INTERNATIONAL TELECOMMUNICATION UNION

TELECOMMUNICATION STANDARDIZATION SECTOR

STUDY PERIOD 2005-2008

STUDY GROUP 15

TD 511 (PLEN/15)

English only

Original: English

Question(s): 9/15 Geneva, 11-22 February 2008

TEMPORARY DOCUMENT

Source: Editors Y.1720

Title: Amendment 1 to Recommendation Y.1720 (Protection switching for MPLS networks) (for consent)

Abstract:

This document provides an amendment to Recommendation Y.1720 "Protection switching for MPLS networks".

The amendment consists of removing the text of Appendix II and Appendix III and a reference to appendix II in clause 7.1.1.4.

Contact:	Yonggang Cheng	Tel: +86-755-28972219
	Huawei Technologies Co., Ltd.	Fax: +86-755-28972935
	P.R.China	Email: yg_cheng@huawei.com
Contact:	Huub van Helvoort	Tel: +31-36-5315076
	Huawei Technologies Co., Ltd.	Fax: +31-84-7122975
	P.R.China	Email: hhelvoort@huawei.com

Amendment to ITU-T Recommendation Y.1720

Protection switching for MPLS networks

In clause 7.1.1.4 the last sentence shall be removed:

Reference model

Figure 9 illustrates a realization of the packet 1+1 protection scheme using sequence numbers as identifiers. After passing through the classifier, each packet that needs to be forwarded on the mated LSPs is assigned a distinct sequence number at the ingress node. This packet with the distinct identification is then duplicated and forwarded onto the two disjoint LSPs. On the egress node, a counter is used to keep track of the expected sequence number of the next packet. ~~The details of an example implementation are described in the appendix.~~

Appendix II shall be removed:

Appendix II

Packet 1+1 example realization

The packet 1+1 scheme can be implemented by using a sequence as an identifier. The sequence number can be carried as the first four bytes inside the shim header of the LSP providing packet 1+1. Since the ingress and egress nodes must be aware of each LSP participating in the packet 1+1, the egress node will recognize that there is a sequence number inside the label. It will use the sequence number for selection purpose and then remove it before forwarding the accepted packet further. Note that packet 1+1 can be provided at any level of the hierarchy of a nested LSP. Figure II.1 illustrates the sequence number position behind the 4-bytes MPLS encapsulation header.

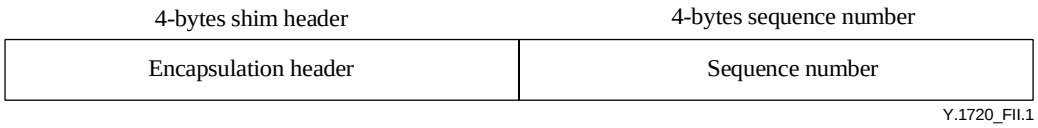

Figure II.1/Y.1720 – An illustration for sequence number transport

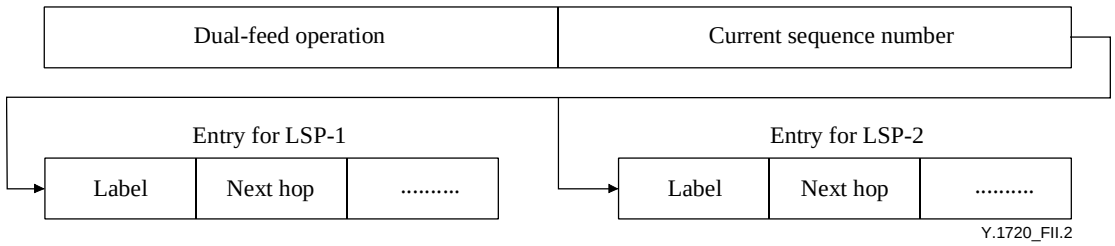

Figure II.2/Y.1720 – Enhanced NHLFE functionality to support dual-feed

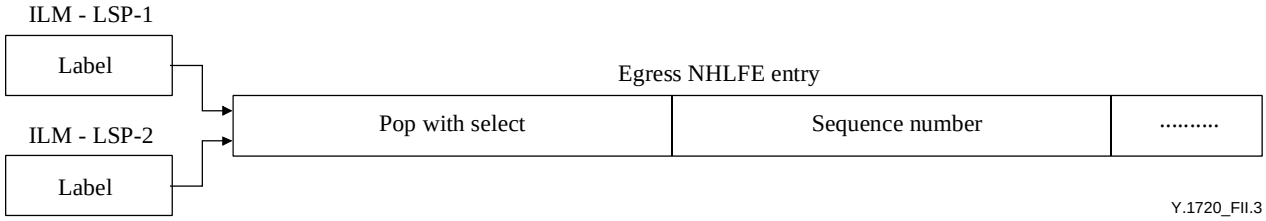

Figure II.3/Y.1720 – Enhanced NHLFE functionality to support selection

Dual-feed and select capabilities can be implemented at the MPLS shim layer by enhancing the Next-Hop-Label-Forward-Entry (NHLFE) entries. At the ingress node, to provide the dual-feed functionality, the NHLFE needs to support two instead of one outgoing LSP. This is easily achieved by using two next-hop/label entries instead of one, each corresponding to one of the mated diverse LSPs. Figure II.2 illustrates this case. Given this, when the client layer packet is forwarded to the NHLFE supporting dual-feed, it first duplicates the packet and then forwards it to the next hops with appropriate labels according to its two next-hop/label entries. In the middle of the network, each copy of the packet traverses the LSP in the standard way, as any other packet would traverse an LSP; thus transparent to the LSRs. On the egress node, the Incoming Label Map (ILM) needs to map the labels of the two diverse LSPs to a single NHLFE entry that enables the receive side to select one of possibly two received copies. Figure II.3 illustrates this case.

II.1 Dual-feed and select mechanism

Two components required for any dual-feed and select mechanism are:

1) the ability to dual-feed at one end; and
2) the ability to select appropriately from the dual-fed signal at the other end. Generally, realization of dual-feed is straightforward whereas, realization of select requires careful and often non-trivial treatment. At the source, packets can be dual-fed by copying on to two packet streams. At the destinations, each packet may be received twice at different times (or once only, or never), once from each of the two LSPs. In order to select each packet once, and once only, the destination must be able to identify the duplicate packets and to then select one, and to handle all possible variations. This selection process at the packet level is non-trivial as the duplicate packets may not arrive at the same time (due to propagation delay and buffering) and also these packets may get lost (due to transmission errors and buffer overflows).

The example algorithm below shows a method that addresses all these issues.

Algorithm

Variables:

N /* number of bits to be used for sequence number */

rec_seq_no /* the sequence number of the received packet */

select_counter /* N bits counter at the receiver that keeps track of the sequence number of next expected packet */

window_sz /* size of the window; must be less than 2^N */

Initialization:

Rec_seq_no = 0;

```
select_counter = 0;
```

Algorithm:

Sender

```
insert rec_seq_no to the inner "label" of the packet;
transmit one copy of the packet on each mated LSPs;
rec_seq_no ++;
```

Selector

```
If(rec_seq_no is outside the sliding window defined by
  [select_counter, select_counter+window_sz])
   reject the packet;
else      /* the rec_seq_no is in the window */
{
          accept the packet;
          select_counter = rec_seq_no +1;
}
```

II.2 Analysis of the packet 1+1 scheme

The ingress node inserts the sequence number. The packet is then duplicated and transported over diverse LSPs. Due to the diversity of the LSPs, there will be a leading LSP and a trailing LSP. The leading LSP will deliver the packets to the egress node faster than the trailing LSP. Therefore, under non-failure conditions, the egress node will select the packets from the leading LSP. The packets received on the trailing LSP will be duplicate packets and will, therefore, be discarded.

The decision whether to accept or discard a received packet is based on the received packet's sequence number and a counter + sliding window at the egress node. The counter indicates the sequence number of the next packet it is expecting. The counter, plus sliding window, provides a window of acceptable sequence numbers. The sliding window is needed to properly accept and reject packets. If the received packet falls in the window, it is considered legitimate and can be accepted: otherwise, it is rejected. The size of the window should be larger than the maximum number of consecutive packets a working (an alive) LSP can lose.

The sliding window is used to solve the problem of losing packets on the leading LSP when the leading LSP's sequence number is very close to the wrap-around point. Figure II.4 illustrates a leading LSP (LSP-1) that delivers a packet with a sequence number 29. The packet is accepted and the counter is incremented to 30. If we assume that 2 consecutive packets are lost (i.e., packets with sequence numbers 30 and 31), the next received packet, on LSP-1 will be 0. Without a sliding window, the egress node will reject the packet since $0 < 30$. By implementing a sliding window that is larger than the maximum number of consecutive packets, a working (an alive) LSP can lose: this problem can be solved. For example, let's say that the maximum number of consecutive packets that a working LSP can lose is 5, then a sliding window of 6 can be defined. Taking the same example as before, however, now using the sliding window, the egress node will accept packets in the range of {30, 31, 0, 1, 2, 3}. Therefore, even if 5 packets are lost (i.e., the maximum number of consecutive packets that can be lost on a working LSP) the next packet received will have a sequence number 3 and the packet will be accepted.

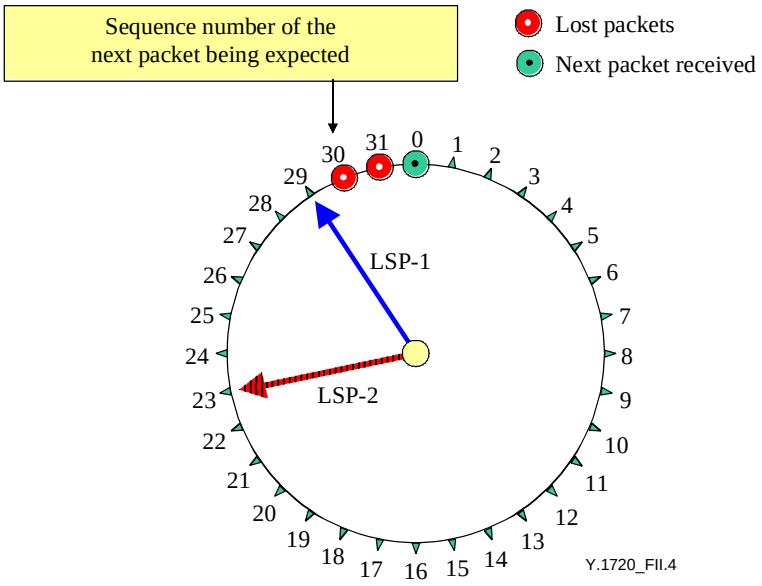

Figure II.4/Y.1720 – Packet loss in conjunction with wrap around

Note that this idea of a sliding window only works if the falling behind LSP cannot fall back in the sliding window range. If a packet with a sequence number in the range of the sliding window is received from the falling behind LSP, then it will be mistakenly accepted. A falling behind LSP can only receive a packet with a sequence number in the range of the sliding window if it falls back by more than (2^N – size of sliding window). Therefore, the number of bits *N* used for the sequence number must support the following equation:

$$2^N > \text{SlidingWindow} + \text{DelayWindow}$$

where:

SlidingWindow > maximum number of consecutive packets that can be lost on a LSP

and

DelayWindow = maximum number of packets the trailing LSP can fall behind the leading LSP

Note that the 4-byte field provides a sequence of more than 4 billion numbers which is large enough to accommodate worst-case consecutive packet losses and delay differentials.

One reasonable way of engineering the size of the sliding and delay windows is to make the size of the sliding window equal to the size of the delay window. (Note that it is assumed that the size of the delay window is generally larger than the size of the sliding window.) This guarantees selection of packets from the leading LSP in all scenarios after a failed LSP gets repaired. This point is further elaborated in the following clause which discusses various failure scenarios.

II.2.1 Operation of select mechanism under various failure scenarios

One way to view the operation of the select mechanism is to picture a clock with 2^N intervals. Figure II.5 illustrates an example where N = 4 (i.e., 4-bit sequence number) and, therefore, the sequence number ranges from 0 through 15. In this example, the sliding window is set equal to the delay window, which is 5.

Figure II.5 shows the leading LSP ahead of the trailing LSP by 3 sequence numbers. The leading LSP delivers a packet with a sequence number = 1 and the counter is now set to 2.

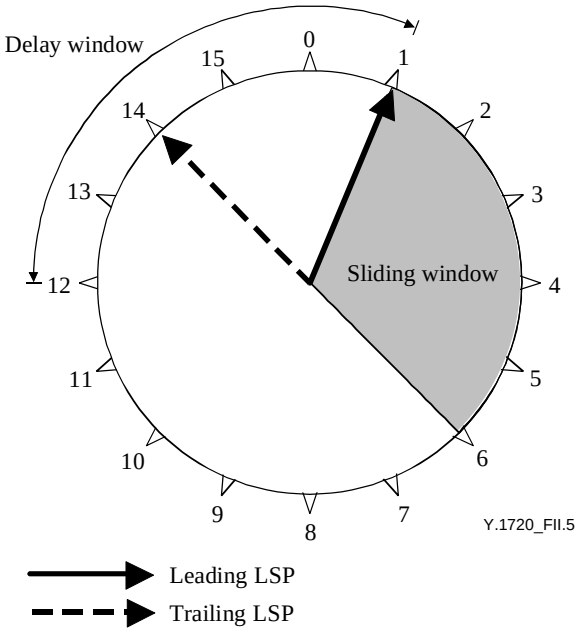

Figure II.5/Y.1720 – Sliding and delay windows concept

Figure II.6 shows that prior to receiving a packet with a sequence number equal to 2 on the leading LSP, the leading LSP fails. Until the packet with a sequence number equal to 2 is delivered from the trailing LSP, the egress node will not select any packets and the counter will remain equal to 2.

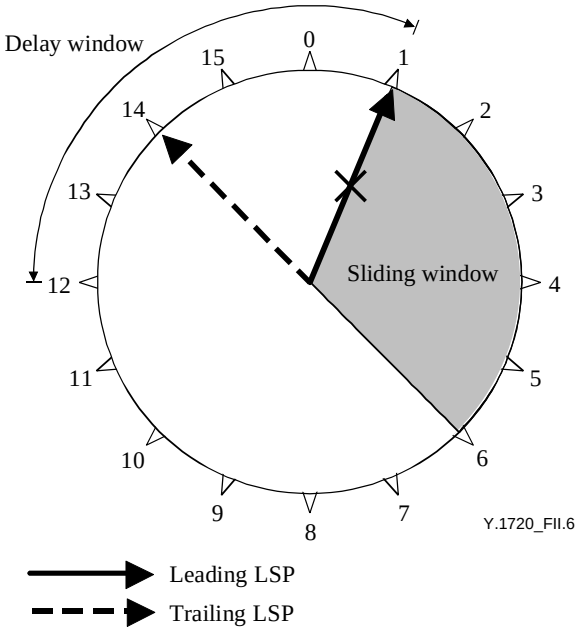

Figure II.6/Y.1720 – Leading LSP failure scenario

Figure II.7 illustrates that when the packet with a sequence number equal to 2 is received on the trailing LSP, the egress node increments the counter to 3 and the sliding window shifts so that a packet with a sequence number in the range of 3 through 7 can be accepted.

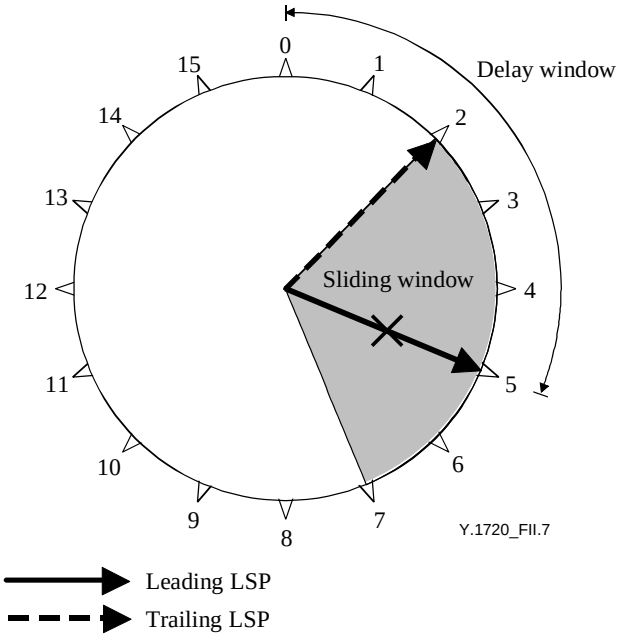

Figure II.7/Y.1720 – Traffic recovery after the leading LSP failure

Figure II.8 illustrates that prior to receiving a packet with a sequence number equal to 3 from the trailing LSP, the leading LSP is repaired and a packet with a sequence number equal to 6 is received from the leading LSP. Since 6 is within the sliding window range, the packet is accepted. Note that it is important that, so long as the leading LSP is working, packets are received from the leading LSP. Therefore, to ensure that when the leading LSP is repaired that it delivers a packet with a sequence number value that is within the sliding window range, the sliding window should be equal to or greater than the delay window which is the case for this example.

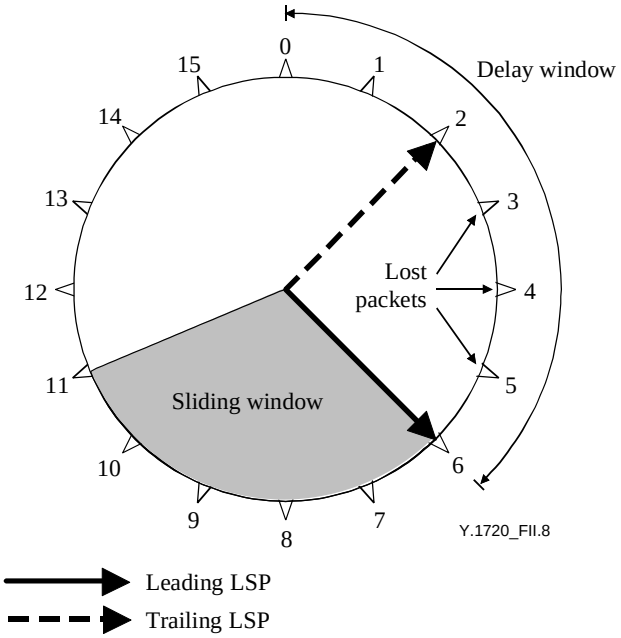

Figure II.8/Y.1720 – Leading LSP repair scenario

Figures II.9, II.10, and II.11 illustrate a problem if the sliding window is set smaller than the delay window. In this case, it is possible that when the leading LSP is repaired, it delivers packets with

sequence numbers that fall outside the sliding window and, therefore, the egress node continues to accept packets from the trailing LSP. If, at a later time, the trailing LSP fails, there is a potential to lose many packets (worst case would be 2^N – *size_of_sliding_window*, where N is the number of bits used for the sequence number).

Figure II.9 shows an example where the sliding window is set to 3 while the delay window can be up to 6. In this example, the trailing LSP trails the leading LSP by 4 sequence numbers. Since the leading LSP has failed, the packets are selected from the trailing LSP.

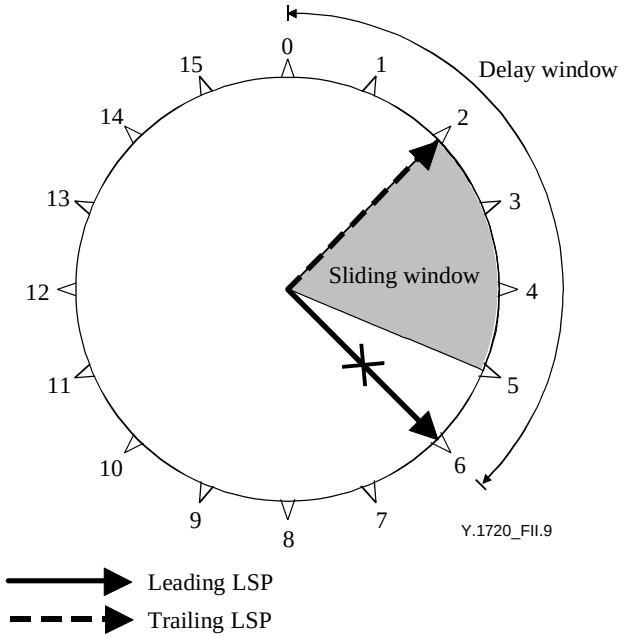

Figure II.9/Y.1720 – Scenario when sliding window < delay window

Figure II.10 illustrates that at the time when the leading LSP is repaired, it delivers a packet with a sequence number equal to 7 which is outside the sliding window and, therefore, rejected. The packets continue to be selected from the trailing LSP.

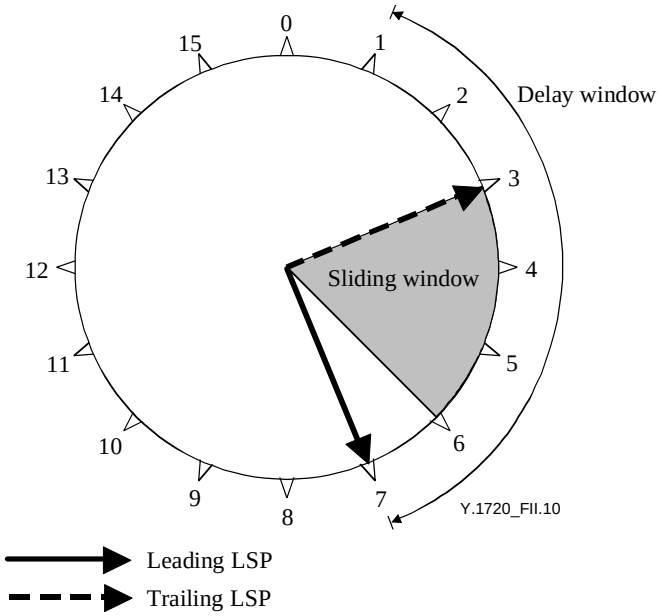

Figure II.10/Y.1720 – LSP repair: sliding window < delay window

Figure II.11 illustrates a failure to the trailing LSP. Since the leading LSP delivers packets outside the sliding window and, therefore, those packets are rejected, the egress node will not start accepting packets until the leading LSP comes all the way around and starts to deliver packets with a sequence number that falls within the sliding window. This can result in a significant loss of packets. Therefore to prevent such an occurrence, it is recommended that this type of selector algorithm set the sliding window equal to the delay window.

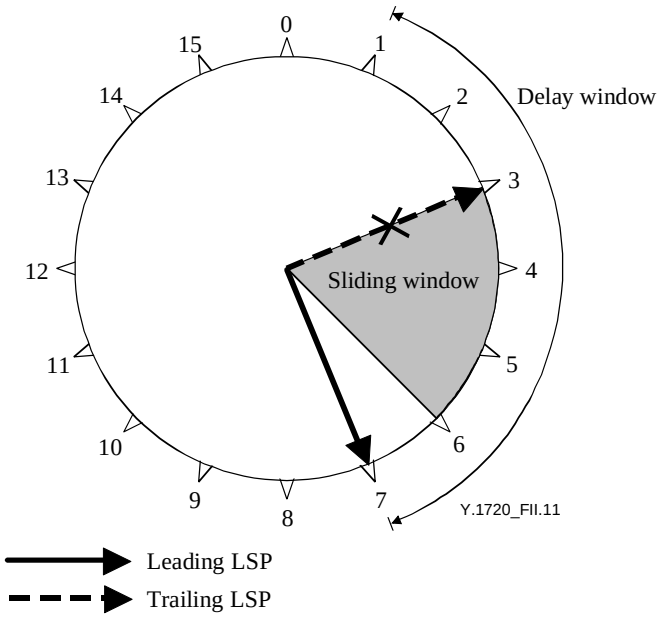

Figure II.11/Y.1720 – Possible problem: sliding window < delay window

II.2.2 Additional remarks

a) The scheme requires intelligence at the edge nodes only. Further, the scheme does not require any explicit fault detection or notification. This is implied by the packet selection

scheme at the egress, which is carried out based on the sequence number and the locally maintained counters.

b) Dual feed requires duplication of packets at ingress. This introduces some additional minimum processing at the ingress. Selection requires comparisons of the sequence number carried in the packet with the counter value maintained at the receiver leading to a packet accept or reject condition. For hardware or software implementation, the processing cost is minimum. Another performance impact is the bandwidth cost due to the sequence number carried in the packets. This introduces some additional packet overhead depending on the length of the sequence number. With a 32-bit sequence number using the whole 4 bytes label, the bandwidth overhead is merely 4% for short 100 bytes packets.

c) The loss performance of the proposed service can be seen as follows. As the selection mechanism at the egress node takes packets from either LSP, the service in fact may compensate, although not required, the packet losses in the network. In the best case, this could result in zero loss, although each LSP may experience losses. On the other hand, in the worst case, the net packet loss would be the sum of the losses of both LSPs. In other words, the loss performance of the service is no worse and of the same order of magnitude as of the worst performing LSP, and sometimes could be much better.

d) The delay performance of the proposed service can be seen as follows. Since the algorithm always selects, without buffering, the first eligible arriving packet of the pair, the delay performance is always better than either of the LSPs.

e) The size of the window should be sized to be larger than the maximum number of consecutive packets a working LSP can lose. As a result, it is assured that the sequence number of the next packet from the same LSP will always fall within the window and will be accepted.

f) The size of the window should be sized such that the delay differential of the packet pairs traversing the mated LSPs, if not lost, is never more than (2^N – size of the window) packets. As a result, it is assured that an old packet will not be mistaken for a new one, thus causing mis-delivery.

g) In case of a single failure in the network, other than the ingress or egress nodes, only one of the mated diverse LSPs will be affected. The surviving LSP will continue delivering the packets. If the surviving LSP is the leading LSP, i.e., the last received and selected packet was from this LSP, then the select function at the egress node will continue to accept packets from it whereas, if the surviving LSP is the trailing LSP, then the select function rejects packets until it sees a packet whose sequence number falls within the sliding window. Upon successful repair of the failed LSP, if so desired, it may be brought back into service. In this "reverted restoration mode", the simplest approach would be to have the first dual-fed packet get the usual next sequence number, next to the one assigned to the last packet fed only on the surviving LSP alone. Various enhancements can be made to manage the service loss performance during this operation, if desired.

h) In case both LSPs have failed, additional mechanisms need to be defined to maintain the service and the LSP associated states to insure robust operations.

Appendix III shall be removed:

Appendix III

Bibliography

IETF, RFC 3469 (2003), Framework for MPLS-based Recovery, Category: Informational.

_____