MPLS Working Group
Internet-Draft
Updates: 6378 (if approved)
Intended status: Standards Track
Expires: May 31, 2014

J. Ryoo, Ed.
ETRI
E. Gray, Ed.
Ericsson
H. van Helvoort
Huawei Technologies
A. D'Alessandro
Telecom Italia
T. Cheung
ETRI
E. Osborne
Cisco Systems, Inc.
November 27, 2013

MPLS Transport Profile (MPLS-TP) Linear Protection in Support of ITU-T's Requirements

draft-ietf-mpls-tp-psc-itu-00.txt

Abstract

This document introduces alternate ways to perform certain operations defined in RFC6378, "MPLS Transport Profile (MPLS-TP) Linear Protection", and also defines additional behaviors. This set of modified and additional behaviors together with the protocol defined in RFC6378 meets the ITU-T's protection switching requirements.

This document introduces capabilities and modes. A capability is an individual behavior. The capabilities of a node are advertised using the method given in this document. A mode is a particular combination of capabilities. Two modes are defined in this document: Protection State Coordination (PSC) mode and Automatic Protection Switching (APS) mode.

This document describes the behavior of the PSC protocol including priority logic and state machine when all the capabilities associated with the APS mode are enabled.

This document updates RFC6378 in that the capability advertisement method defined here is an addition to that document.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute

working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on May 31, 2014.

Table of Contents

1. Introduction

This document introduces alternate ways to perform certain operations defined in [RFC6378], "MPLS Transport Profile (MPLS-TP) Linear Protection", and also defines additional behaviors. This set of modified and additional behaviors together with the protocol defined in [RFC6378] meets the ITU-T's protection switching requirements.

Alternative behaviors are defined for the following capabilities:

1. Priority modification,

2. non-revertive behavior modification,

and the following capabilities have been added to define additional behaviors:

3. support of Manual Switch to Working (MS-W) command,

4. support of protection against Signal Degrade (SD), and

5. support of Exercise command.

Priority modification includes priority swapping between Signal Fail on the Protection path (SF-P) and Forced Switch (FS), and raising the priority level of Clear SF.

Non-revertive behavior is modified to align with the behavior defined in [RFC4427] as well as to meet the ITU-T's protection switching requirements.

Support of Manual Switch to Working (MS-W) command to revert traffic to the working path in non-revertive operation is covered in this document.

Support of protection switching protocol against Signal Degrade (SD) is covered in this document. The specifics for the method of identifying SD is out of the scope of this document similarly to SF for [RFC6378].

Support of Exercise command to test if the Protection State Coordination (PSC) communication is operating correctly is also covered in this document. More specifically, the Exercise tests and validates the linear protection mechanism and PSC protocol including the aliveness of the Local Request logic, the PSC state machine and the PSC message generation and reception, and the integrity of the protection path, without triggering the actual traffic switching.

This document introduces capabilities and modes. A capability is an individual behavior, The capabilities of a node are advertised using the method given in this document. A mode is a particular combination of capabilities. Two modes are defined in this document: PSC mode and Automatic Protection Switching (APS) mode.

This document describes the behavior of the PSC protocol including priority logic and state machine when all the capabilities associated with the APS mode are enabled.

This document updates [RFC6378] in that the capability advertisement method defined here is an addition to that document. For an existing implementation of [RFC6378], it is recommended to be updated with the bug-fixes in [I-D.ietf-mpls-psc-updates] and the capability adevertisement in this document.

2. Conventions Used in This Document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

3. Acronyms

This document uses the following acronyms:

APS	Automatic Protection Switching
EXER	Exercise
FS	Forced Switch
LO	Lockout of protection
MS	Manual Switch
MS-P	Manual Switch to Protection
MS-W	Manual Switch to Working
MPLS-TP	MPLS Transport Profile
NR	No Request
OC	Operator Clear
PSC	Protection State Coordination
RR	Reverse Request
SD	Signal Degrade
SD-P	Signal Degrade on the Protection path
SD-W	Signal Degrade on the Working path
SF	Signal Fail
SFc	Clear Signal Fail
SF-P	Signal Fail on the Protection path
SF-W	Signal Fail on the Working path
WTR	Wait to Restore

4. Capability 1: Priority Modification

In this document, the priorities of Forced Switch (FS) and Signal Fail on the Protection path (SF-P) are swapped and the priority of Clear SF (SFc) is raised. In addition to the priority modification, this document introduces the use of a Freeze command in Appendix C. The reasons for these changes are explained in the following sub-sections from technical and network operational aspects.

4.1. Motivations for swapping priorities of FS and SF-P

Defining the priority of FS higher than that of Signal Fail on the Protection path (SF-P) can result in a situation where the protected traffic is taken out-of-service. Setting the priority of any input that is supposed to be signalled to the other end to be higher than that of SF-P can result in unpredictable protection switching state, when the protection path has failed and consequently the PSC communication stopped. An example of the out-of-service scenarios is shown in Appendix A

According to Section 2.4 of [RFC5654] it MUST be possible to operate an MPLS-TP network without using a control plane. This means that external switch commands, e.g., FS, can be transferred to the far end only by using the PSC communication channel and should not rely on the presence of a control plane.

As the priority of SF-P has been higher than FS in optical transport networks and Ethernet transport networks, for network operators it is important that the MPLS-TP protection switching preserves the network operation behavior to which network operators have become accustomed. Typically, the FS command is issued before network maintenance jobs, (e.g., replacing optical cables or other network components). When an operator pulls out a cable on the protection path by mistake, the traffic should be protected and the operator expects this behavior based on his/her experience on the traditional transport network operations.

4.2. Motivation for raising the priority of Clear SF

The priority level of SFc defined in [RFC6378] can cause traffic disruption when a node that has experienced local signal fails on both working and protection paths is recovering from these failures.

An example of sequence diagram showing the problem with the priority level of SFc as defined in [RFC6378] is shown in Appendix B.

4.3. Motivation for introducing Freeze command

With the priority swapping between FS and SF-P, the traffic is always moved back to the working path when SF-P occurs in Protecting Administrative state. In the case that network operators need an option to control their networks so that the traffic can remain on the protection path even when the PSC communication channel is broken, the Freeze command, which is a local command (i.e., not signalled to the other end) can be used. The use of the Freeze command is described in Appendix C.

4.4. Updates to the PSC RFC

The list of local requests in order of priority should be modified as follows:

(from higher to lower)

- Clear Signal Fail/Degrade
- Signal Fail on the Protection path

- Forced Switch
- Signal Fail on the Working path

The change of the PSC control logic including state machine due to this priority modification is incorporated in the PSC control logic description when all the capabilities are enabled in Section 10 and Section 11.

5. Capability 2: Modification of Non-revertive Operation

Non-revertive mode of protection switching is defined in [RFC4427]. In this mode, the traffic does not return to the working path when switch-over requests are terminated.

However, PSC protocol defined in [RFC6378] supports this operation only when recovering from a defect condition, but does not operate as non-revertive when an operator's switch-over command such as Forced Switch or Manual Switch is cleared. To be aligned with legacy transport network behavior and [RFC4427], a node should go into the Do-not-Revert (DNR) state not only when a failure condition on a working path is cleared but also when an operator command requesting switch-over is cleared.

The change of the PSC control logic including state machine due to the modification of non-revertive operation is incorporated into the PSC control logic description when all the capabilities are enabled in Section 10 and Section 11.

6. Capability 3: Support of Manual Switch to Working Command

6.1. Motivation for adding Manual Switch to Working

Changing the non-revertive operation introduces necessity of a new operator command to revert traffic to the working path when in Do-not-Revert (DNR) state. When the traffic is on the protection path in DNR state, a Manual Switch to Working (MS-W) command is issued to switch the normal traffic back to the working path. According to Section 4.3.3.6 (Do-not-Revert State) in [RFC6378], "to revert back to Normal state, the administrator SHALL issue a Lockout of protection (LO) command followed by a Clear command." However, using LO command introduces the potential risk of an unprotected situation while the Lockout of protection is in effect.

Manual Switch-over for recovery LSP/span command, defined in [RFC4427] and also defined in [RFC5654], Requirement 83, as one of the mandatory external commands, should be used for this purpose, but is not included in [RFC6378]. Note that the "Manual Switch-over for recovery LSP/span" command is the same as MS-W command.

6.2. Terms modified to support MS-W

The term "Manual Switch" and its acronym "MS" used in [RFC6378] are replaced respectively by "Manual Switch to Protection" and "MS-P" by this document to avoid confusion with "Manual Switch to Working" and its acronym "MS-W".

Also, the term "Protecting administrative state" used in [RFC6378] is replaced by "Switching administrative state" by this document to include the case where traffic is switched back to the working path by administrative Manual Switch to Working command.

6.3. Behavior of MS-P and MS-W

The MS-P and MS-W commands SHALL have the same priority. If one of these commands is already issued and accepted, and the other command that is issued afterwards SHALL be ignored. If two LERs are requesting opposite operations simultaneously, i.e. one LER is sending MS-P while the other LER is sending MS-W, the MS-W SHALL be considered to have a higher priority than MS-P, and MS-P SHALL be ignored.

Two commands, MS-P and MS-W are represented by the same Request Field value, but differentiated by the FPath value. When traffic is switched to the protection path, the FPath field SHALL indicate that the working path is being blocked (i.e., FPath set to 1), and the Path field SHALL indicate that user data traffic is being transported on the protection path (i.e., Path set to 1). When traffic is switched to the working path, the FPath field SHALL indicate that the protection path is being blocked (i.e., FPath set to 0), and the Path field SHALL indicate that user data traffic is being transported on the working path (i.e., Path set to 0).

6.4. Equal priority resolution for MS

[RFC6378] defines only one rule for equal priority condition in Section 4.3.2 as "The remote message from the far-end LER is assigned a priority just below the similar local input." In order to support the manual switch behavior described in Section 6.3, additional rules for equal priority resolution are required. Since the support of protection against signal degrades also requires a similar equal priority resolution, the rules are described in Section 7.4.

The change of the PSC control logic including state machine due to the support of MS-W command is incorporated into the PSC control logic description when all the capabilities are enabled in Section 10 and Section 11.

7. Capability 4: Support of protection against Signal Degrade

7.1. Motivation for supporting protection against Signal Degrade

In MPLS-TP survivability framework [RFC6372], fault conditions include both Signal Fail (SF) and Signal Degrade (SD) that can be used to trigger protection switching.

[RFC6378], which defines the Protection State Coordination (PSC) protocol, does not specify how the SF and SD are declared and specifies the protection switching protocol associated with SF only.

The protection switching protocol associated with SD is covered in this document, and the specifics for the method of identifying SD is out of the scope of PSC protocol similarly to how to detect SF and how MS and FS commands are initiated in a management system and signalled to PSC.

7.2. Terms modified to support SD

Clear Signal Fail (SFc) includes the clearance of a degraded condition in addition to the clearance of a failure condition

The second paragraph of Section 4.3.3.2 Unavailable State in [RFC6378] shows the intention of including Signal Degrade on the Protection path (SD-P) in the Unavailable state. Even though the protection path can be partially available under the condition of the Signal Degrade on the Protection path, this document follows the same state grouping as [RFC6378] for SD on the protection path.

The bullet item "Protecting failure state" in Section 3.6. PSC Control States in [RFC6378] includes the degraded condition in Protection Failure state. This document follows the same state grouping as [RFC6378] for Signal Degrade on the Working path (SD-W).

7.3. Behavior of protection against SD

In order to maintain the network operation behavior to which transport network operators have become accustomed, the priorities of SD-P and SD-W are defined to be equal as in other transport networks, such as OTN and Ethernet. Once a switch has been completed due to Signal Degrade on one path, it will not be overridden by Signal Degrade on the other path (first come, first served behavior), to

avoid protection switching that cannot improve signal quality and flapping.

Signal Degrade (SD) indicates that the transmitting end point has identified a degradation of the signal, or integrity of the packet transmission on either the working or protection path. The FPath field SHALL identify the path that is reporting the degrade condition (i.e., if protection path, then FPath is set to 0; if working path, then FPath is set to 1), and the Path field SHALL indicate where the data traffic is being transported (i.e., if working path is selected, then Path is set to 0; if protection path is selected, then Path is set to 1).

The Wait to Restore (WTR) timer is used when the protected domain is configured for revertive behavior and started at the node that recovers from a local degraded condition on the working path.

If the detection of a SD depends on the presence of user data packets, such a condition declared on the working path is cleared following protection switching to the protection path if a selector bridge is used, possibly resulting in flapping. To avoid flapping, the selector bridge should duplicate the user data traffic and feed it to both working and protection paths under SD condition. In revertive mode, when WTR timer expires the packet duplication will be stopped and the user data traffic will be transported on the working path only. In non-revertive mode, when SD is cleared the packet duplication will be stopped and the user data traffic will be transported on the protection path only.

When multiple SDs are detected simultaneously, either as local or remote requests on both working and protection paths, the SD on the standby path (the path from which the selector does not select the user data traffic) is considered as having higher priority than the SD on the active path (the path from which the selector selects the user data traffic). Therefore, no unnecessary protection switching is performed and the user data traffic continues to be selected from the active path.

In the preceding paragraph, "simultaneously" relates to the occurrence of SD on both the active and standby paths at input to the Protection State Control Logic in Figure 1 of [RFC6378] at the same time, or as long as a SD request has not been acknowledged by the remote end in bidirectional protection switching. In other words, when a local node that has transmitted a SD message receives a SD message that indicates a different value of data path (Path) field than the value of the Path field in the transmitted SD message, both the local and the remote SD requests are considered to occur simultaneously.

7.4. Equal priority resolution

In order to support the manual switch behavior described in Section 6.3 and the protection against Signal Degrade described in Section 7.3, the rules to resolve the equal priority requests are required.

For local inputs with same priority, such as MS and SD, first-come, first-served rule is applied. Once a local input is determined as the highest priority local input, then a subsequent equal priority local input requesting a different action, i.e., the same PSC Request Field but different FPath value, to the PSC control logic will not be presented to the PSC control logic as the highest local request. Furthermore, in the case of MS, the subsequent MS local input requesting a different action will be cancelled.

The remote message from the far-end LER is assigned a priority just below the similar local input. For example, a remote Forced Switch would have a priority just below a local Forced Switch but above a local Signal Fail on working input assuming that the priority modification is in place as in Section 4.4

However, if the LER is in a remote state due to a remote message, a subsequent local input having the same priority but requesting different action to the control logic, will be considered as having lower priority than the remote message, and will be ignored. For example, if the LER is in remote Unavailable state due to a remote SD-P, then subsequent local SD-W input will be ignored. Likewise, if the LER is in remote Switching administrative state due to a remote MS-P, then subsequent local MS-W will be ignored and automatically cancelled.

It should be noted that there is a reverse case where one LER receives a local input and the other LER receives, simultaneously, an input with the same priority but requesting different action. In this case, each of the two LERs receives a subsequent remote message having the same priority but requesting different action, while the LER is in a local state due to the local input. In this case, a priority must be set for the inputs with the same priority regardless of its origin (local input or remote message). For example, one LER receives SD-P as a local input and the other LER receives SP-W as a local input, simultaneously. Likewise, one LER receives MS-P as a local input and the other LER receives MS-W as a local input, simultaneously.

When MS-W and MS-P occur simultaneously at both LERs, MS-W SHALL be considered as having higher priority than MS-P at both LERs.

When SD-W and SD-P occur simultaneously at both LERs, In this case, the SD on the standby path (the path from which the selector does not select the user data traffic) is considered as having higher priority than the SD on the active path (the path from which the selector selects the user data traffic) regardless of its origin (local or remote message). Therefore, no unnecessary protection switching is performed and the user data traffic continues to be selected from the active path. Giving the higher priority to the SD on the standby path SHALL also be applied to the Local Request logic when two SDs for different paths happen to be presented to the Local Request logic exactly at the same time.

The change of the PSC control logic including state machine due to the support of protection against SD is incorporated into the PSC control logic description when all the capabilities are enabled in Section 10 and Section 11.

8. Capability 5: Support of Exercise Command

Exercise is a command to test if the PSC communication is operating correctly. More specifically, the Exercise is to test and validate the linear protection mechanism and PSC protocol including the aliveness of the Local Request logic, the PSC state machine and the PSC message generation and reception, and the integrity of the protection path, without triggering the actual traffic switching. It is used while the working path is either carrying the traffic or not. It is lower priority than any "real" switch request. It is only valid in bidirectional switching, since this is the only place where one can get a meaningful test by looking for a response.

This command is documented in R84 of [RFC5654] and it has been identified as a requirement from ITU-T.

A received EXER message indicates that the remote end point is operating under an operator command to validate the protection mechanism and PSC protocol including the aliveness of the Local Request logic, the PSC state machine and the PSC message generation and reception, and the integrity of the protection path, without triggering the actual traffic switching. The valid response to EXER message will be an Reverse Request (RR) with the corresponding FPath and Path numbers. The near end will signal a Reverse Request (RR) only in response to an EXER command from the far end.

When Exercise commands are input at both ends, an EXER, instead of RR, is transmitted from both ends.

The following PSC Requests should be added to PSC Request field to support Exercise:

(TBD2) Exercise - indicates that the transmitting end point is exercising the protection channel and mechanism. FPath and Path are set to the same value of the NR, RR or DNR request that EXER replaces.

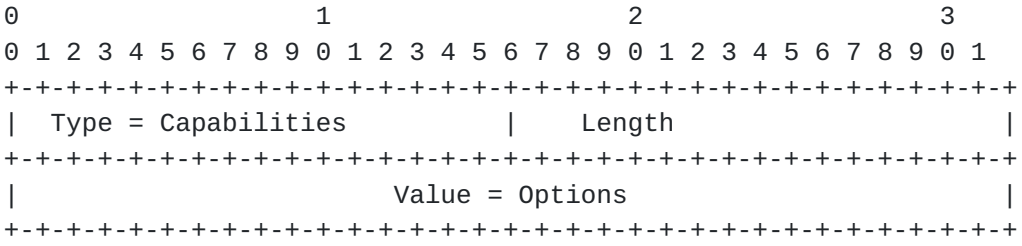
(TBD1) Reverse Request - indicates that the transmitting end point is responding to an EXER command from the far end. FPath and Path are set to the same value of the NR, RR or DNR request that EXER replaces.

The priority of Exercise should be inserted between the priorities of WTR Expires and No Request.

9. Capabilities and Modes

9.1. Capabilities

A Capability is an individual behavior whose use is signalled in a Capabilities TLV, which is placed in Optional TLVs field inside PSC messages shown in Figure 2 of [RFC6378]. The format of the Capabilities TLV is:



The value of the Type field is TBD3 pending IANA allocation.

The value of the Length field is the length of the Options Value, and is in octets.

The Value of the Capabilities TLV can be any length, as long as it is a multiple of 4 octets. The length of the Value field MUST be the minimum required to signal all the required capabilities. Section 4 to Section 8 discuss five capabilities that are signalled using the 5 most significant bits; if a node wishes to signal these five capabilities, it MUST send an Options Value of 4 octets. A node would send an Options Value greater than 4 octets only if it had more than 32 Capabilities to indicate. All unused bits MUST be set to zero.

If the bit assigned for an individual capability is set to 1, it indicates the sending node's intent to use that capability in the protected domain. If a bit is set to 0, the sending node does not

intend to use the indicated capability in the protected domain. Note that it is not possible to distinguish between the intent not to use a capability and a node's complete non-support (i.e. lack of implementation) of a given capability.

This document defines five specific capabilities that are described from Section 4 to Section 8. Each capability is assigned bit as follows:

0x80000000: priority modification

0x40000000: modification of non-revertive behavior

0x20000000: support of Manual Switch to Working (MS-W) command

0x10000000: support of protection against Signal Degrade (SD)

0x08000000: support of Exercise command

9.1.1. Sending the Capabilities TLV

PSC sends messages in response to external events and in periodic retransmission of current status. It may be expensive to send and to parse an Capabilities TLV attached to a packet intended to trigger a protection switch or other real- time behavior. However, if a node does not periodically send its Capabilities TLV, the receiving node cannot discriminate a deliberate omission of the Capabilities TLV for performance reasons from an accidental omission due to an implementation issue. To guard against this, a node MUST include its Capabilities TLV in every PSC message that it sends.

9.1.2. Receiving the Capabilities TLV

A node MUST establish a receive timer for the Capabilities TLV. By default this MUST be 3.5 times the periodic retransmission timer of five seconds - i.e., 17.5 seconds. Both the periodic retransmission time and the timeout SHOULD be configurable by the operator. When a node receives a Capabilities TLV it resets the timer to 17.5 seconds. If the timer expires, the node behaves as in Section 9.1.3.

[Editor's note: In other packet transport protection technologies, Failure of Protocol defect (dFOP) is declared when no protocol message is received on the protection path during at least 3.5 times the periodic message transmission interval (i.e., at least 17.5 seconds) and there is no defect on the protection transport entity. As the "Capabilities TLV" is included in the PSC message, this error of not receiving the Capabilities TLV can be covered by dFOP. To be discussed.]

When a node receives a Capabilities TLV it MUST compare it to its most recent transmitted Capabilities TLV. If the two are equal, the protected domain is said to be running in the mode indicated by that set of capabilities (see Section 9.2). If the sent and received Capabilities TLVs are not equal, this indicates a capabilities mismatch. When this happens, the node MUST alert the operator and MUST behave as in Section 9.1.3.

9.1.3. Handling Capabilities TLV errors

This section covers the two possible errors - a TLV timeout and a TLV mismatch - and the error handling procedures in both cases.

9.1.3.1. Capabilities TLV Timeout

If the Capabilities TLV receive timer expires, a node is said to have timed out. When this happens, the node MUST alert the operator and MUST behave as in Section 9.1.3.3.

9.1.3.2. Capabilities TLV Mismatch

If the sent and received Capabilities TLVs are not equal, this indicates a capabilities mismatch. When this happens, the node MUST alert the operator and MUST behave as in Section 9.1.3.3. A node MAY retain the received TLV for logging, alert or debug purposes.

9.1.3.3. Handling Capabilities TLV error conditions

When a node enters in Capabilities protocol error conditions, the following actions MUST be taken:

1. Indicate the error condition (e.g., either mismatch or timeout) to the operator by the usual alert mechanisms (e.g., syslog).

2. Not make any state transitions based on the contents of any PSC Messages

To expand on point 2 - assume node A is receiving NR(0,0) from its PSC peer node Z and is also receiving a mismatched set of capabilities (e.g., received 0x4, transmitted 0x5). If node Z detects a local SF-W and wants to initiate a protection switch (that is, by sending SF(1,1)), node A MUST NOT react to this input by changing its state. A node MAY increase the severity or urgency of its alarms to the operator, but until the operator resolves the mismatch in the Capabilities TLV the protected domain will likely operate in an inconsistent state.

9.2. Modes

A Mode is a given set of Capabilities. Modes are shorthand; referring to a set of capabilities by their individual values or by the name of their mode does not change the protocol behavior. This document defines two modes - PSC and APS.

9.2.1. PSC Mode

PSC Mode is defined as the lack of any Capabilities - that is, a Capabilities set of 0x0. It is the behavior specified in RFC6378. There are two ways to declare PSC Mode. A node can send a Capabilities TLV of 0x0, or it can send no Capabilities TLV at all. This is further explored in Section 9.3.

9.2.2. APS Mode

APS Mode is defined as the use of all of the five specific capabilities, which are described from Section 4 to Section 8 in this document. APS Mode is indicated with a Value of 0xF8000000.

9.3. Backward compatibility

As defined in Section 9.2.1, PSC Mode is indicated either with a Capabilities TLV of 0x0 or the lack of Capabilities TLV. This is to allow backward compatibility between two nodes - one which can send the Capabilities TLV, and one which cannot.

[RFC6378] does not define how to handle an unrecognized TLV. There may be some implementations that silently discard an unrecognized TLV, and some that take more drastic steps like refusing to allow PSC to operate. Thus, a node which has the ability to send and receive the PSC Mode Capabilities TLV MUST be able to both send the PSC Mode Capabilities TLV and send no Capabilities TLV at all. An implementation MUST be configurable between these two choices.

One question that arises from this dual definition of PSC Mode is, what happens if a node which was sending a non-null Capabilities TLV (e.g., APS Mode) sends PSC packets without any Capabilities TLV? This case is handled as follows:

If a node has never, during the life of a PSC session, received a Capabilities TLV from a neighbour, the lack of a Capabilities TLV is treated as receipt of a PSC Capabilities TLV. This allows for interop between nodes which support the PSC Mode TLV and nodes which do not, and are thus implicitly operating in PSC Mode.

If a node has received a non-null Capabilities TLV (e.g., APS Mode) during the life of a PSC session and then receives a PSC packet with no Capabilities TLV, the receiving node MUST treat the lack of Capabilities TLV as simply a lack of refresh. That is, the receipt of a PSC packet with no Capabilities TLV simply does not reset the receive timer defined in Section 9.1.2.

10. PSC Protocol in APS Mode

This section and Section 11 defines the behavior of PSC protocol when all of the aforementioned capabilities are enabled, i.e., APS mode.

10.1. Request field in PSC protocol message

The values of "Request" field in the PSC protocol message, which is shown in Figure 2 of [RFC6378], are defined as follows:

(14) Lockout of protection

(12) Forced Switch

(10) Signal Fail

(7) Signal Degrade

(5) Manual Switch

(4) Wait-to-Restore

(TBD2) Exercise

(TBD1) Reverse Request

(1) Do-not-Revert

(0) No Request

10.2. Priorities of local inputs and remote requests

Based on the description in Section 3 and Section 4.3.2 in [RFC6378], the priorities of multiple outstanding local inputs are evaluated in Local Request logic unit, where the highest priority local request is determined. This high-priority local request is passed to the PSC Control logic, that will determine the higher priority input (top priority global request) between the highest priority local input and the last received remote message. When a remote message comes to the PSC Control logic, the top priority global request is determined between this remote message and the highest priority local input

which is present. The top priority global request is used to determine the state transition, which is described in Section 11.

The priorities for both local and remote requests are defined as follows from highest to lowest:

- Operator Clear (Local only)
- Lockout of protection (Local and Remote)
- Clear Signal Fail/Degrade (Local only)
- Signal Fail on Protection path (Local and Remote)
- Forced Switch (Local and Remote)
- Signal Fail on Working path (Local and Remote)
- Signal Degrade on either Protection path or Working path (Local and Remote)
- Manual Switch to either Protection path or Working path (Local and Remote)
- WTR Expires (Local only)
- WTR (Remote only)
- Exercise (Local and Remote)
- Reverse Request (Remote only)
- Do-Not-Revert (Remote only)
- No Request (Remote and Local)

The remote request from the far-end LER is assigned a priority just below the same local request. However, for the equal priority requests, such as Signal Degrade on either Working or protection and Manual Switch to either Protection or Working path, the following equal priority resolution rules are defined:

- If two local inputs having same priority but requesting different action come to the Local Request logic, then the input coming first SHALL be considered to have a higher priority than the other coming later (first-come, first-served).

- If the LER receives both a local input and a remote message with the same priority and requesting the same action, i.e., the same PSC Request Field and the same FPath value, then the local input SHALL be considered to have a higher priority than the remote message.

- If the LER receives both a local input and a remote message with the same priority but requesting different actions, i.e., the same PSC Request Field but different FPath value, then the first-come, first-served rule SHALL be applied. If the remote message comes first, then the state SHALL be a remote state and subsequent local input is ignored. However, if the local input comes first, the first-come, first-served rule cannot be applied and must be viewed as simultaneous condition. This is because the subsequent remote message will not be an acknowledge of the local input by the far-end node. In this case, the priority SHALL be determined by rules for each simultaneous condition.

- If the LER receives both MS-P and MS-W requests as both local input and remote message and the LER is in a local Switching administrative state, then the MS-W request SHALL be considered to have a higher priority than the MS-P request.

- If the LER receives both SD-P and SD-W requests as both local input and remote message and the LER is in a local state, then the SD on the standby path (the path from which the selector does not select the user data traffic) SHALL be considered as having higher priority than the SD on the active path (the path from which the selector selects the user data traffic) regardless of its origin (local or remote message). This rule of giving the higher priority to the SD on the standby path SHALL also be applied to the Local Request logic when two SDs for different paths happen to be presented to the Local Request logic exactly at the same time.

11. State Transition Tables in APS Mode

When there is a change in the highest-priority local request or in remote PSC messages, the top priority global request is evaluated and the state transition tables are looked up in PSC control logic. The following rules are applied to the operation related to the state transition table lookup.

- If the top priority global request, which determines the state transition, is the highest priority local input, the local state transition table SHALL be used to decide the next state of the LER. Otherwise, remote messages state transition table SHALL be used.

- If in remote state, the highest local defect condition (SF-P, SF-W, SD-P or SD-W) SHALL always be reflected in the Request Field and Fpath.

- Operator Clear command, Clear SF/SD (SFc) and WTR Expires are not persistent. Once they appear to the local priority logic and complete the operation, they will be disappeared.

- For the LER currently in the local state, if the top priority global request is changed to OC or SFc causing the next state to be Normal, WTR or DNR, then all the local and remote requests should be re-evaluated as if the LER is in the state specified in the footnotes to the state transition tables, before deciding the final state. This re-evaluation is an internal operation confined within the local LER, and PSC messages are generated according to the final state.

- The WTR timer is started only when the LER which has recovered from a local failure/degradation enters the WTR state. An LER which is entering into the WTR state due to a remote WTR message does not start the WTR timer.

The extended states, as they appear in the table, are as follows:

N	Normal state
UA:LO:L	Unavailable state due to local LO command
UA:P:L	Unavailable state due to local SF-P
UA:DP:L	Unavailable state due to local SD-P
UA:LO:R	Unavailable state due to remote LO message
UA:P:R	Unavailable state due to remote SF-P message
UA:DP:L	Unavailable state due to local SD-P
PF:W:L	Protecting failure state due to local SF-W
PF:DW:L	Protecting failure state due to local SD-W
PF:W:R	Protecting failure state due to remote SF-W message
PF:DW:R	Protecting failure state due to remote SD-W message
SA:F:L	Switching administrative state due to local FS command
SA:MW:L	Switching administrative state due to local MS-W command
SA:MP:L	Switching administrative state due to local MS-P command
SA:F:R	Switching administrative state due to remote FS message
SA:MW:R	Switching administrative state due to remote MS-W message
SA:MP:R	Switching administrative state due to remote MS-P message
E::L	Exercise state due to local EXER command
E::R	Exercise state due to remote EXER message
WTR	Wait-to-Restore state
DNR	Do-not-Revert state

Each state corresponds to the transmission of a particular set of Request, FPath and Path bits. The table below lists the message that

is generally sent in each particular state. If the message to be sent in a particular state deviates from the table below, it is noted in the footnotes to the state transition tables.

State	REQ(FP,P)
N	NR(0,0)
UA:LO:L	LO(0,0)
UA:P:L	SF(0,0)
UA:DP:L	SD(0,0)
UA:LO:R	highest local request(local FPath,0)
UA:P:R	highest local request(local FPath,0)
UA:DP:R	highest local request(local FPath,0)
PF:W:L	SF(1,1)
PF:DW:L	SD(1,1)
PF:W:R	highest local request(local FPath,1)
PF:DW:R	highest local request(local FPath,1)
SA:F:L	FS(1,1)
SA:MW:L	MS(0,0)
SA:MP:L	MS(1,1)
SA:F:R	highest local request(local FPath,1)
SA:MW:R	highest local request(local FPath,0)
SA:MP:R	highest local request(local FPath,1)
WTR	WTR(0,1)
DNR	DNR(0,1)
E::L	EXER(0,x), where x is the existing Path value when Exercise command is issued.
E::R	RR(0,x), where x is the existing Path value when RR message is generated.

11.1. State transition by local inputs

	OC	LO	SFc	SF-P	FS	SF-W
N	i	UA:LO:L	i	UA:P:L	SA:F:L	PF:W:L
UA:LO:L	(1)	i	i	i	i	i
UA:P:L	i	UA:LO:L	(1)	i	i	i
UA:DP:L	i	UA:LO:L	(1)	UA:P:L	SA:F:L	PF:W:L
UA:LO:R	i	UA:LO:L	i	UA:P:L	i	PF:W:L
UA:P:R	i	UA:LO:L	i	UA:P:L	PF:W:L	PF:W:L
UA:DP:R	i	UA:LO:L	i	UA:P:L	SA:F:L	PF:W:L
PF:W:L	i	UA:LO:L	(2)	UA:P:L	SA:F:L	i
PF:DW:L	i	UA:LO:L	(2)	UA:P:L	SA:F:L	PF:W:L
PF:W:R	i	UA:LO:L	i	UA:P:L	SA:F:L	PF:W:L
PF:DW:R	i	UA:LO:L	i	UA:P:L	SA:F:L	PF:W:L
SA:F:L	(3)	UA:LO:L	i	UA:P:L	i	i
SA:MW:L	(1)	UA:LO:L	i	UA:P:L	SA:F:L	PF:W:L
SA:MP:L	(3)	UA:LO:L	i	UA:P:L	SA:F:L	PF:W:L

SA:F:R	i	UA:LO:L	i	UA:P:L	SA:F:L	PF:W:L
SA:MW:R	i	UA:LO:L	i	UA:P:L	SA:F:L	PF:W:L
SA:MP:R	i	UA:LO:L	i	UA:P:L	SA:F:L	PF:W:L
WTR	i	UA:LO:L	i	UA:P:L	SA:F:L	PF:W:L
DNR	i	UA:LO:L	i	UA:P:L	SA:F:L	PF:W:L
E::L	(4)	UA:LO:L	i	UA:P:L	SA:F:L	PF:W:L
E::R	i	UA:LO:L	i	UA:P:L	SA:F:L	PF:W:L

	SD-P	SD-W	MS-W	MS-P	WTRExp	EXER
N	UA:DP:L	PF:DW:L	SA:MW:L	SA:MP:L	i	E::L
UA:LO:L	i	i	i	i	i	i
UA:P:L	i	i	i	i	i	i
UA:DP:L	i	i	i	i	i	i
UA:LO:R	UA:DP:L	PF:DW:L	i	i	i	i
UA:P:R	UA:DP:L	PF:DW:L	i	i	i	i
UA:DP:R	UA:DP:L	PF:DW:L	i	i	i	i
PF:W:L	i	i	i	i	i	i
PF:DW:L	i	i	i	i	i	i
PF:W:R	UA:DP:L	PF:DW:L	i	i	i	i
PF:DW:R	UA:DP:L	PF:DW:L	i	i	i	i
SA:F:L	i	i	i	i	i	i
SA:MW:L	UA:DP:L	PF:DW:L	i	i	i	i
SA:MP:L	UA:DP:L	PF:DW:L	i	i	i	i
SA:F:R	UA:DP:L	PF:DW:L	i	i	i	i
SA:MW:R	UA:DP:L	PF:DW:L	SA:MW:L	i	i	i
SA:MP:R	UA:DP:L	PF:DW:L	i	SA:MP:L	i	i
WTR	UA:DP:L	PF:DW:L	SA:MW:L	SA:MP:L	(6)	i
DNR	UA:DP:L	PF:DW:L	SA:MW:L	SA:MP:L	i	E::L
E::L	UA:DP:L	PF:DW:L	SA:MW:L	SA:MP:L	i	i
E::R	UA:DP:L	PF:DW:L	SA:MW:L	SA:MP:L	i	E::L

11.2. State transition by remote messages

	LO	SF-P	FS	SF-W	SD-P	SD-W
N	UA:LO:R	UA:P:R	SA:F:R	PF:W:R	UA:DP:R	PF:DW:R
UA:LO:L	i	i	i	i	i	i
UA:P:L	UA:LO:R	i	i	i	i	i
UA:DP:L	UA:LO:R	UA:P:R	SA:F:R	PF:W:R	i	(10)
UA:LO:R	i	UA:P:R	SA:F:R	PF:W:R	UA:DP:R	PF:DW:R
UA:P:R	UA:LO:R	i	SA:F:R	PF:W:R	UA:DP:R	PF:DW:R
UA:DP:R	UA:LO:R	UA:P:R	SA:F:R	PF:W:R	i	PF:DW:R
PF:W:L	UA:LO:R	UA:P:R	SA:F:R	i	i	i
PF:DW:L	UA:LO:R	UA:P:R	SA:F:R	PF:W:R	(11)	i
PF:W:R	UA:LO:R	UA:P:R	SA:F:R	i	UA:DP:R	PF:DW:R
PF:DW:R	UA:LO:R	UA:P:R	SA:F:R	PF:W:R	UA:DP:R	PF:DW:R
SA:F:L	UA:LO:R	UA:P:R	i	i	i	i

SA:MW:L	UA:LO:R	UA:P:R	SA:F:R	PF:W:R	UA:DP:R	PF:DW:R
SA:MP:L	UA:LO:R	UA:P:R	SA:F:R	PF:W:R	UA:DP:R	PF:DW:R
SA:F:R	UA:LO:R	UA:P:R	i	PF:W:R	UA:DP:R	PF:DW:R
SA:MW:R	UA:LO:R	UA:P:R	SA:F:R	PF:W:R	UA:DP:R	PF:DW:R
SA:MP:R	UA:LO:R	UA:P:R	SA:F:R	PF:W:R	UA:DP:R	PF:DW:R
WTR	UA:LO:R	UA:P:R	SA:F:R	PF:W:R	UA:DP:R	PF:DW:R
DNR	UA:LO:R	UA:P:R	SA:F:R	PF:W:R	UA:DP:R	PF:DW:R
E::L	UA:LO:R	UA:P:R	SA:F:R	PF:W:R	UA:DP:R	PF:DW:R
E::R	UA:LO:R	UA:P:R	SA:F:R	PF:W:R	UA:DP:R	PF:DW:R

	MS-W	MS-P	WTR	EXER	RR	DNR	NR
N	SA:MW:R	SA:MP:R	i	E::R	i	i	i
UA:LO:L	i	i	i	i	i	i	i
UA:P:L	i	i	i	i	i	i	i
UA:DP:L	i	i	i	i	i	i	i
UA:LO:R	SA:MW:R	SA:MP:R	i	E::R	i	i	N
UA:P:R	SA:MW:R	SA:MP:R	i	E::R	i	i	N
UA:DP:R	SA:MW:R	SA:MP:R	i	E::R	i	i	N
PF:W:L	i	i	i	i	i	i	i
PF:DW:L	i	i	i	i	i	i	i
PF:W:R	SA:MW:R	SA:MP:R	(7)	E::R	i	(8)	(5)
PF:DW:R	SA:MW:R	SA:MP:R	(7)	E::R	i	(8)	(5)
SA:F:L	i	i	i	i	i	i	i
SA:MW:L	i	i	i	i	i	i	i
SA:MP:L	i	i	i	i	i	i	i
SA:F:R	SA:MW:R	SA:MP:R	i	E::R	i	DNR	N
SA:MW:R	i	SA:MP:R	i	E::R	i	i	N
SA:MP:R	SA:MW:R	i	i	E::R	i	DNR	N
WTR	SA:MW:R	SA:MP:R	i	i	i	i	(9)
DNR	SA:MW:R	SA:MP:R	i	E::R	i	i	i
E::L	SA:MW:R	SA:MP:R	i	i	i	i	i
E::R	SA:MW:R	SA:MP:R	i	i	i	DNR	N

NOTES:

(1) Re-evaluate to determine final state as if the LER is in the Normal state.

(2) In the case that both local input and the last received remote message are no request after the occurrence of SFc, the LER enters into the WTR state when the domain is configured for revertive behavior, or the LER enters into the DNR state when the domain is configured for non-revertive behavior. In all the other cases, re-evaluate to determine the final state as if the LER is in the Normal state.

(3) Re-evaluate to determine final state as if the LER is in the Normal state when the domain is configured for revertive behavior, or as if the LER is in the DNR state when the domain is configured for non-revertive behavior,

(4) If Path value is 0, re-evaluate to determine final state as if the LER is in the Normal state. If Path value is 1, re-evaluate to determine final state as if the LER is in the DNR state

(5) If the received NR message has Path=1, transition to WTR if domain configured for revertive behavior, else transition to DNR.

(6) Remain in WTR, send NR(0,1).

(7) Transition to WTR state and continue to send the current message.

(8) Transition to DNR state and continue to send the current message.

(9) If the receiving LER's WTR timer is running, maintain current state and message. If the WTR timer is not running, transition to N.

(10) If the active path just before the SD is selected as the highest local input was the working path, then ignore. Otherwise, go to PF:DW:R and transmit SD(0,1)

(11) If the received SD-P message has Path=1, ignore the message. If the received SD-P message has Path=0 and the active path just before the SD is selected as the highest local input was the working path, then go to UA:DP:R and transmit SD(1,0). If the received SD-P message has Path=0 and the active path just before the SD is selected as the highest local input was the protection path, then ignore the received SD-P message.

12. Security considerations

No specific security issue is raised in addition to those ones already documented in [RFC6378]

13. IANA considerations

13.1. PSC Request Field

This document defines two new values in the "MPLS PSC Request Registry".

The PSC Request Field is 4 bits, and the two new values have been allocated as follows:

Value	Description	Reference
TBD1	Reverse Request	[this document]
TBD2	Exercise	[this document]

[to be removed upon publication: It is requested to assign 2 (=TBD1)for the Reverse Request value and 3 (=TBD2) for the Exercise value to be aligned with the priority levels of those two requests defined in this document.]

13.2. PSC TLV

This document defines a new value for the Capabilities TLV type in the "MPLS PSC TLV Registry".

Type	TLV Name	Reference
TBD3	Capabilities	[this document]

[Editor's note: Need to specify a registry for Value (=options) inside the Capabilities TLV in a later version of this draft]

14. Acknowledgements

15. References

15.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC5654] Niven-Jenkins, B., Brungard, D., Betts, M., Sprecher, N., and S. Ueno, "Requirements of an MPLS Transport Profile", RFC 5654, September 2009.

[RFC6378] Weingarten, Y., Bryant, S., Osborne, E., Sprecher, N., and A. Fulignoli, "MPLS Transport Profile (MPLS-TP) Linear Protection", RFC 6378, October 2011.

[I-D.ietf-mpls-psc-updates] Osborne, E., "Updates to PSC", draft-ietf-mpls-psc-updates-00 (work in progress), October 2013.

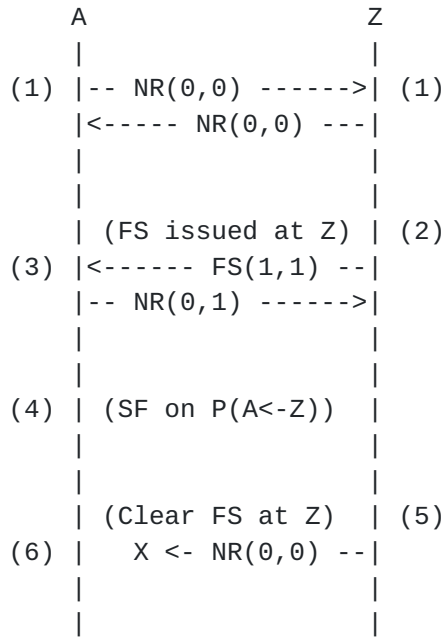
15.2. Informative References

[RFC4427] Mannie, E. and D. Papadimitriou, "Recovery (Protection and Restoration) Terminology for Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4427, March 2006.

[RFC6372] Sprecher, N. and A. Farrel, "MPLS Transport Profile (MPLS-TP) Survivability Framework", RFC 6372, September 2011.

Appendix A. An example of out-of-service scenarios

The sequence diagram shown is an example of the out-of-service scenerios based on the priority level defined in [RFC6378]. The first PSC message which differs from the previous PSC message is shown.



(1) Each end is in Normal state, and transmits NR (0,0) messages.

(2) When a Forced Switch command is issued at node Z, node Z goes into local Protecting Administrative state (PA:F:L) and begins transmission of an FS (1,1) messages.

(3) A remote Forced Switch message causes node A to go into remote Protecting Administrative state (PA:F:R), and node A begins transmitting NR (0,1) messages.

(4) When node A detects a unidirectional Signal Fail on the Protection path, node A keeps sending NR (0,1) message because SF-P is ignored under the state PA:F:R.

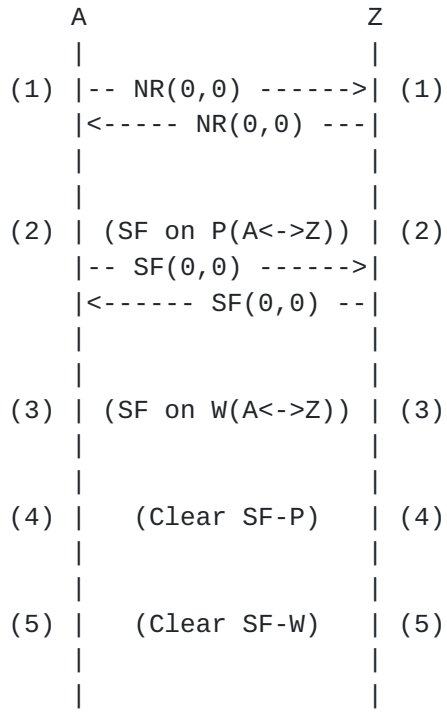
(5) When a Clear command is issued at node Z, node Z goes into Normal state and begins transmission of NR (0,0) messages.

(6) But node A cannot receive PSC message because of local unidirectional Signal Fail on the Protection path. Because no valid PSC message is received, over a period of several successive message intervals, the last valid received message remains applicable and the node A continue to transmit an NR (0,1) message in the state of PA:F:R.

Now, there exists a mismatch between the bridge/selector positions of node A (transmitting an NR (0,1)) and node Z (transmitting an NR (0,0)). It results in out-of-service even when there is neither signal fail on working path nor FS.

Appendix B. An example of sequence diagram showing the problem with the priority level of Clear SF

An example of sequence diagram showing the problem with the priority level of Clear SF defined in [RFC6378] is given below. The following sequence diagram is depicted for the case of bidirectional signal fails. However, other cases with unidirectional signal fails can result in the same problem. The first PSC message which differs from the previous PSC message is shown.



(1) Each end is in Normal state, and transmits NR (0,0) messages.

(2) When signal fail on protection (SF-P) occurs, each node enters into [UA:P:L] state and transmits SF (0,0) messages. Traffic remains on working path.

(3) When signal fail on working (SF-W) occurs, each node remains in [UA:P:L] state as SF-W has a lower priority than SF-P. Traffic is still on the working path. Traffic cannot be delivered as both working and protection paths are experiencing signal fails.

(4) When the signal fail on protection is cleared, local "Clear SF-P" request cannot be presented to the PSC control logic, which takes the highest priority local request and runs PSC state machine, as the priority of "Clear SF-P" is lower than that of SF-W. Consequently, there is no change in state, and the selector and/or bridge keep pointing at the working path, which has signal fail condition.

Now, traffic cannot be delivered while the protection path is recovered and available. It should be noted that the same problem will occur in the case that the sequence of SF-P and SF-W events is changed.

If we further continue with this sequence to see what will happen after SF-W is cleared,

(5) When the signal fail on working is cleared, local "Clear SF-W" request can be passed to the PSC control logic (state machine) as there is no higher priority local request, but this will be ignored in the PSC control logic according to the state transition definition in [RFC6378]. There will be no change in state or protocol message transmitted.

As the signal fail on working is now cleared and the selector and/or bridge are still pointing at the working path, traffic delivery is resumed. However, each node is in [UA:P:L] state and transmitting SF(0,0) message, while there exists no outstanding request for protection switching. Moreover, any future legitimate protection switching requests, such as SF-W, will be rejected as each node thinks the protection path is unavailable.

Appendix C. Freeze Command

The "Freeze" command applies only to the near end (local node) of the protection group and is not signalled to the far end. This command freezes the state of the protection group. Until the Freeze is cleared, additional near end commands are rejected and condition changes and received PSC information are ignored.

"Clear Freeze" command clears the local freeze. When the Freeze command is cleared, the state of the protection group is recomputed based on the persistent condition of the local triggers.

Because the freeze is local, if the freeze is issued at one end only, a failure of protocol can occur as the other end is open to accept any operator command or a fault condition.

Authors' Addresses

Jeong-dong Ryoo (editor)
ETRI
218 Gajeongno
Yuseong-gu, Daejeon 305-700
South Korea

Phone: +82-42-860-5384
Email: ryoo@etri.re.kr

Eric Gray (editor)
Ericsson

Email: eric.gray@ericsson.com

Huub van Helvoort
Huawei Technologies
Karspeldreef 4,
Amsterdam 1101 CJ
the Netherlands

Phone: +31 20 4300936
Email: huub.van.helvoort@huawei.com

Alessandro D'Alessandro
Telecom Italia
via Reiss Romoli, 274
Torino 10148
Italy

Phone: +39 011 2285887
Email: alessandro.dalessandro@telecomitalia.it

Taesik Cheung
ETRI
218 Gajeongno
Yuseong-gu, Daejeon 305-700
South Korea

Phone: +82-42-860-5646
Email: cts@etri.re.kr

Eric Osborne
Cisco Systems, Inc.

Email: eosborne@cisco.com