. Control Packet Format

```
           0                   1                   2                   3
           0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
          +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
          | Time to Live |R| MOD | PRI |P|                               |
          +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+       Destination Address       |
          |                                                               |
          +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
          |                                                               |
          +    Source Address              +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
          |                                |     Protocol Type = 0x2007    |
          +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
          | Control Ver  | Control Type   |    Control Checksum           |
          +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
          |   Control TTL                  |                               |
          +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+                               +
          .                                                               .
          .   Payload                                                     .
          .                                                               .
          +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

   The priority (PRI) value should be set to 0x7 (all one's) when
   sending control packets and should be queued to the highest priority
   transmit queue available. The Time to Live is not relevant since all

packets will be received and stripped by the nearest downstream
neighbor and can be set to any value (preferably this should be set
to 001).

4.5.1. Control Ver

This one octet field is the version number associated with the
control type field. Initially, all control types will be version 0.

4.5.2. Control Type

This one octet field represents the control message type. Table 4
contains the currently defined control types.

TABLE 4. Control Types

Control Type	Description
0x01	Topology Discovery
0x02	IPS message
0x03-0xFF	Reserved

4.5.3. Control TTL

The Control TTL is a control layer hop-count that must be decremented
every time a node forwards a control packet. If a node receives a
control packet with a control TTL <= 1, then it should accept the
packet but not forward it.

Note that the control layer hop count is separate from the SRP L2 TTL
which is always set to 1 for control messages.

The originator of the control message should set the initial value of
the control TTL to the SRP L2 TTL normally used for data packets.

4.5.4. Control Checksum

The checksum field is the 16 bit one's complement of the one's
complement sum of all 16 bit words starting with the control version.
If there are an odd number of octets to be checksummed, the last
octet is padded on the right with zeros to form a 16 bit word for
checksum purposes. The pad is not transmitted as part of the
segment. While computing the checksum, the checksum field itself is
replaced with zeros. This is the same checksum algorithm as that
used for TCP. The checksum does not cover the 32 bit SRP FCS.

4.5.5. Payload

    The payload is a variable length field dependent on the control type.

4.5.6. Addressing

    All nodes must have a globally unique IEEE 48 bit MAC address. A
    multicast bit is defined using canonical addressing conventions i.e.
    the multicast bit is the least significant bit of the most
    significant octet in the destination address. It is acceptable but
    not advisable to change a node's MAC address to one that is known to
    be unique within the administrative layer 2 domain (that is the SRP
    ring itself along with any networks connected to the SRP ring via a
    layer 2 transparent bridge).

    FIGURE 12. Multicast bit position

```
                    Destination Address
        0                   1                   2                   3
        0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
       |           |M|                                               |
       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
                    ^
                    |----Multicast bit
```
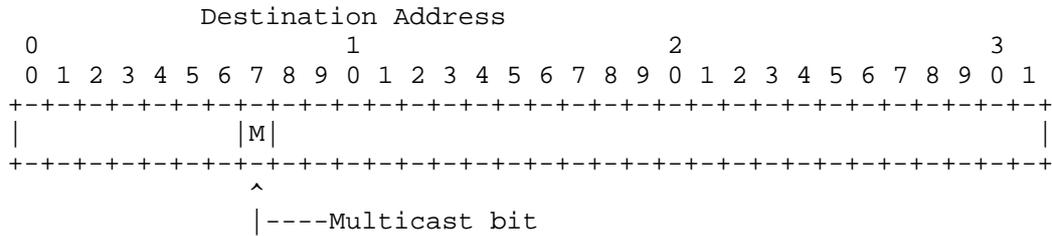
    Note that for SONET media, the network order is MSB of each octet
    first, so that as viewed on the line, the multicast bit will be the
    8th bit of the destination address sent. (For SRP on Ethernet media,
    the multicast bit would be sent first).

4.6. Topology Discovery

    Each node performs topology discovery by sending out topology
    discovery packets on one or both rings. The node originating a
    topology packet marks the packet with the egressing ring id, appends
    the node's mac binding to the packet and sets the length field in the
    packet before sending out the packet. This packet is a point-to-point
    packet which hops around the ring from node to node. Each node
    appends its mac address binding, updates the length field and sends
    it to the next hop on the ring. If there is a wrap on the ring, the
    wrapped node will indicate a wrap when appending its mac binding and
    wrap the packet. When the topology packets travel on the wrapped
    section with the ring identifier being different from that of the
    topology packet itself, the mac address bindings are not added to the
    packet.

Eventually the node that generated the topology discovery packet gets
back the packet. The node makes sure that the packet has the same
ingress and egress ring id before excepting the packet. A topology
map is changed only after receiving two topology packets which
indicate the same new topology (to prevent topology changes on
transient conditions).

Note that the topology map only contains the reachable nodes. It does
not correspond to the failure-free ring in case of wraps and ring
segmentations.

FIGURE 13. Topology Packet Format

```
     Topology

      0                   1                   2                   3
      0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     |  Time to Live  |R| MOD | PRI |P|                             |
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+        Destination Address    |
     |                                                              |
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     |                                                              |
     +     Source Address             +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     |                                |    Protocol Type = 0x2007    |
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     | Control Ver=0 | Control Type=1|     Control Checksum          |
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     |  Control TTL                   |      Topology Length         |
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     |        Originator's Globally Unique                          |
     +        MAC Address  (48 bits)  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     |                                |    MAC Type     |           |
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+           +
     |                  MAC Address (48 bits)                       |
     +              +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     |              |    Other MAC bindings                         |
     +-+-+-+-+-+-+-+-+                                              +
     |                                                              |
     +                                                              +
```
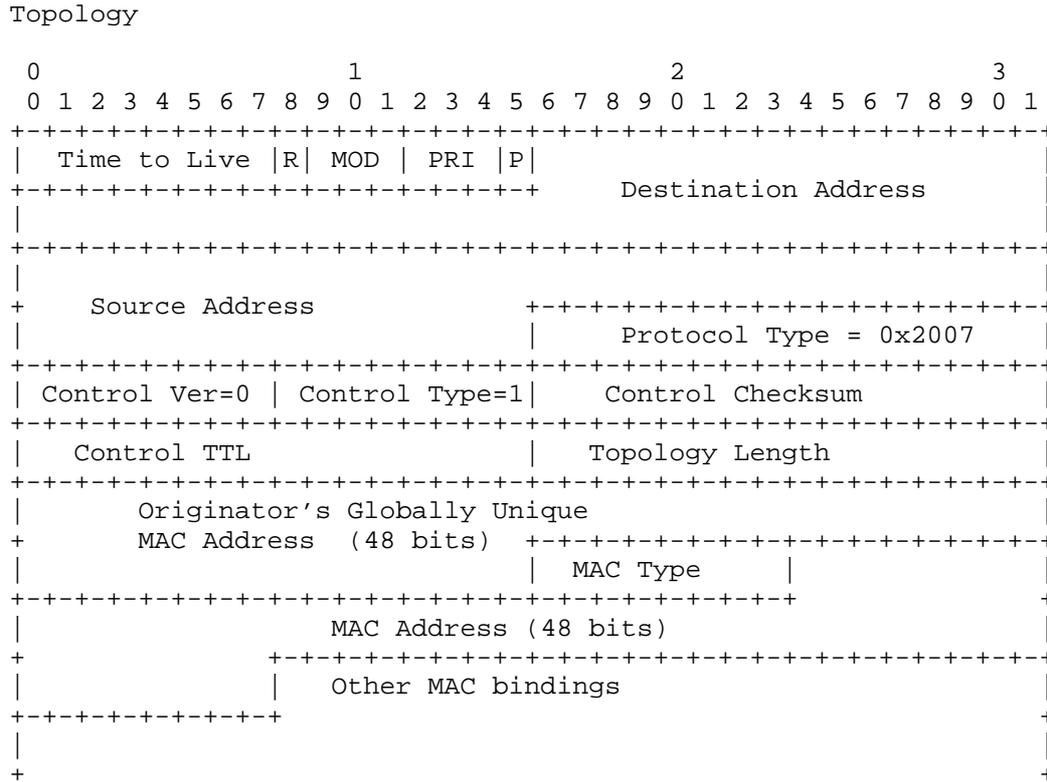
Note that the Source address should be set to the source address of
the TRANSMITTING node (which is not necessarily the ORIGINATING
node).

4.6.1. Topology Length

   This two octet field represents the length of the topology message in
   octets starting with the first MAC Type/MAC Address binding.

4.6.2. Topology Originator

   A topology discovery packet is determined to have been originated by
   a node if the originator's globally unique MAC address of the packet
   is that node's globally unique MAC address (assigned by the IEEE).

   Because the mac addresses could be changed at a node, the IEEE MAC
   address ensures that a unique identifier is used to determine that
   the topology packet has gone around the ring and is to be consumed.

4.6.3. MAC bindings

   Each MAC binding shall consist of a MAC Type field followed by the
   node's 48 bit MAC address. The first MAC binding shall be the MAC
   binding of the originator. Usually the originator's MAC address will
   be it's globally unique MAC Address but some implementations may
   allow this value to be overridden by the network administrator.

4.6.4. MAC Type Format

   This 8 bit field is encoded as follows:

        TABLE 5. MAC Type Format

        Bit   Value

        0   Reserved
        1   Ring ID (1 or 0)
        2   Wrapped Node (1) / Unwrapped Node (0)
        3-7   Reserved

   Determination of whether a packet's egress and ingress ring ID's are
   a match should be done by using the Ring ID found in the MAC Type
   field of the last MAC binding as the ingress ring ID rather than the
   R bit found in the SRP header. Although they should be the same, it
   is better to separate the two functions as some implementations may
   not provide the SRP header to upper layer protocols.

   The topology information is not required for the IPS protection
   mechanism. This information can be used to calculate the number of
   nodes in the ring as well as to calculate hop distances to nodes to
   determine the shortest path to a node (since there are two counter-
   rotating rings).

The implementation of the topology discovery mechanism could be a
periodic activity or on "a need to discover" basis. In the periodic
implementation, each node generates the topology packet periodically
and uses the cached topology map until it gets a new one. In the need
to discover implementation, each node generates a topology discovery
packet whenever they need one e.g., on first entering a ring or
detecting a wrap.

4.7. Intelligent Protection Switching (IPS)

   IPS is a method for automatically recovering from various ring
   failures and line degradation scenarios. The IPS packet format is
   outlined in Figure 14 below.

   FIGURE 14. IPS Packet Format

```
      0                   1                   2                   3
      0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     | Time to Live |R| MOD | PRI |P|                               |
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+    Destination Address          |
     |                                                              |
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     |                                                              |
     +     Source Address               +-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     |                                  |    Protocol Type = 0x2007  |
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     | Control Ver=0 | Control Type=2|   Control Checksum            |
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     |    Control TTL            |                                   |
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+                                   +
     |              Originator MAC Address                          |
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     |    Ips Octet   |   Rsvd Octet   |
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
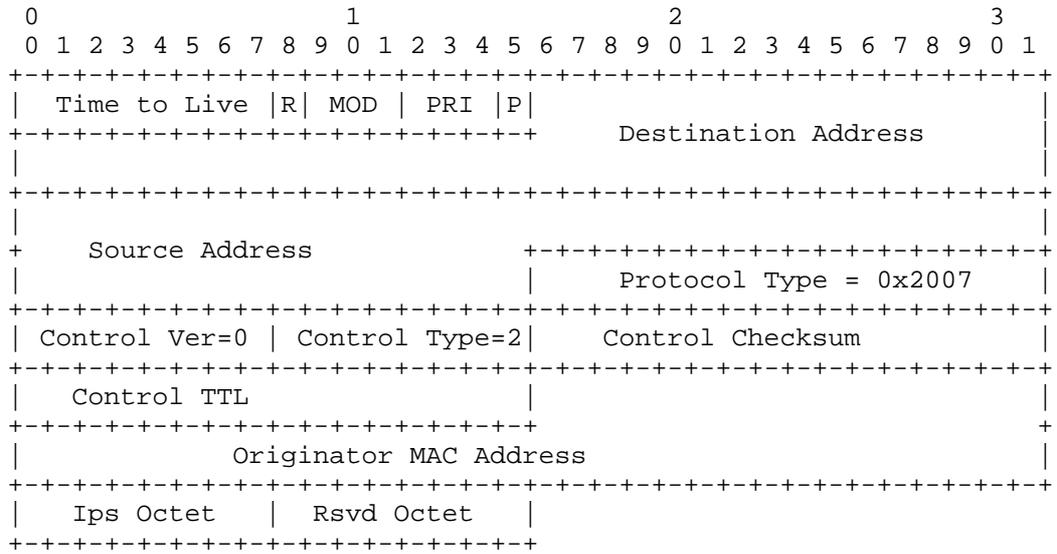
   The IPS specific fields are detailed below.

4.7.1. Originator MAC Address

   This is the MAC address of the originator of the IPS message. It is
   not necessarily the same as the SRP Header Source Address as a node
   may be simply propagating an IPS message (see the section "SRP IPS
   Protocol Rules" Rule P.8 as an example).

4.7.2. IPS Octet

   The IPS octet contains specific protection information. The format of
   the IPS octet is as follows:

   FIGURE 15. IPS Octet Format:

   Bits   Values (values not listed are reserved)

   0-3   IPS Request Type

           1101 - Forced Switch (FS)
           1011 - Signal Fail (SF)
           1000 - Signal Degrade (SD)
           0110 - Manual Switch (MS)
           0101 - Wait to Restore (WTR)
           0000 - No Request (IDLE)

   4   Path indicator

           0 - short (S)
           1 - long (L)

   5-7   Status Code

           010 - Protection Switch Completed - traffic Wrapped (W)
           000 - Idle (I)

   The currently defined request types with values, hierarchy and
   interpretation are as used in SONET BLSR [3], [4], except as noted.

4.8. Circulating packet detection (stripping)

   Packets continue to circulate when transmitted packets fail to get
   stripped. Unicast packets are normally stripped by the destination
   station or by the source station if the destination station has
   failed. Multicast packets are only stripped by the source station. If
   both the source and destination stations drop out of the ring while a
   unicast packet is in flight, or if the source node drops out while
   its multicast packet is in flight, the packet will rotate around the
   ring continuously.

   The solution to this problem is to have a TTL or Time To Live field
   in each packet that is set to at least twice the number of nodes in
   the ring. As each node forwards the packet, it decrements the TTL. If
   the TTL reaches zero it is stripped off of the ring.

The ring ID is used to qualify all stripping and receive decisions.
This is necessary to handle the case where packets are being wrapped
by some node in the ring. The sending node may see its packet on the
reverse ring prior to reaching its destination so must not source
strip it. The exception is if a node is in wrap. Logically, a node
in wrap "sees" the packet on both rings. However the usual
implementation is to receive the packet on one ring and to transmit
it on the other ring. Therefore, a node that is in the wrap state
ignores the ring ID when making stripping and receiving decisions.

A potential optimization would be to allow ring ID independent
destination stripping of unicast packets. One problem with this is
that packets may be delivered out of order during a transition to a
wrap condition. For this reason, the ring ID should always be used as
a qualifier for all strip and receive decisions.

5. Packet acceptance and stripping

A series of decisions based on the type of packet (mode), source and
destination addresses are made on the MAC incoming packets. Packets
can either be control or data packets. Control packets are stripped
once the information is extracted. The source and destination
addresses are checked in the case of data packets. The rules for
reception and stripping are given below as well as in the flow chart
in Figure 16.

1. Decrement TTL on receipt of a packet, discard if it gets to
   zero; do not forward.

2. Strip unicast packets at the destination station. Accept and
   strip "control" packets.

3. Do not process packets other than for TTL and forwarding if
   they have the "wrong" ring_id for the direction in which they
   are received unless the node is in wrap. If the node is in
   wrap then ignore the ring_id.

4. Do not process packets other than for TTL and forwarding if the
   mode is not supported by the node (e.g. reserved modes, or ATM
   cell mode for packet nodes).

5. Packets accepted by the host because of the destination address
   should be discarded at the upper level if there is CRC error.

6. Control messages are point to point between neighbors and
   should always be accepted and stripped.

        7. Packets whose source address is that of the receiving station
           and whose ring_id matches should be stripped. If a node is in
           wrap then ignore the ring_id.

    FIGURE 16. SRP Receive Flowchart (Packet node)

```
   if (MODE == 4,5)------------------------------->[to host]--->|
          |                                                     |
          v                                                     |
   if (MODE == 6)--------------------------------->[strip]----->|
          |                                                     |
          v                                                     |
   if (!WRAPPED                                                 |
      & WRONG_RING_ID)---------------------------------------|--->|
          |                                                     |  |
          v                                                     |  |
   if (MODE == 0,1,2,3)----------------------------------------|--->|
          |                                                     |  |
          v                                                     |  |
   if (DA MATCH)--------------->if !(SA MATCH)----->[to host]--->|  |
          |                          |                          |  |
          |                          v                          |  |
          |                     if (unicast)------->[to host]--->|  |
          |                          |                          |  |
          |                          v                          |  |
   if (SA MATCH)-------------------->[strip]------------------->|  |
          |                                                     |  |
          |                                                     |  v
          |--------------------------->|<---------------------|----|
                                       |                      |
                                       v                      |
                     if (ttl < 2)------->[strip]----->|        |
                                       |                      |
                                       v                      |
                     [decrement ttl]                          |
                                       |                      |
                     [fwd pkt to tb]                          |
                                       |                      v
                                       |<---------------------|
                                       v
                     [back to top]
```
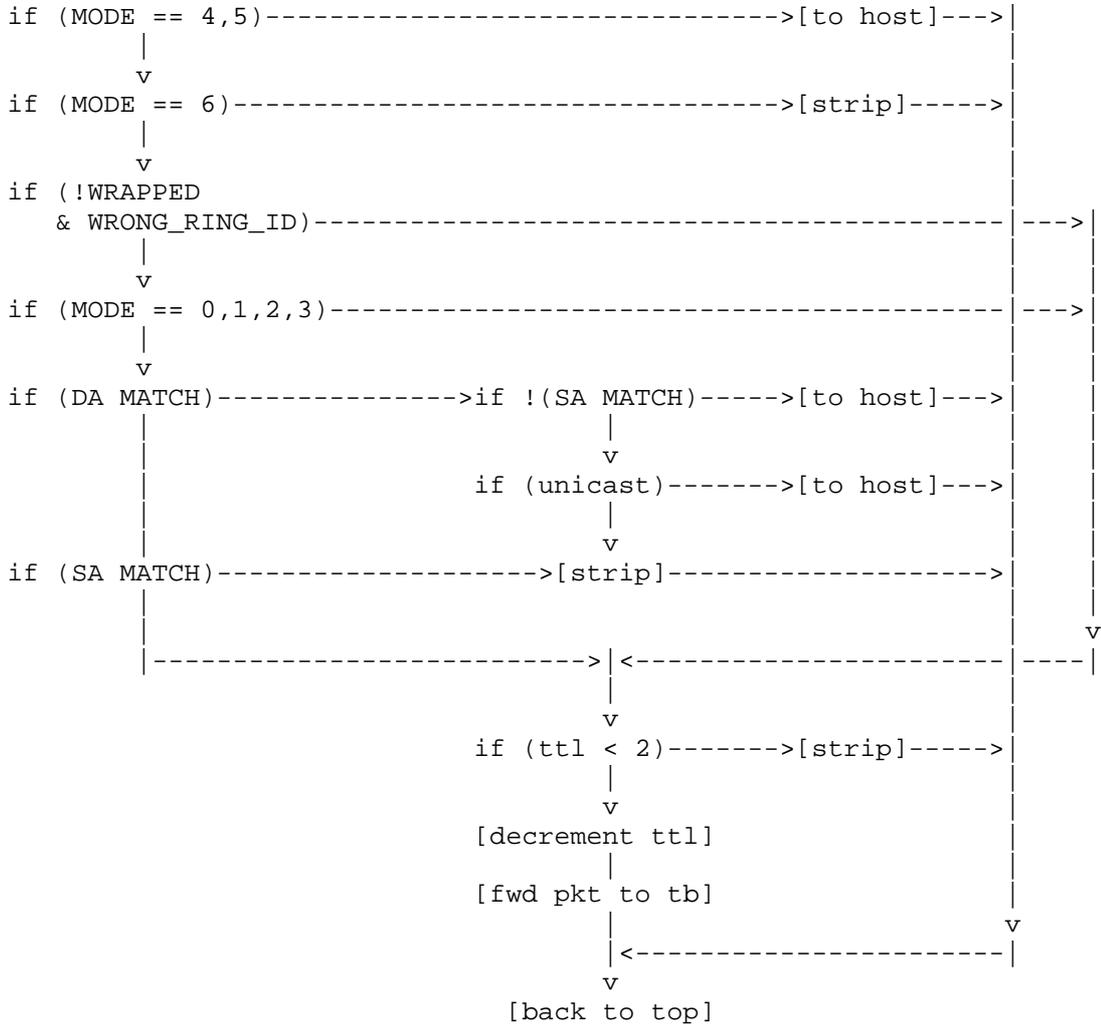
    Notes: Host is responsible for discarding CRC errored packets.
            Conditionals (if statements) branch to the right if true
            and branch down if false.

5.1. Transmission and forwarding with priority

    A node can transmit four types of packets:

        1. High priority packets from the high priority transit
           buffer.

        2. Low priority packets from the low priority transit buffer.

        3. High priority packets from the host Tx high priority fifo.

        4. Low priority packets from the host Tx low priority fifo.

    High priority packets from the transit buffer are always sent first.
    High priority packets from the host are sent as long as the low
    priority transit buffer is not full. Low priority packets are sent
    as long as the transit buffer has not crossed the low priority
    threshold and the SRP-fa rules allow it (my_usage < allowed_usage).
    If nothing else can be sent, low priority packets from the low
    priority transit buffer are sent.

    This decision tree is shown in Figure 17.

    FIGURE 17. SRP transmit flowchart

    if (TB_High has pkt)----------->[send pkt from TB_high]-->|
            | |
            v |
    if (TB_Low full)---|---->|
            | | |
            v | |
    if (Tx_High has pkt)----------->[send pkt from Tx_high]-->| |
            | | |
            v | |
    if (TB_Low > Hi threshold)------------------------------|---->|
            | | |
            v | |
    if (my_usage >= allowed_usage)--------------------------|---->|
            | | |
            v | |
    if (Tx_Low has pkt)------------>[send pkt from Tx_low]--->| |
            | | |
            | | v
            |<--|-----|
            | |
            v |
    if (TB_Low has pkt)------------>[send pkt from TB_low]--->|
            | v
            |<--|
            |
            v
        [Go to Top]

    Notes: Conditionals (if statements) branch to the right if true
            and branch down if false.

5.2. Wrapping of Data

    Normally, transmitted data is sent on the same ring to the downstream
    neighbor. However, if a node is in the wrapped state, transmitted
    data is sent on the opposite ring to the upstream neighbor.

6. SRP-fa Rules Of Operation

    The SRP-fa governs access to the ring. The SRP-fa only applies to
    low priority traffic. High priority traffic does not follow SRP-fa
    rules and may be transmitted at any time as long as there is
    sufficient transit buffer space.

The SRP-fa requires three counters which control the traffic
forwarded and sourced on the SRP ring. The counters are my_usage
(tracks the amount of traffic sourced on the ring), forward_rate
(amount of traffic forwarded on to the ring from sources other than
the host) and allowed_usage (the current maximum transmit usage for
that node).

With no congestion all nodes build up allowed usage periodically.
Each node can send up to max_usage. Max_usage is a per node
parameter than limits the maximum amount of low priority traffic a
node can send.

When a node sees congestion it starts to advertise its my_usage which
has been low pass filtered (lp_my_usage).

Congestion is measured by the transit buffer depth crossing a
congestion threshold.

A node that receives a non-null usage message (rcvd_usage) will set
its allowed usage to the value advertised. However, if the source of
the rcvd_usage is the same node that received it then the rcvd_usage
shall be treated as a null value. When comparing the rcvd_usage
source address the ring ID of the usage packet must match the
receiver's ring ID in order to qualify as a valid compare. The
exception is if the receive node is in the wrap state in which case
the usage packet's ring ID is ignored.

Nodes that are not congested and that receive a non-null rcvd_usage
generally propagate rcvd_usage to their upstream neighbor else
propagate a null value of usage (all 1's). The exception is when an
opportunity for local reuse is detected. Additional spatial reuse
(local reuse) is achieved by comparing the forwarded rate (low pass
filtered) to allow_usage. If the forwarded rate is less than the
allowed usage, then a null value is propagated to the upstream
neighbor.

Nodes that are congested propagate the smaller of lp_my_usage and
rcvd_usage.

Convergence is dependent upon number of nodes and distance.
Simulation has shown simulation convergence within 100 msec for rings
of several hundred miles.

6.1. SRP-fa pseudo-code

   A more precise definition of the fairness algorithm is shown below:

Variables:

lo_tb_depth   low priority transit buffer depth

my_usage   count of octets transmitted by host
lp_my_usage   my_usage run through a low pass filter
my_usage_ok   flag indicating that host is allowed to transmit

allow_usage   the fair amount each node is allowed to transmit

fwd_rate   count of octets forwarded from upstream
lp_fwd_rate   fwd_rate run through a low pass filter

congested   node cannot transmit host traffic without the TB buffer
                  filling beyond its congestion threshold point.

rev_usage   the usage value passed along to the upstream neighbor

Constants:

MAX_ALLOWANCE = configurable value for max allowed usage for this node

DECAY_INTERVAL = 8000 octet times @ OC-12, 32,000 octet times @ OC-48

AGECOEFF = 4   // Aging coeff for my_usage and fwd_rate

LP_FWD = 64   // Low pass filter for fwd_rate
LP_MU   = 512   // Low pass filter for my usage
LP_ALLOW = 64   // Low pass filter for allow usage auto increment

NULL_RCVD_INFO = All 1's in rcvd_usage field

TB_LO_THRESHOLD // TB depth at which no more lo-prio host traffic
                // can be sent

MAX_LRATE = AGECOEFF * DECAY_INTERVAL = 128,000 for OC-48, 32000 for
           OC-12

```
THESE ARE UPDATED EVERY CLOCK CYCLE:
====================================

my_usage          is incremented by 1 for every octet that is
                  transmitted by the host (does not include data
                  transmitted from the Transit Buffer).

fwd_rate          is incremented by 1 for every octet that enters the
                  Transit Buffer

if ((my_usage < allow_usage) &&
        !((lo_tb_depth > 0) && (fwd_rate < my_usage)) &&
              (my_usage < MAX_ALLOWANCE))
        // true means OK to send host packets
        my_usage_ok = true;


UPDATED WHEN USAGE_PKT IS RECEIVED:
===================================

if (usage_pkt.SA == my_SA) &&
        [(usage_pkt.RI == my_RingID) || (node_state == wrapped)]
        rcvd_usage = NULL_RCVD_INFO;
else
        rcvd_usage = usage_pkt.usage;


THE FOLLOWING IS CALCULATED EVERY DECAY_INTERVAL:
=================================================

congested = (lo_tb_depth > TB_LO_THRESHOLD/2)

lp_my_usage = ((LP_MU-1) * lp_my_usage + my_usage) / LP_MU

my_usage is decremented by min(allow_usage/AGECOEFF, my_usage/AGECOEFF)

lp_fwd_rate = ((LP_FWD-1) * lp_fwd_rate + fwd_rate) / LP_FWD

fwd_rate is decremented by fwd_rate/AGECOEFF

(Note: lp values must be calculated prior to decrement of non-lp
values).

if (rcvd_usage != NULL_RCVD_INFO)
        allow_usage = rcvd_usage;
else
        allow_usage += (MAX_LRATE - allow_usage) / (LP_ALLOW);
```

```
if (congested)
      {
        if (lp_my_usage < rcvd_usage)
                rev_usage = lp_my_usage;
        else
                rev_usage =  rcvd_usage;
      }
else if ((rcvd_usage != NULL_RCVD_INFO) &&
         (lp_fwd_rate > allow_usage)
    rev_usage = rcvd_usage;
else
    rev_usage = NULL_RCVD_INFO

if (rev_usage > MAX_LRATE)
        rev_usage = NULL_RCVD_INFO;
```

6.2. Threshold settings

   The low priority transit buffer (TB_LO_THRESHOLD) is currently sized
   to about 4.4 msec or 320 KB at OC12 rates. The TB_HI_THRESHOLD is
   set to about 870 usec higher than the TB_LO_THRESHOLD or at 458 KB at
   OC12 rates.

   The high priority transit buffer needs to hold 2 to 3 MTUs or about
   30KB.

7. SRP Synchronization

   Each node operates in "free-run" mode. That is, the receive clock is
   derived from the incoming receive stream while the transmit clock is
   derived from a local oscillator. This eliminates the need for
   expensive clock synchronization as required in existing SONET
   networks. Differences in clock frequency are accommodated by
   inserting a small amount of idle bandwidth at each nodes output.

   The clock source for the transmit clock shall be selected to deviate
   by no more than 20 ppm from the center frequency. The overall
   outgoing rate of the node shall be rate shaped to accommodate the
   worst case difference between receive and transmit clocks of adjacent
   nodes. This works as follows:

   A transit buffer slip count (tb_cnt) keeps track of the amount of
   octets inserted into the TB minus the amount of octets transmitted
   and is a positive integer.

   To account for a startup condition where a packet is being inserted
   into an empty TB and the node was otherwise idle the tb_cnt is reset
   if the transmit interface is idle. Idle is defined as no data being
   sent even though there is opportunity to send (i.e. the transmit
   interface is not prohibited from transmitting by the physical layer).

   An interval counter defines the sample period over which rate shaping
   is performed. This number should be sufficiently large to get an
   accurate rate shaping.

   A token_bucket counter implements the rate shaping and is a signed
   integer. We increment this counter by one of two fixed values called
   quantums each sample period. Quantum1 sets the rate at (Line_rate -
   Delta) where delta is the clock inaccuracy we want to accommodate.

   Quantum2 sets the rate at (Line_rate + Delta). If at the beginning
   of a sample period, tb_cnt >= sync_threshold, then we set the rate to
   Quantum2. This will allow us to catch up and causes the TB slip count
   to eventually go < sync_threshold. If tb_cnt is < sync_threshold
   then we set the rate to Quantum1.

   When the input rate and output rates are exactly equal, the tb_cnt
   will vary between sync_threshold > tb_cnt >= 0. This will vary for
   each implementation dependent upon the burst latencies of the design.
   The sync_threshold value should be set so that for equal transmit and
   receive clock rates, the transmit data rate is always Line_rate-Delta
   and will be implementation dependent.

   The token_bucket is decremented each time data is transmitted. When
   token_bucket reaches a value <= 0, a halt_transmit flag is asserted
   which halts further transmission of data (halting occurs on a packet
   boundary of course which can cause token_bucket to become a negative
   number).

7.1. SRP Synchronization Examples

   Assume an interval of 2^^18 or 262144 clock cycles. A Quantum1 value
   must be picked such that the data rate will = (LINE_RATE - DELTA). A
   Quantum2 value must be picked and used if the tb_cnt shows that the
   incoming rate is greater than the outgoing rate and is = (LINE_RATE +
   DELTA). Assume that the source of the incoming and outgoing rate
   clocks are +/- 100 ppm.

   For an OC12c SPE rate of 600 Mbps and a system clock rate of 800 Mbps
   (16 bits @ 50 Mhz). The system clock rate is the rate at which the
   system transmits bytes to the framer (in most cases the framer
   transmit rate is asynchronous from the rate at which the system
   transfers data to the framer).

```
Quantum1/Interval * 800 Mbps = 600 Mbps(1 - Delta)
Quantum1 = Interval * (600/800) * (1 - Delta)
Quantum1 = Interval * (600/800) * (1 - 1e-4) = 196588

Quantum2/Interval * 800 Mbps = 600 Mbps(1 + Delta)
Quantum2 = Interval * (600/800) * (1 + Delta)
Quantum2 = Interval * (600/800) * (1 + 1e-4) = 196628
```

Note: The actual data rate for OC-12c is 599.04 Mbps.

8. IPS Protocol Description

An SRP ring is composed of two counter-rotating, single fiber rings.
If an equipment or fiber facility failure is detected, traffic going
towards and from the failure direction is wrapped (looped) back to go
in the opposite direction on the other ring. The wrap around takes
place on the nodes adjacent to the failure, under software control.
This way the traffic is re-routed from the failed span.

Nodes communicate between themselves using IPS signaling on both
inner and outer ring.

The IPS octet contains specific protection information. The format of
the IPS octet is as follows:

FIGURE 18. IPS Octet format:

0-3   IPS Request Type

        1101 - Forced Switch (FS)
        1011 - Signal Fail (SF)
        1000 - Signal Degrade (SD)
        0110 - Manual Switch (MS)
        0101 - Wait to Restore (WTR)
        0000 - No Request (IDLE)

4   Path indicator

        0 - short (S)
        1 - long (L)

5-7   Status Code

        010 - Protection Switch Completed -traffic Wrapped (W)
        000 - Idle (I)

   The IPS control messages are shown in this document as:

   {REQUEST_TYPE, SOURCE_ADDRESS, WRAP_STATUS, PATH_INDICATOR}

8.1. The IPS Request Types

   The following is a list of the request types, from the highest to the
   lowest priority. All requests are signaled using IPS control
   messages.

      1. Forced Switch (FS - operator originated)

         This command performs the ring switch from the working channel
         to the protection, wrapping the traffic on the node at which
         the command is issued and at the adjacent node to which the
         command is destined. Used for example to add another node to
         the ring in a controlled fashion.

      2. Signal Fail (SF - automatic)

         Protection caused by a media "hard failure" or SRP keep- alive
         failure. SONET examples of SF triggers are: Loss of Signal
         (LOS), Loss of Frame (LOF), Line Bit Error Rate (BER) above a
         preselected SF threshold, Line Alarm Indication Signal (AIS).
         Note that the SRP keep-alive failure provides end-to-end
         coverage and as a result SONET Path triggers are not necessary.

      3. Signal Degrade (SD - automatic)

         Protection caused by a media "soft failure". SONET example of a
         SD is Line BER or Path BER above a preselected SD threshold.

      4. Manual Switch (MS - operator originated)

         Like the FS, but of lower priority. Can be used for example to
         take down the WTR.

      5. Wait to Restore (WTR - automatic)

         Entered after the working channel meets the restoration
         threshold after an SD or SF condition disappears. IPS waits WTR
         timeout before restoring traffic in order to prevent protection
         switch oscillations.

8.2. SRP IPS Protocol States

   Each node in the IPS protocol is in one of the following states for
   each of the rings:

8.2.1. Idle

   In this mode the node is ready to perform the protection switches and
   it sends to both neighboring nodes "idle" IPS messages, which include
   "self" in the source address field {IDLE, SELF, I, S}

8.2.2. Pass-through

   Node participates in a protection switch by passing the wrapped
   traffic and long path signaling through itself. This state is entered
   based on received IPS messages. If a long path message with not null
   request is received and if the node does not strip the message (see
   Protocol Rules for stripping conditions) the node decrements the TTL
   and retransmits the message without modification. Sending of the
   Idle messages is stopped in the direction in which the message with
   not null request is forwarded.

8.2.3. Wrapped

   Node participates in a protection switch with a wrap present. This
   state is entered based on a protection request issued locally or
   based on received IPS messages.

8.3. IPS Protocol Rules

8.3.1. SRP IPS Packet Transfer Mechanism

   R T.1:

   IPS packets are transferred in a store and forward mode between
   adjacent nodes (packets do not travel more than 1 hop between nodes
   at a time). Received packet (payload portion) is passed to software
   based on interrupts.

   R T.2:

   All IPS messages are sent to the neighboring nodes periodically on
   both inner and outer rings. The timeout period is configurable 1-600
   sec (default 1 sec). It is desirable (but not required) that the
   timeout is automatically decreased by a factor of 10 for the short
   path protection requests.

8.3.2. SRP IPS Signaling and Wrapping Mechanism

   R S.1:

   IPS signaling is performed using IPS control packets as defined in
   Figure 14 "IPS Packet Format".

   R S.2:

   Node executing a local request signals the protection request on both
   short (across the failed span) and long (around the ring) paths after
   performing the wrap.

   R S.3:

   Node executing a short path protection request signals an idle
   request with wrapped status on the short (across the failed span)
   path and a protection request on the long (around the ring) path
   after performing the wrap.

   R S.4:

   A node which is neither executing a local request nor executing a
   short path request signals IDLE messages to its neighbors on the ring
   if there is no long path message passing through the node on that
   ring.

   R S.5:

   Protection IPS packets are never wrapped.

   R S.6:

   If the protocol calls for sending both short and long path requests
   on the same span (for example if a node has all fibers disconnected),
   only the short path request should be sent.

   R S.7:

   A node wraps and unwraps only on a local request or on a short path
   request. A node never wraps or unwraps as a result of a long path
   request. Long path requests are used only to maintain protection
   hierarchy. (Since the long path requests do not trigger protection,
   there is no need for destination addresses and no need for topology
   maps)

In Figure 19, Node A detects SF (local request/ self-detected
request) on the span between Node A and Node B and starts sourcing
{SF, A, W, S} on the outer ring and {SF, A, W, L} on the inner ring.
Node B receives the protection request from Node A (short path
request) and starts sourcing {IDLE, B, W, S} on the inner ring and
{SF, B, W, L} on the outer ring.

FIGURE 19. SRP IPS Signaling

```
    {SF,A,W,S}
              -------------------------------
             |  -----X--------------------   |
             | |       fiber                | |
             | v       cut        {IDLE,B,W,S}| v
            -----                     -----
           | A |                     | B |
           |   |                     |   |
            -----                     -----
             ^ | {SF,A,W,L}        i ^ | o {SF,B,W,L}
             | |                   n | | u
             | |                   n | | t
             | |                   e | | e
             | v                   r | v r
```

8.4.   SRP IPS Protocol Rules

   R P.1:

   Protection Request Hierarchy is as follows (Highest priority to the
   lowest priority). In general a higher priority request preempts a
   lower priority request within the ring with exceptions noted as
   rules. The 4 bit values below correspond to the REQUEST_TYPE field in
   the IPS packet.

        1101 - Forced Switch (FS)
        1011 - Signal Fail (SF)
        1000 - Signal Degrade (SD)
        0110 - Manual Switch (MS)
        0101 - Wait to Restore (WTR)
        0000 - No Request (IDLE): Lowest priority

   R P.2:

   Requests >= SF can coexist.

R P.3:

Requests < SF can not coexist with other requests.

R P.4:

A node always honors the highest of {short path request, self
detected request} if there is no higher long path message passing
through the node.

R P.5:

When there are more requests of priority < SF, the first request to
complete long path signaling will take priority.

R P.6:

A Node never forwards an IPS packet received by it which was
originally generated by the node itself (it has the node's source
address).

R P.7:

Nodes never forward packets with the PATH_INDICATOR set to SHORT.

R P.8:

When a node receives a long path request and the request is >= to the
highest of {short path request, self detected request}, the node
checks the message to determine if the message is coming from its
neighbor on the short path. If that is the case then it does not
enter pass-thru and it strips the message.

R P.9:

When a node receives a long path request, it strips (terminates) the
request if it is a wrapped node with a request >= than that in the
request; otherwise it passes it through and unwraps.

R P.10:

Each node keeps track of the addresses of the immediate neighbors
(the neighbor node address is gleaned from the short path IPS
messages).

R P.11:

When a wrapped node (which initially detected the failure) discovers
disappearance of the failure, it enters WTR (user-configurable WTR
time-period). WTR can be configured in the 10-600 sec range with a
default value of 60 sec.

R P.12:

When a node is in WTR mode, and detects that the new neighbor (as
identified from the received short path IPS message) is not the same
as the old neighbor (stored at the time of wrap initiation), the node
drops the WTR.

R P.13:

When a node is in WTR mode and long path request Source is not equal
to the neighbor Id on the opposite side (as stored at the time of
wrap initiation), the node drops the WTR.

R P.14:

When a node receives a local protection request of type SD or SF and
it cannot be executed (according to protocol rules) it keeps the
request pending. (The request can be kept pending outside of the
protection protocol implementation).

R P.15:

If a local non-failure request (WTR, MS, FS) clears and if there are
no other requests pending, the node enters idle state.

R P.16:

If there are two failures and two resulting WTR conditions on a
single span, the second WTR to time out brings both the wraps down
(after the WTR time expires a node does not unwrap automatically but
waits till it receives idle messages from its neighbor on the
previously failed span)

R P.17:

If a short path FS request is present on a given side and a SF/SD
condition takes place on the same side, accept and process the SF/SD
condition ignoring the FS. Without this rule a single ended wrap
condition could take place. (Wrap on one end of a span only).

8.5. State Transitions

   Figure 20 shows the simplified state transition diagram for the IPS
   protocol:

   FIGURE 20. Simplified State Transitions Diagram

                          Local FS,SF,SD,MS req.
            ---------   or Rx{REQ,SRC,W,S} from mate
           | IDLE |---
           | |<------------------------------------ |
            ---------   Local REQ clears | |
              ^ | or Rx{IDLE,SRC,I,S} | |
              | | | |
              | | | |
              | | | |
              | | | |
Rx{IDLE,SRC,I,S}| | Rx{REQ,SRC,W,L} | |
              | | | |
              | | | |
              | v   Local FS,SF,SD,MS REQ > Active req. | v
            --------- or Rx{REQ,SRC,W,S},REQ > Active req. ---------
           | PASS |------------------------------------>| WRAPPED |
           | THRU |<------------------------------------| |
            ---------   ---------
            Forwards   Tx{REQ,SELF,W,S} for local REQ
            {REQ,SRC,W,L}   Tx{IDLE,SELF,W,S} for mate REQ
                                        & Tx{REQ,SELF,W,L}
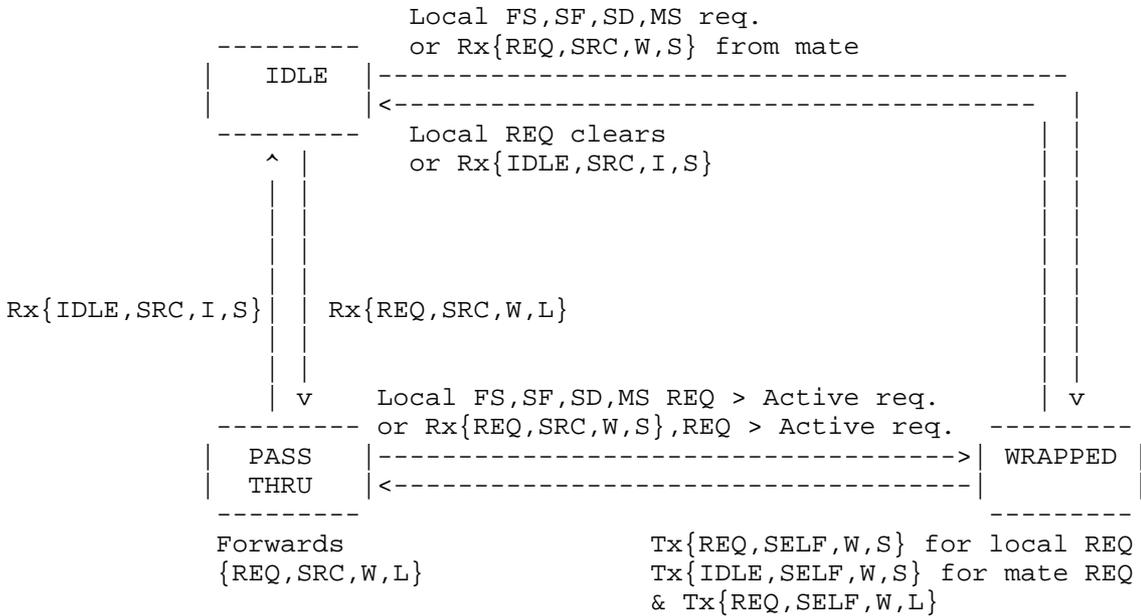
   Legend: Mate = node on the other end of the affected span
           REQ = {FS | SF | SD | MS}

8.6. Failure Examples
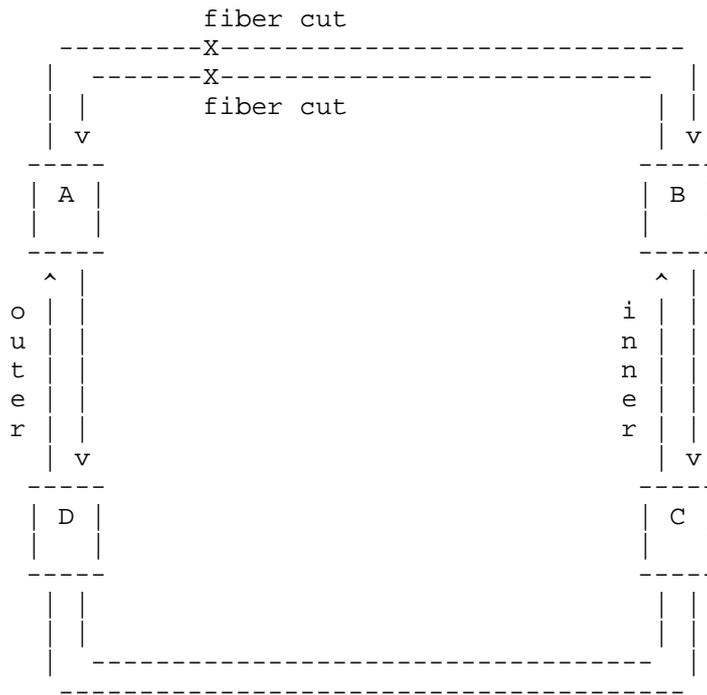
8.6.1. Signal Failure - Single Fiber Cut Scenario

   Sample scenario in a ring of four nodes A, B, C and D, with
   unidirectional failure on a fiber from A to B, detected on B. Ring is
   in the Idle state (all nodes are Idle) prior to failure.

   Signal Fail Scenario

   1. Ring in Idle, all nodes transmit (Tx) {IDLE, SELF, I, S} on both
        rings (in both directions)

   FIGURE 21. An SRP Ring with outer ring fiber cut

```
                fiber cut
         ---------X----------------------------
        |  ----------------------------------  |
        | |                                  | |
        | v                                  | v
       -----                              -----
      | A |                              | B |
      |   |                              |   |
       -----                              -----
       ^ |                                ^ |
     o | |                              i | |
     u | |                              n | |
     t | |                              n | |
     e | |                              e | |
     r | |                              r | |
       | v                                | v
       -----                              -----
      | D |                              | C |
      |   |                              |   |
       -----                              -----
        | |                                | |
        | |                                | |
        |  ----------------------------------  |
         --------------------------------------
```
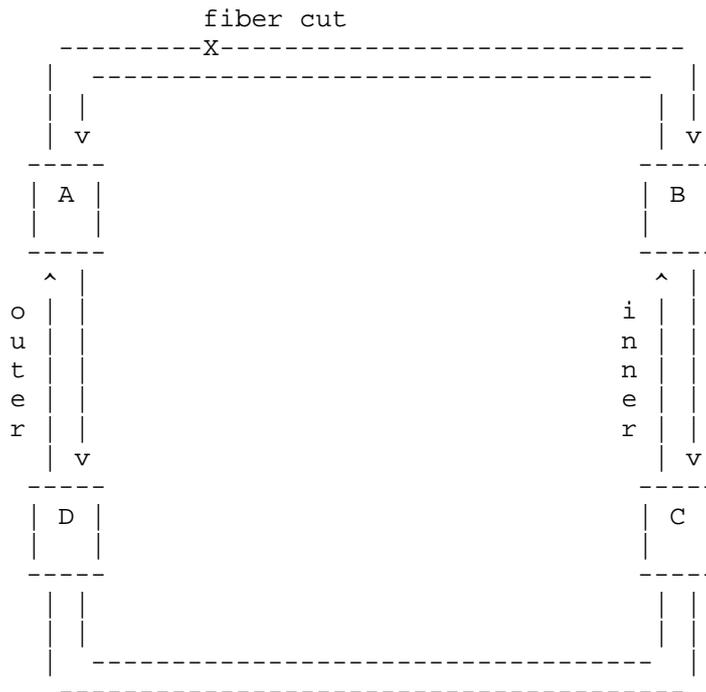
   2. B detects SF on the outer ring, transitions to Wrapped state
      (performs a wrap), Tx towards A on the inner ring/short path:
      {SF, B, W, S} and on the outer ring/long path: Tx {SF, B, W, L}

   3. Node A receives protection request on the short path,
      transitions to Wrapped state, Tx towards B on short path:
      {IDLE, A, W, S} (message does not go through due to the
      failure) and on the long path: Tx {SF, A, W, L}

   4. As the nodes D and C receive a switch request, they enter a
      pass-through mode (in each direction) which mean they stop
      sourcing the Idle messages and start passing the messages
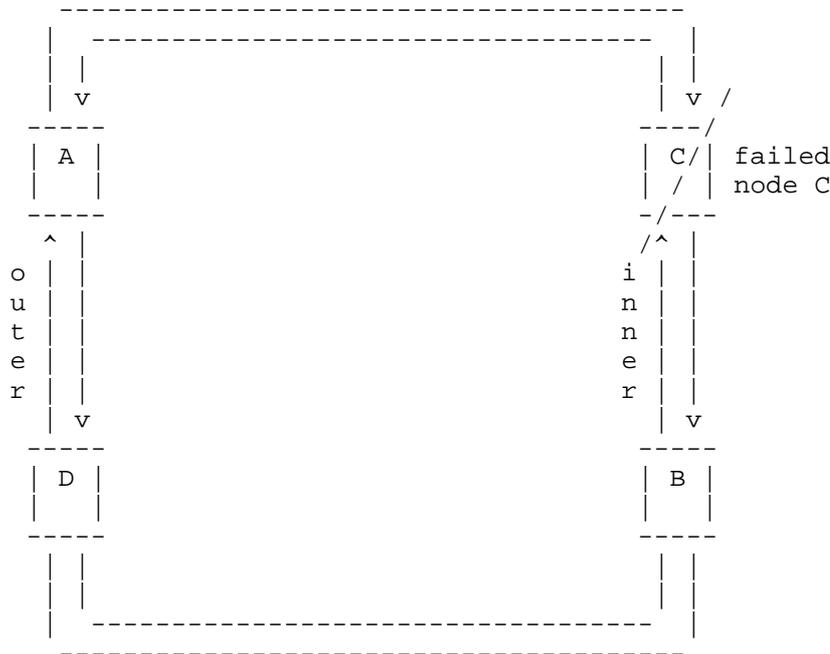   between A an B

   5. Steady state is reached

Signal Fail Clears

   1. SF on B clears, B does not unwrap, sets WTR timer, Tx {WTR, B,
      W, S} on inner and Tx {WTR, B, W, L}

   2. Node A receives WTR request on the short path, does not unwrap,
      Tx towards B on short path: {IDLE, A, W, S} (message does not
      go through due to the failure) and on the long path: Tx {WTR,
      A, W, L}

   3. Nodes C and D relay long path messages without changing the IPS
      octet

   4. Steady state is reached

   5. WTR times out on B. B transitions to idle state (unwraps) Tx
      {IDLE, B, I, S} on both inner and outer rings

   6. A receives Rx {IDLE, B, I, S} and transitions to Idle

   7. As idle messages reach C and D the nodes enter the idle state
      (start sourcing the Idle messages)

   8. Steady state it reached

8.6.2. Signal Failure - Bidirectional Fiber Cut Scenario

   Sample scenario in a ring of four nodes A, B, C and D, with a
   bidirectional failure between A and B. Ring is in the Idle state
   (all nodes are Idle) prior to failure.

   Signal Fail Scenario

   1. Ring in Idle, all nodes transmit (Tx) {IDLE, SELF, I, S} on
      both rings (in both directions)

   2. A detects SF on the outer ring, transitions to Wrapped state
      (performs a wrap), Tx towards B on the inner ring/short path:
      {SF, A, W, S} and on the outer ring/long path: Tx {SF, A, W, L}

   3. B detects SF on the outer ring, transitions to Wrapped state
      (performs a wrap), Tx towards A on the inner ring/short path:
      {SF, B, W, S} and on the outer ring/long path: Tx {SF, B, W, L}

   FIGURE 22. An SRP Ring with bidirectional fiber cut

```
                        fiber cut
            ---------X----------------------------
           | -------X--------------------------- |
           | |           fiber cut              | |
           | v                                  | v
           -----                                -----
          | A |                                | B |
          |   |                                |   |
           -----                                -----
           ^ |                                  ^ |
           o | |                                i | |
           u | |                                n | |
           t | |                                n | |
           e | |                                e | |
           r | |                                r | |
           | v                                  | v
           -----                                -----
          | D |                                | C |
          |   |                                |   |
           -----                                -----
           | |                                  | |
           | |                                  | |
           | ---------------------------------   |
           --------------------------------------
```

   4. As the nodes D and C receive a switch request, they enter a
      pass-through mode (in each direction) which mean they stop
      sourcing the Idle messages and start passing the messages
      between A an B

   5. Steady state is reached

   Signal Fail Clears

   1. SF on A clears, A does not unwrap, sets WTR timer, Tx {WTR, A,
      W, S} towards B and Tx {WTR, A, W, L} on the long path

   2. SF on B clears, B does not unwrap. Since it now has a short
      path WTR request present from A it acts upon this request. It
      keeps the wrap, Tx {IDLE, B, W, S} towards A and Tx {WTR, B, W,
      L} on the long path

   3. Nodes C and D relay long path messages without changing the IPS
      octet

   4. Steady state is reached

   5. WTR times out on A. A enters the idle state (drops wraps) and
      starts transmitting idle in both rings

   6. B sees idle request on short path and enters idle state

   7. Remaining nodes in the ring enter the idle state

   8. Steady state is reached

8.6.3. Failed Node Scenario

   FIGURE 23. An SRP Ring with a failed node

```
              --------------------------------------
             | ---------------------------------    |
             | |                               | |  |
             | v                               | v /
            -----                             ----/
           | A |                             | C/| failed
           |   |                             | / | node C
            -----                             -/---
             ^ |                              /^ |
           o | |                             i | | |
           u | |                             n | | |
           t | |                             n | | |
           e | |                             e | | |
           r | |                             r | | |
             | v                               | v
            -----                             -----
           | D |                             | B |
           |   |                             |   |
            -----                             -----
             | |                               | |
             | |                               | |
             | ---------------------------------    |
              --------------------------------------
```

   Sample scenario in a ring where node C fails. Ring is in the Idle
   state (all nodes are Idle) prior to failure.

Node Failure (or fiber cuts on both sides of the node)

   1. Ring in Idle, all nodes transmit (Tx) {IDLE, SELF, I, S} on
      both rings (in both directions)

   2. Based on the source field of the idle messages, all nodes
      identify the neighbors and keep track of them

   3. B detects SF on the outer ring, transitions to Wrapped state
      (performs a wrap), Tx towards C on the inner ring/short path:
      {SF, B, W, S} and on the outer ring/long path: Tx {SF, B, W, L}

   4. A detects SF on the inner ring, transitions to Wrapped state
      (performs a wrap), Tx towards C on the outer ring/short path:
      {SF, A, W, S} and on the inner ring/long path: Tx {SF, A, W, L}

   5. As the nodes on the long path between A and B receive a SF
      request, they enter a pass-through mode (in each direction),
      stop sourcing the Idle messages and start passing the messages
      between A an B

   6. Steady state is reached

Failed Node and One Span Return to Service

Note: Practically the node will always return to service with one
span coming after the other (with the time delta potentially close to
0). Here, a node is powered up with the fibers connected and fault
free.

   1. Node C and a span between A and C return to service (SF between
      A and C disappears)

   2. Node C, not seeing any faults starts to source idle messages
      {IDLE, C, I, S} in both directions.

   3. Fault disappears on A and A enters a WTR (briefly)

   4. Node A receives idle message from node C. Because the long path
      protection request {SF, B, W, L} received over the long span is
      not originating from the short path neighbor (C), node A drops
      the WTR and enters a PassThrough state passing requests between
      C and B

   5. Steady state is reached

   Second Span Returns to Service

   The scenario is like the Bidirectional Fiber Cut fault clearing
   scenario.

8.6.4. Bidirectional Fiber Cut and Node Addition Scenarios

   FIGURE 24. An SRP Ring with a failed node

```
                   wrap
            ----->|--------------------------------
            |  -<--|---------------------------    |
            | |    |                           |   | |
            | v                                |   | v
          -----                              ----
          | A |                              | C | Added
          |   |                              |   | node
          -----                              -----
           ^ |                                ^ |
         o | |                              i | |
         u | |                              n | |
         t | |                              n | |
         e | |                              e --- wrap
         r | |                              r ^ |
           | v                                | v
          -----                              -----
          | D |                              | B |
          |   |                              |   |
          -----                              -----
           | |                                | |
           | |                                | |
           | ---------------------------------   |
           ------------------------------------
```
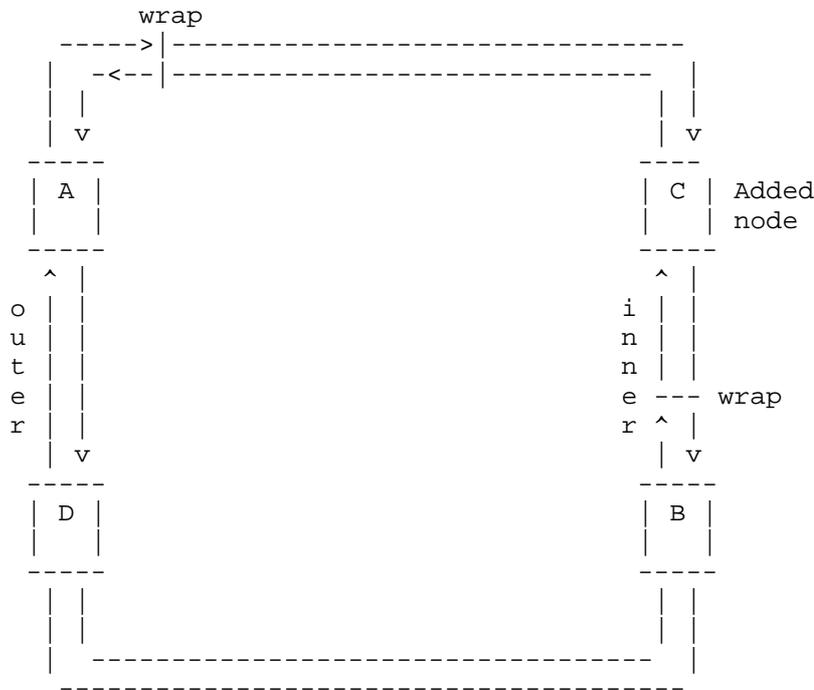
   Sample scenario in a ring where initially nodes A and B are
   connected. Subsequently fibers between the nodes A and B are
   disconnected and a new node C is inserted.

   Bidirectional Fiber Cut

       1. Ring in Idle, all nodes transmit (Tx) {IDLE, SELF, I, S} on
          both rings (in both directions)

       2. Fibers are removed between nodes A and B

       3. B detects SF on the outer ring, transitions to Wrapped state
          (performs a wrap), Tx towards A on the inner ring/short path:
          {SF, B, W, S} and on the outer ring/long path: Tx {SF, B, W, L}

   4. A detects SF on the inner ring, transitions to Wrapped state
      (performs a wrap), Tx towards B on the inner ring/short path:
      {SF, A, W, S} and on the outer ring/long path: Tx {SF, A, W, L}

   5. As the nodes on the long path between A and B receive a SF
      request, they enter a pass-through mode (in each direction),
      stop sourcing the Idle messages and start passing the messages
      between A an B

   6. Steady state is reached

   Node C is Powered Up and Fibers Between Nodes A and C are Reconnected

   This scenario is identical to the returning a Failed Node to Service
   scenario.

   Second Span Put Into Service

   Nodes C and B are connected. The scenario is identical to
   Bidirectional Fiber Cut fault clearing scenario.

9. SRP over SONET/SDH

   Although SRP is media independent it is worth noting how SRP is used
   with a layer 1 media type. SRP over SONET/SDH is the first media type
   perceived for SRP applications.

   Flag delimiting on SONET/SDH uses the octet stuffing method defined
   for POS. The flags (0x7E) are packet delimiters required for
   SONET/SDH links but may not be necessary for SRP on other media
   types. End of a packet is delineated by the flag which could also be
   the same as the next packet's starting flag. If the flag (0x7E) or
   an escape character (0x7D) are present anywhere inside the packet,
   they have to be escaped by the escape character when used over
   SONET/SDH media.

   SONET/SDH framing plus POS packet delimiting allows SRP to be used
   directly over fiber or through an optical network (including WDM
   equipment).

   SRP may also connect to a SONET/SDH ring network via a tributary
   connection to a SONET/SDH ADM (Add Drop Multiplexor). The two SRP
   rings may be mapped into two STS-Nc connections. SONET/SDH networks
   typically provide fully redundant connections, so SRP mapped into two
   STS-Nc connections will have two levels of protection. The SONET/SDH
   network provides layer 1 protection, and SRP provides layer 2
   protection. In this case it is recommended to hold off the SRP Signal
   Fail IPS triggers (which correspond to failures which can be

protected by SONET/SDH) for about 100 msec in order to allow the
SONET/SDH network to protect. Only if a failure persists for over 100
msec (indicating SONET/SDH protection failure) should the IPS
protection take place.

Since multiple protection levels over the same physical
infrastructure are not very desirable, an alternate way of connecting
SRP over a SONET/SDH network is configuring SONET/SDH without
protection. Since the connection is unprotected at layer 1, SRP would
be the sole protection mechanism.

Hybrid SRP rings may also be built where some parts of the ring
traverse over a SONET/SDH network while other parts do not.

Connections to a SONET/SDH network would have to be synchronized to
network timing by some means. This can be accomplished by locking
the transmit connection to the frequency of the receive connection
(called loop timing) or via an external synchronization technique.

Connections made via dark fiber or over a WDM optical network should
utilize internal timing as clock synchronization is not necessary in
this case.

10. Pass-thru mode

An optional mode of operation is pass-thru mode. In pass-thru mode,
a node transparently forwards data. The node does not source
packets, and does not modify any of the packets that it forwards.
Data should continue to be sorted into high and low priority transit
buffers with high priority transit buffers always emptied first. The
node does not source any control packets (e.g. topology discovery or
IPS) and basically looks like a signal regenerator with delay (caused
by packets that happened to be in the transit buffer when the
transition to pass-thru mode occurred).

A node can enter pass-thru mode because of an operator command or due
to a error condition such as a software crash.

11. References

   [1] ANSI X3T9 FDDI Specification

   [2] IEEE 802.5 Token Ring Specification

   [3] Bellcore GR-1230, Issue 4, Dec. 1998, "SONET Bidirectional
        Line-Switched Ring Equipment Generic Criteria".

   [4] ANSI T1.105.01-1998 "Synchronous Optical Network (SONET)
        Automatic Protection Switching"

   [5] Malis, A. and W. Simpson, "PPP over SONET/SDH", RFC 2615, June
        1999.

   [6] Simpson, W., "PPP in HDLC-like Framing", STD 51, RFC 1662, July
        1994.

12. Security Considerations

   As in any shared media, packets that traverse a node are available to
   that node if that node is misconfigured or maliciously configured.
   Additionally, it is possible for a node to not only inspect packets
   meant for another node but to also prevent the intended node from
   receiving the packets due to the destination stripping scheme used to
   obtain spatial reuse. Topology discovery should be used to detect
   duplicate MAC addresses.

13. IPR Notice

   The IETF has been notified of intellectual property rights claimed in
   regard to some or all of the specification contained in this
   document. For more information consult the online list of claimed
   rights.

14. Acknowledgments

   The authors would like to acknowledge Hon Wah Chin who came up with
   the original version of the SRP-fa. Besides the authors, the
   original conceivers of SRP include Hon Wah Chin, Graeme Fraser, Tony
   Bates, Bruce Wilford, Feisal Daruwalla, and Robert Broberg.

15. Authors' Addresses

   Comments should be sent to the authors at the following addresses:

   David Tsiang
   Cisco Systems
   170 W. Tasman Drive
   San Jose, CA 95134

   Phone: (408) 526-8216
   EMail: tsiang@cisco.com

   George Suwala
   Cisco Systems
   170 W. Tasman Drive
   San Jose, CA 95134

   Phone: (408) 525-8674
   EMail: gsuwala@cisco.com

16. Full Copyright Statement

Acknowledgement