P802.17/D0.1, January 2002

Proposed Draft Standard for Resilient Packet Ring Access Method & Physical Layer Specifications

(Protection and Topology)

Draft 0.1

Comments on this proposal can be directed to the contributing editors:

Italo Busi
Alcatel
Via Trento, 30
20059 Vimercate (MI)
Italy
Phone: +39 039 686 7054
FAX: +39 039 686 3590
Email: Italo.Busi@alcatel.it

Anoop Ghanwani
Lantern Communications
Email: anoop@lanterncom.com

Jeanne De Jaegher
Alcatel
Email: Jeanne.De_Jaegher@alcatel.be

Pietro Grandi
Alcatel
Via Trento, 30
20059 Vimercate (MI)
Italy
Phone: +39 039 686 4930
FAX: +39 039 686 3590
Email: PietroVittorio.Grandi@netit.alcatel.it

Vittorio Mascolo
Alcatel
Via Trento, 30
20059 Vimercate (MI)
Italy
Phone: +39 039 686 3196
FAX: +39 039 686 3590
Email: Vittorio.Mascolo@alcatel.it

NOTE — The editors recognize that the some text of this document has been taken from the draft Alladin. The purpose of this draft is to update the proposal Alladin.

Table of contents

Abbreviations

MAC	Medium Access Control
PHY	Physical Interface

References

[B1] IEEE 802.3 – 2000 Edition

Carrier sense multiple access with collision detection (CSMA/CD) MAC and physical layer specification.

13 Topology Discovery

13.1 Scope

This section describes the RPR Topology Discovery Protocol, which implements a reliable and accurate means for all RPR stations on a ring to discover the initial topology of the stations on the ring and any changes to that topology. The protocol is intended to scale up to hundreds of stations, to cause minimal overhead for ring traffic, and to cause minimal impact on software and ASICs.

The services and features provided are:

a) Determine/validate connectivity and ordering of stations on the ring

b) Ensure that all stations on the ring have a uniform and current image of the topology

c) Tolerant of message loss

d) Operate without any master station on the ring

e) Operate independently of and in the absence of any management systems

f) Usable with all supported topologies: ring, linear (broken ring), and “star” (single station)

g) Support dynamic addition and removal of stations to/from the ring

h) Detect mis-cabling between stations

i) Provide means of sharing additional information between stations

j) Cause minimal overhead

The RPR Topology Discovery Protocol is used to discover the physical link configuration between stations. The RPR Topology Discovery Protocol also provides a mean to auto-negotiate the support of some options (e.g. wrapping or steering protection) by all the stations on the ring.

It is not within the scope of the RPR Topology Discovery Protocol to determine the dynamic link status information, i.e. which ringlet links are up or down, ring segment failures, etc. The ring topology changes only because of a node insertion or removal. Link failures do not cause any topology change. The discovered topology is used by other protocols such as the RPR Protection Protocol and the RPR Fairness Algorithm.

13.2 Algorithm Overview

The RPR Topology Discovery Protocol provides each station on the ring with knowledge of the number and arrangement of other stations on the ring. This collection of information is referred to as the topology image. Each station maintains its own local copy of the topology image for the entire ring. Initially, the station’s topology image contains information only about itself.

Ring topology discovery is initiated only as needed. Local topology validation eliminates the need for acknowledgements or periodic broadcasts. No station acts as a master for the topology image or for the protocol. All ringlet segments that can be discovered are included. A fully connected ring is not needed for the protocol.

In addition to station identifiers and physical connectivity relationships, the topology discovery protocol is also used to propagate additional station information, both that which is used for other parts of this standard, and optionally information beyond the standard.

The messages sent as part of the RPR Topology Discovery Protocol are indicated in the RPR frame header as control frames.

13.2.1 At initialization

At station initialization, the local topology image is initialized to contain only the local station and no links, and the version of image is initialized to 0. The station starts the topology algorithm by broadcasting a Topology_Status message on all ringlets and sending Neighbor_Hello messages to its neighbors. Then it continually listens for Topology_Status messages broadcast on the ring, listens for Neighbor_Hello messages from its neighbors, broadcasts a Topology_Status messages whenever there is a local topology change or validation failure, and sends Neighbor_Hello messages to each of the neighbors periodically.

13.2.2 Topology Discovery Messages

The periodically sent Neighbor_Hello messages allow stations to learn the status of their neighbor links, and to announce their presence to their neighbors. The Neighbor_Hello message contains a summary of the local topology image that is used to validate that neighbors share the same topology image. When a station receives a Neighbor_Hello message from a neighbor, it checks the image information received against its own image. If the 2 images are different, a topology exchange is initiated.

At any point that a station receives a change in status from a neighbor station, or detects that it and a neighbor station are out of synchronization with each other, it initiates a topology exchange. Topology exchange is done by sending a Topology_Status broadcast message to all stations on the ring. The Topology_Status message contains all the information about the local station, including its links to its neighbors. The combination of Topology_Status messages whenever there is any change in status, and periodic validation checks via Neighbor_Hello messages allows every node to both learn the full ring topology, and to assure that it has the current, correct topology.

It can be easily determined when an image is complete and consistent by examining the image contents. When the contents of the local topology image show station information for each station described in the link information of another station, then the image is complete. When all stations show for every ringlet that all stations on each ringlet have neighbors only in one direction on each ringlet, then the topology image is consistent.

A canonical form for the topology image allows all the stations to eventually arrive at the same image for the topology and to easily compare images. A (rolling) generation counter allows detection of changes to the topology image.

13.2.3 Station_Image_Version

Each station maintains a version number for its local topology image, called the Station_Image_Version. The Station_Image_Version is initialized to 0 to indicate no valid image (other than itself). It is incremented by the local station whenever a change in local status occurs, and sent out in the resulting Topology_Status message. Change in local status is defined by change in neighbor ID. Each station maintains an independent Station_Image_Version.

13.2.4 Ring_Image_Version

Each station retains the Station_Image_Version sent in the latest Topology_Status message from each other station on the ring. Each time a new Station_Image_Version is received, the receiving station calculates a checksum of all the Station_Image_Version values in its local topology image (including itself). This checksum is called the Ring_Image_Version. The Ring_Image_Version should be the same in all stations that are mutually reachable. (A set of stations are mutually reachable if a message from one station can reach any other station.) A mismatch between mutually reachable neighbors indicates a need to update the topology image. However, mismatches are ignored during the time immediately following a change in topology to avoid excessive messages while the topology stabilizes. This period of time is the set by the configurable Topology_Stabilization_Timer.

13.2.5 Determination And Validation Of Ringlet ID

Each station determines which interface is associated with which ringlet and assigns the corresponding ringlet ID either through fixed mapping between hardware locations or through configuration. Each Neighbor_Hello topology control message is sent separately on each ringlet, identifying the ringlet on which it is being sent. Any Neighbor_Hello message received on a ringlet different from the ringlet on which it is identified as being sent shall trigger a mis-configuration alarm.

13.3 Topology Discovery Process

13.3.1 Topology Discovery Process Description

1) Neighbor station change

   — Trigger

   No Neighbor_Hello messages in 3 Neighbor_Hello Periods (indicating a lost neighbor), or a Neighbor_Hello message from a new neighbor.

   — Action

   a) Increment the local Station_Image_Version.
   b) Broadcast a Topology_Status message.
   c) Replace the station information in the local topology image.
   d) Update the local Ring_Image_Version.
   e) Start Topology_Stabilization_Timer.

2) Non-neighbor station change

   — Trigger

   A higher Station_Image_Version is received in a Topology_Status message.

   — Action

   a) Replace the remote station information in the local topology image.
   b) Update the remote Station_Image_Version.
   c) Update the local Ring_Image_Version.
   d) Start Topology_Stabilization_Timer.

3) Neighbor validation failure or new station

   — Trigger

   A Ring_Image_Version in a Neighbor_Hello doesn't match the local one AND the sending station and the local station are mutually reachable AND the local Topology_Stabilization_Timer is not running, or the local Ring_Image_Version is 0 (a new station).

   — Action

   a) Set the local and all the remote Station_Image_Versions = 0.
   b) Send a Topology_Status message.
   c) Send a Neighbor_Hello message.
   d) Increment Neighbor_Validation_Failure counter.
   e) Start Topology_Stabilization_Timer.

4) Non-neighbor validation failure

   — Trigger

   A Topology_Status message with Station_Image_Version = 0.

— Action

a) Update the remote Station_Image_Version to 0.
b) Broadcast a Topology_Status message.
c) Update the local Ring_Image_Version.
d) Start Topology_Stabilization_Timer.

13.3.2 Topology Discovery Process State Diagram

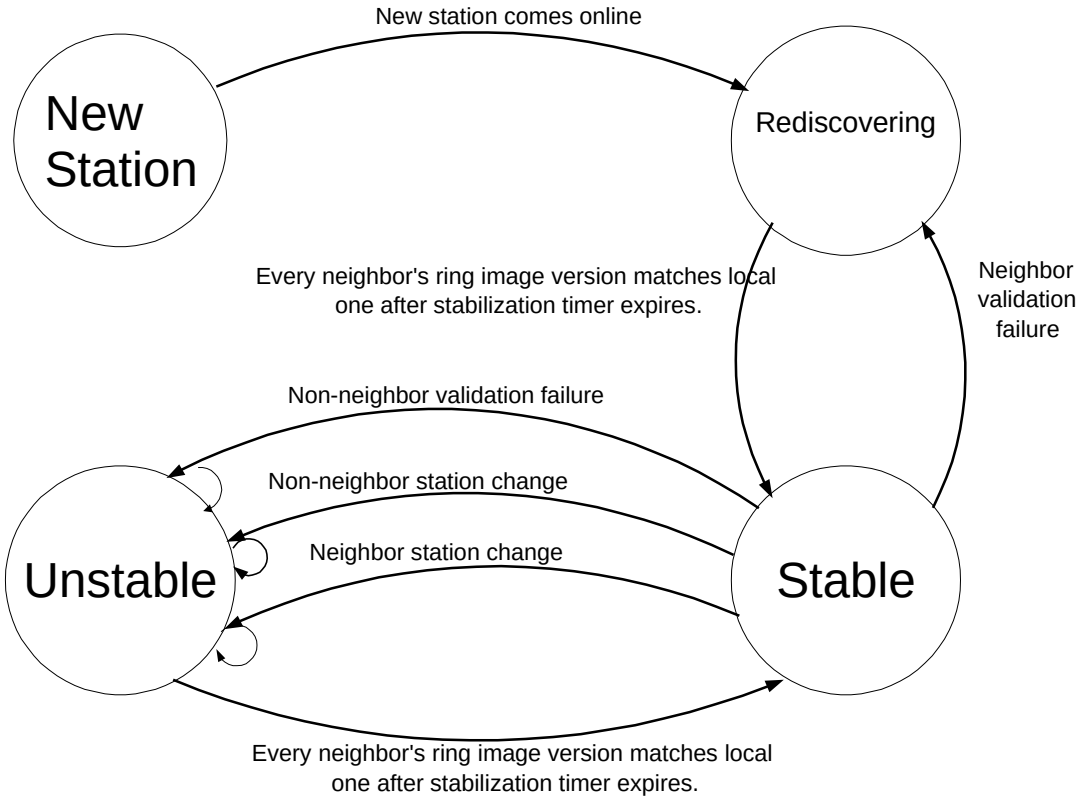

Figure 13.1 – Topology Discovery State Diagram

13.4 Topology Discovery Messages

13.4.1 Topology_Status

Topology_Status messages report changes in neighbor identity. They are sent as MAC Control messages, as broadcast frames on all ringlets, and with TTL of Max_Ring_Size (255). They are removed by the source station. The source MAC is set to the actual MAC of the sending station.

The information field of the message is as follows:

Bytes 1..4	station_image_version: unsigned 32-bit integer
Byte 5	cw_ringlets: unsigned 8-bit integer
Byte 6	ccw_ringlets: unsigned 8-bit integer

Byte 7	cw_station_ringlet_id[0]: unsigned 8-bit integer
Bytes 8..13	cw_station_address[0]: IEEE-48 MAC address
	Above 3 fields repeated as necessary for cw_ringlets
Byte 14	ccw_station_ringlet_id[0]: unsigned 8-bit integer
Bytes 15..20	ccw_station_address[0]:
	Above 3 fields repeated as necessary for ccw_ringlets
Byte 21	Autonegotiation
Byte 22	Weight
Byte 23	Organization ID
Byte 24	private_length
Bytes 25…	private_data

Table 13.1 – Topology_Status message format

Parameters (see table above for codings)

NOTE —Byte displacement values shown are for 1 clockwise ringlet and 1 counter clockwise ringlet.

Topology_Status opcode: The MAC Control opcode value for a Topology_Status message.

station_image_version: The station_image_version parameter shall be set to the current value of the Station_Image_Version of the sending station. If there is no current local topology image, Station_Image_Version shall be set to 0.

cw_ringlets: The cw_ringlets parameter indicates the number of ringlets in the clockwise direction.

ccw_ringlets: The ccw_ringlets parameter indicates the number of ringlets in the counterclockwise direction.

cw_station_ringlet_id: The cw_station_ringlet_id parameter carries the ID of the ringlet on which the corresponding station is connected.

cw_station_address: The cw_station_address parameters carry the MAC addresses of the stations clockwise to the sending station. If the station's MAC address is unknown, it shall be all 0's.

cw_station_in_link_status: The cw_station_in_link_status parameters carry the current status of the in coming links from the stations clockwise to the sending station. Valid values are DISCONNECTED, CONNECTED, and UNKNOWN.

ccw_station_ringlet_id: The ccw_station_ringlet_id parameter carries the ID of the ringlet on which the corresponding station is connected.

ccw_station_address: The ccw_station_address parameters carry the MAC addresses of the stations counterclockwise to the sending station. If the station's MAC address is unknown, it shall be all 0's.

ccw_station_in_link_status: The ccw_station_in_link_status parameters carry the current status of the in coming links from the stations counterclockwise to the sending station. Valid values are DISCONNECTED, CONNECTED, and UNKNOWN.

Autonegotiation: The Autonegotiation field is used to indicate the characteristic of a node for auto-negotiation purposes. The field is encoded as follows.

Bit 0	Wrapping protection capable (1)
Bit 1	Steering protection capable (1)
Bits 2…7	Reserved

Table 13.2 – Autonegotiation field format

Weight: The weight field represents the weight assigned to the station for the weighted fairness algorithm.

Organization ID: The organization ID field represents the organization (i.e. a particular vendor) responsible to define the structure of the private data.

private_length: The private_length parameter carries the length, in bytes, of the private_data parameter.

private_data: The private_data parameter carries any private data desired beyond the data required by the protocol.

13.4.1.2 When generated

The Topology_Status message is broadcast on the initial start of the RPR topology discovery, upon any change in the neighboring nodes, and upon any validation failure.

13.4.1.3 Effect of receipt

The receipt of this message from another station causes the MAC sublayer to update its current local topology image.

13.4.2 Neighbor_Hello

The Neighbor_Hello message reports the presence, identity, and topology version of a source station to a neighbor station. It is resent every time the Neighbor_Hello_Timer expires.

Neighbor_Hello messages are sent as MAC Control messages, as broadcast frames, and with TTL set to 1. This guarantees that they will be received by any neighbor and removed from the ring immediately. The source MAC address is set to the actual MAC address of the sending station.

The information field of the message is as follows:

Byte 1	ringlet_id: unsigned 8-bit integer
Bytes 2..5	ring_image_version: unsigned 32-bit integer

Table 13.3 – Neighbor_Hello message format

Parameters:

Neighbor_Hello opcode: The MAC Control opcode value for a Neighbor_Hello message.

ringlet_id: The ringlet_id parameter carries the ID of the ringlet on which the request is sent.

ring_image_version: The ring_image_version parameter carries the current value of the Ring_Image_Version checksum calculated by the sending station. If there is no current local topology image, Ring_Image_Version shall be set to 0.

13.4.2.2 When generated

The Neighbor_Hello message is generated on the initial start of the topology discovery and upon expiration of the Neighbor_Hello_Timer.

13.4.2.3 Effect of receipt

The receipt of this message causes the MAC sublayer to validate its current local topology image and to broadcast a Topology_Status message if it discovers that the image has changed.

14 Ring Protection

14.1 Scope

This RPR Protection Protocol implements a reliable, accurate, efficient, and quick means for all RPR stations on a ring to discover a broken segment in a ringlet of an RPR and to re-route RPR frames away from the break, until the break is healed. The protocol is intended to scale from 1 to 100's of stations, to cause minimal overhead for ring traffic, and to cause minimal impact on software and ASICs. The services and features provided are:

a) Quick dissemination of loss of connectivity information on the ring

b) Tolerance of message loss

c) Operation without any master station on the ring

d) Operation independent of and in the absence of any management systems

e) Operation with dynamic addition and removal of stations to/from the ring

f) Independency from the protection type (wrapping/steering)

g) Minimal overhead

14.2 Algorithm Overview

The end result of the RPR Protection Protocol is that each station on the ring knows of a ring segment failure and protects ring traffic away from the failure within 50 ms of the failure.

The protocol is always initiated by the network elements that detect a fault. The fault notification (detected via physical layer trigger or keep alive control) is propagated hop by hop on each ringlet, within the keep alive notification.

The keep alive notification message is a single hop (TTL = 1) broadcast message that is regenerated periodically.

When faults arise or disappear, the fault notifications must be generated and propagated immediately in an asynchronous way.

The broadcast messages sent as part of the RPR Protection Protocol are indicated in the RPR frame header as control packets with the highest priority class of service.

14.2.1 Keep_Alive message

The Keep_Alive message is used to detect MAC layer failure. They are exchanged periodically (1 msec.) between neighbors to check continuity at the data link level. They are sent as MAC Control messages, as broadcast frames, and with TTL of 1. This guarantees that they will be received by any neighbor and removed from the ring immediately.

When no faults are detected all the keep alive fields are set to “0”.

Byte 0..5	MAC address = 0
Byte 1	upstream_link_status: = 0
Byte 2	Flags

Table 14.1 – Keep_Alive Message format

MAC address field: the MAC address of the node that detects the fault. It is set to 0 in case of no fault on the ring.

Flags field: contains 8 flag-bit: one bit can be used to specify the direction of the fault. It is set to 0 on the same ringlet on which the fault is detected and it is set to one on the opposite ringlet. It is set to 0 in case of no fault on the ring.

upstream_link_status field: It contains the kind of fault. Values are (according to SDH/SONET G.841 notation):

0x00	No Fault
0x05	WTR
0x06	Manual Switch
0x08	Signal Degrade
0x0B	Signal Fail
0x0D	Forced Switch

Table 14.2 – Upstream link status field values

14.2.1.2 General rules

a) The RPR Ring consists of two counter-rotating ringlets

b) Protection messages are continuously sent

c) Fault is always considered bi-directional.

d) In case of multiple failures the ring must form islands.

e) Network elements detect the faults monitoring the receive direction on each ringlet

Notifications are emitted as keep alive messages on all the ringlets and must be propagated along the ringlets.

Notifications are emitted and propagated in an asinchronous way. (that is without waiting for the expiration of the keep-alive sending timer)

14.2.1.3 Basic Principle and priority

Local event notifications have precedence (in term of processing) over incoming notifications whenever:

a) locally detected either SF (signal fail) or FS (force switch)

b) locally detected priority is higher or equal than the incoming one (remote signalling).

Local events withdrawn because of priority, do not originate notifications (e.g. if a node locally detects a SD and on the ring there was already a SF, the SD is not notified).

Commands withdrawn because of priority are not re-issued.

In case of two locally detected events that can co-exist each event is notified in only one direction (the opposite direction to the fault).

The only kind of faults that can co-exist in the ring are:

a) Signal Fail and Force Switch

b) Fault with the same priority

In case of multiple fails with the same priority, if they are SD or manual switch, the failure is notified but the protection is not performed.

Notification integration is performed after forwarding the notifications.

Notification integration should be at least long as the ring roundtrip time (no false switches in case of contemporary faults with different priorities).

Because of the fault priorities WTR works only when re-entering from a single fault.

14.2.2 Initialization

At station initialization, the local topology image is learned through the RPR Topology Discovery Protocol. No protection state is enabled until the local topology image is generated. Once the topology is learned, any link failure detected locally or received from a remote station initiates the protection state.

14.3 Protection Triggers

14.3.1 Physical Layer Triggers

Link failure and restoration are initially detected by the physical layer, according to its capability, and then indicated to the MAC layer.

14.3.1.1 SONET/SDH Triggers

Link failure on a SONET/SDH link is triggered by Signal Failure (SF) or Signal Degradation (SD) defects provided by the SONET/SDH physical layer. Clearing of these defects clears the link failure indication.

14.3.1.2 Ethernet Triggers

Link failure on an Ethernet link is triggered by Signal Failure (SF) or Signal Degradation (SD) defects provided by the Ethernet physical layer, or inferred from statistics from the Ethernet physical layer. For Ethernet physical layers that do not provide SD indications, these shall be calculated from threshold values of the coding violations. Clearing of these defects clears the link failure indication.

14.3.2 Manual Triggers

For the purpose of maintenance, an RPR link can be taken off-line. The operator shall be able to signal the off-line state via a forced or a manual switch command indicating the link down. The operator command will be translated into a protection message that simulates a link failure.

14.3.3 Keep_Alive Triggers

The Keep_Alive message provides a means of detecting a link failure or station failure. Any time a configurable number of Keep_Alive messages in a row are missed (as indicated by receiving 0 Keep_Alive messages within the configured number of expirations of the Keep_Alive_Timer), the link upon which they were to have been received is indicated to be down. This kind of failure is considered as Signal Fail. The default value of the number of messages that must be missed is 3. 2 Keep_Alives must be received in succession (within 3 expirations of the Keep_Alive_Timer) to clear the link failure indication.

14.4 Protection Hierarchy

The protection switching events hierarchy is used to handle multiple, concurrent events. The hierarchy is shown below from the highest priority to the lowest one:

FS, Force Switch - operator originated (highest priority)
SF, Signal Fail (e.g LOS, LOF, EXBER, LOK (Loss Of Keep_Alive)) - automatically originated
SD, Signal Degrade - automatically originated
MS, Manual Switch - operator originated
WTR, Wait Time To Restore - automatically originated
NR, No Request present (lowest priority)

14.4.1 Hold-Off Time

For RPR deployments over existing a physical medium providing its own protection (such as SONET/SDH Automatic Protection Switching or optical protection), the underlying protection for the RPR bandwidth cannot always be turned off. In such cases, the underlying protection already in place will be used instead of the RPR protection.

In order not to use RPR protection when an underlying protection should be used as first resort, a configurable Hold-Off Time (HOT) is provided for the RPR protection. By setting the HOT to a value greater than zero, the underlying protection switching is given a chance to work. RPR protection will take place only if the link is not declared up before the HOT expires.

When the RPR protection is the “lowest layer” protection scheme, the HOT is simply set to zero.

The Hold-Off Time is configurable on a per span base.

14.4.2 Wait To Restore

To avoid frequent changes in reported link status over unstable links, a configurable Wait To Restore (WTR) timer is provided. By setting the WTR to a value greater than zero, the failed link is required to show no failure for the given period of time before being declared working.

The WTR is configurable per node.

14.5 Multiple Link Failure

Multiple link failures between two neighbor stations are treated in the same manner as a single link failure. Each link failure triggers a separate protection for the ringlet that experienced the failure. For dual-ringlet rings, each of the two ringlets protects the other ringlet. Each flow, conversation, aggregate, or whatever means is used to segregate and route traffic (henceforth called the routable unit) is protected by a single ringlet. Protection does not need to be one to one if 100% protection is not needed.

14.6 Algorithm Details

14.6.1 Normal Operation

During normal operation, each station listens for keep alive messages and examines them for any change in link status. In case of change any network element must first of all propagate the fault to its neighbor on the same ringlet where the fault notification is received.

A network element that detects a fault must immediately originate a fault notification on both ringlets unless two local faults that can co-exist are detected by the same network element. In this last case the behavior is described in paragraph 3.1.2.

14.6.2 Unicast Protection

During protection, for each routable unit, if a sourcing station can reach the intended destination through the normal route (either through the assigned ringlet or locally), then it uses the normal route. Otherwise, if the routable unit is protected and it can reach the intended destination through an alternate route (through the protected ringlet), then it uses the protection route.

14.6.3 Multicast Protection

If a sourcing station can reach all the intended destinations through the normal route (either through the assigned ringlet or locally), then it uses the normal route. Otherwise, if the routable unit is protected and it can reach all intended destinations through the alternate route (through the protected ringlet), then it uses the protection route. Otherwise, it uses both routes (both directions). In all of these cases, the TTL would be set to the distance to the failed link on each ringlet. A station can make a local optimization to set the TTL to the shorter of distance of the farthest station for the routable unit or the distance to the failed link, for each ringlet.

14.6.4 Recovery From Protection

During protection operation, each station listens for keep alive messages and examines them for any change in link status. When a fault recovers, the network element must now evaluate whether to enter WTR state or not depending from other fault present on the network. In case WTR state is entered the link is declared recovered at WTR expiration; otherwise the link is declared recovered immediately. At this point the system returns to its state prior to the protection event and the Keep_Alive messages are sent with all 0s.