

Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: September 8, 2016

Q. Wang, Ed.
ZTE
March 7, 2016

Framework and Requirements for GMPLS-based Control of Flexible Ethernet
Network
draft-wang-ccamp-flex-e-fwk-00

Abstract

Flex Ethernet (FlexE) technology, which is defined by Optical Internetworking Forum (OIF), is a new kind of data plane technology and can be used to provide a generic mechanism for supporting a variety of Ethernet MAC rates that may or may not correspond to any existing Ethernet PHY rate. This includes MAC rates that are both greater than (through bonding) and less than (through sub-rate and channelization) the Ethernet PHY rates used to carry FlexE.

Based on the applications/use cases given in the Flex Ethernet Implementation Agreement, this document defines a framework and control plane requirements for the application of existing GMPLS architecture and control plane protocols to the control of flexible Ethernet network. The actual extensions to the GMPLS protocols will be defined in companion documents.

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 September 8, 2016.

Copyright Notice

Copyright (c) 2016 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	3
1.1. Requirements Language	3
2. Overview of FlexE Networks	3
2.1. FlexE Terminology	3
2.2. Scenarios Supported by FlexE	4
2.3. FlexE Layer Model	5
2.3.1. Layer Model in FlexE Unaware Case	5
2.3.2. Layer Model in FlexE Terminating Case	6
2.3.3. Layer Model in FlexE Aware Case	7
3. GMPLS Applicability	8
3.1. General Considerations	9
3.2. Consideration of LSPs in FlexE	9
3.3. Control-Plane Modelling of FlexE Network Elements	9
3.4. FlexE Layer Resource Allocation Considerations	10
3.5. Neighbour Discovery and Link Property Correlation	11
3.6. Routing and Topology Dissemination	11
4. Control-Plane Requirements	11
4.1. Support for Signalling of FlexE	11
4.2. Support for Routing of FlexE	12
4.3. Support for Neighbour Discovery and Link Property and Link Correlation	12
5. Security Considerations	12
6. Manageability Considerations	12
7. References	13
7.1. Normative References	13
7.2. Informative References	14
Author's Address	14

1. Introduction

As defined in Flex Ethernet (FlexE) Implementation Agreement, FlexE can provide a generic mechanism for supporting a variety of Ethernet MAC rates that may or may not correspond to any existing Ethernet PHY rate. In a more detailed description, FlexE employs more than one Ethernet PHYs as server layer and these Ethernet PHYs are bonded together as a FlexE group to carry FlexE client signal. The general capabilities supported by FlexE implementation includes:

Bonding of Ethernet PHYs, e.g., supporting a 200G MAC over two bonded 100GBASE-R PHYs.

Sub-rates of Ethernet PHYs, e.g., supporting a 50G MAC over a 100GBASE-R PHY.

Channelization within a PHY or a group of bonded PHYs, e.g., support a 150G and two 25G MACs over two bonded 100GBASE-R PHYs.

Note that hybrids are also possible, for example a sub-rate of a group of bonded PHYs, for example, a 250G MAC over three bonded 100GBASE-R PHYs.

In order to operate on the Ethernet PHYs, FlexE mechanism operates using a calendar which assigns slot positions on sub-calendars on each PHY of the FlexE group to each of the FlexE clients. The calendar has a granularity of 5G, and has a length of 20 slots per 100G of FlexE group capacity.

Based on the FlexE Implementation Agreement, this document defines the framework for GMPLS-based control of flexible Ethernet network to depict the layer model of Flex Ethernet as well as a set of associated control-plane requirements.

1.1. Requirements Language

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 RFC 2119 [RFC2119].

2. Overview of FlexE Networks

2.1. FlexE Terminology

This section presents the definitions of the terms used in FlexE networks. More details about these terms can be found in OIF Flex Ethernet (FlexE) Implementation Agreement.

FlexE group: the FlexE Group refers to a group of from 1 to n bonded Ethernet PHYs. This version of the Implementation Agreement supports FlexE groups composed of one or more bonded 100GBASE-R PHYs.

FlexE Client: a FlexE Client is an Ethernet flow based on a MAC data rate that may or may not correspond to any Ethernet PHY rate. The FlexE client MAC rates supported by this implementation agreement are 10, 40, and $m * 25$ Gb/s.

FlexE Shim: the FlexE Shim is the layer that maps or demaps the FlexE clients carried over a FlexE group. The FlexE mux refers to the transmit direction which maps the FlexE clients over the FlexE group. The FlexE demux refers to the receive direction which demaps the FlexE clients from the FlexE group.

2.2. Scenarios Supported by FlexE

According to the FlexE Implementation Agreement, FlexE can support a variety of cases. A non-exhaustive list includes:

One case of router to transport connection is where the transport network is unaware of FlexE. This may be used with legacy transport equipment that provides PCS-codeword transparent transport of 100GbE, but provides no special support for FlexE. In this case, all PHYs of the FlexE group are carried independently, but over the same fiber route, over the transport network.

Another case of router to transport connection is where the transport network equipment terminates the FlexE group. In the FlexE terminating case, FlexE group is terminated before crossing the transport network and FlexE client is extracted from the FlexE signal and then carried over the transport network.

The final router to transport example described is one where the transport network is aware that it is carrying FlexE PHYs (as opposed to 100GbE), but the FlexE group is not terminated on the transport equipment. Transport network equipment may "crunch" the PHY of the FlexE group by allowing bits or bytes to be discarded from the unavailable calendar slots at the transport network ingress and these bits or bytes re-inserted with fixed values at the transport network egress. This may be used to support cases where the Ethernet PHY rate is greater than the wavelength rate, the wavelength rate is not an integral multiple of the PHY rate, or there is a reason (for example, wavelengths terminated on different transponder line cards) that it is not possible to terminate the FlexE group in the transport equipment. This kind

of equipment is a kind of special transport equipment which can support partial-rate transport.

2.3. FlexE Layer Model

Based on the cases addressed in section 3.2, FlexE has different kinds of mapping hierarchy accordingly. This section gives description of FlexE layer model in different cases. Figure 1 depicts a FlexE layered network scenario. In this network, B and E are FlexE capable nodes, C and D are OTN ODUflex/ODU4 capable nodes. Node B, C are mainly used to encapsulate the client layer signal into the server layer, while node D, E are mainly used to extract the client layer signal from the server layer signal.

As defined in FlexE Implementation Agreement, a FlexE client may be generated internally within a system, received from an Ethernet PHY or from another FlexE shim. In this network scenario, we suppose the FlexE client is generated in router B.

Feature of cases can be found in section 3.2.

In all cases, FlexE client at node B has a path setup request from source B to destination E.

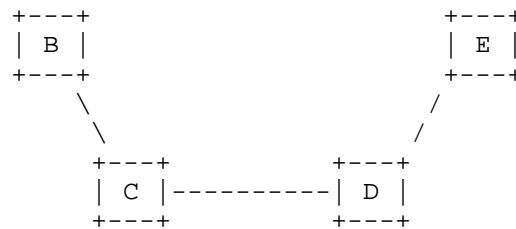


Figure 1: FlexE Layer Network

2.3.1. Layer Model in FlexE Unaware Case

In this case which is depicted in Figure 2, there exist four network layers. FlexE client layer represents an end-to-end connection, which is from the source B to destination E. When the FlexE client signal is generated inside node B, the FlexE client signal is first placed into the slots of FlexE, then the FlexE signal is carried by Ethernet PHYs towards the destination E. When the Ethernet PHYs arrive at node C, each PHY will be mapped into a separate ODU4 connection and then transferred towards the ODU layer connection destination D.

Four layers exist in this case, and the mapping hierarchy between node C and node D can be seen in Figure 3.

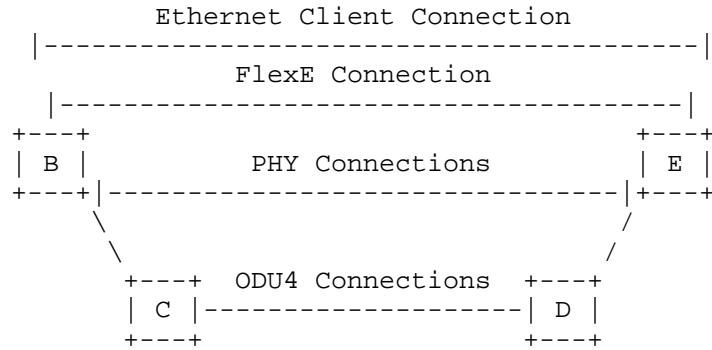


Figure 2: FlexE Unaware Layer Network

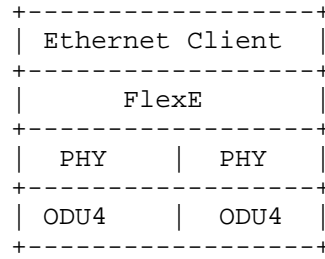


Figure 3: FlexE Unaware Layer Hierarchy

2.3.2. Layer Model in FlexE Terminating Case

In this case, FlexE client layer represents an end-to-end connection, which is from the source B to destination E. When the FlexE client signal is generated inside node B, the Ethernet signal is first placed into the slots of FlexE, then the FlexE signal is carried by Ethernet PHYs towards the destination C. When the FlexE signal arrives at node C, node C first extracts the FlexE client signal, then maps the Ethernet client signal into ODU signal and transfers towards destination node D. Node D will first extract the FlexE client signal from the ODU signal, then map the Ethernet client signal into FlexE signal, which will then be carried by Ethernet PHYs towards destination node E.

Two segments of FlexE connection exist in this case. one is from node B to node C, and the other is from node D to node E.

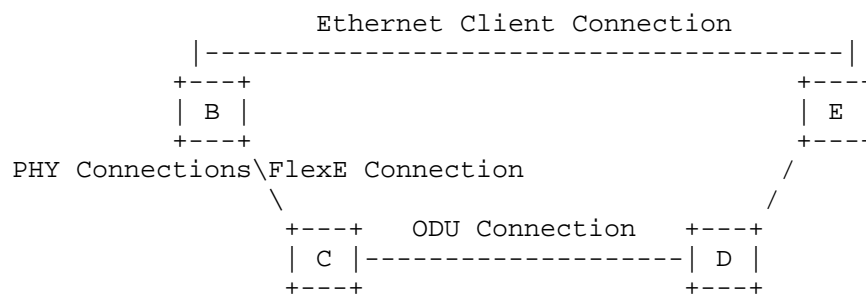


Figure 4: FlexE Terminating Layer Network

Two kinds of mapping hierarchy exist. For the signal transferred on the links between B and C, D and E, the mapping hierarchy can be seen in Figure 5. For the signal transferred on the links between C and D, the mapping hierarchy can be seen in Figure 6.

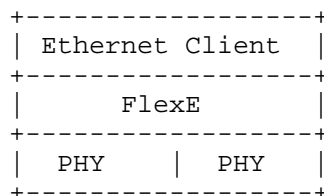


Figure 5: Mapping Hierarchy of FlexE

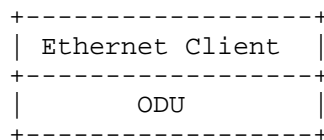


Figure 6: Mapping Hierarchy on OTN Link

2.3.3. Layer Model in FlexE Aware Case

FlexE client layer represents an end-to-end connection, which is from the source B to destination E. When the FlexE client signal is generated inside node B,, the Ethernet signal is first placed into the slots of FlexE, then the FlexE signal is carried by Ethernet PHYs towards the destination E. When the FlexE signal arrives at node C, node C will first discards unavailable slots, then transfers the left slots to ODU Connection.

In this scenario, Ethernet PHYs connection exist between node B and node C, node D and node E.

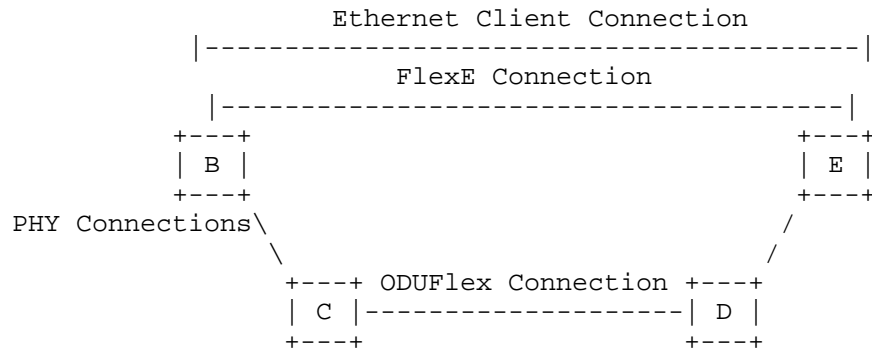


Figure 7: FlexE Aware Layer Network

Two kinds of mapping hierarchy exist. For the signal transferred on the links between B and C, D and E, the mapping hierarchy can be seen in Figure 8. For the signal transferred on the links between C and D, the mapping hierarchy can be seen in Figure 9.

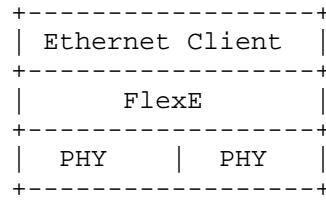


Figure 8: Mapping Hierarchy of FlexE

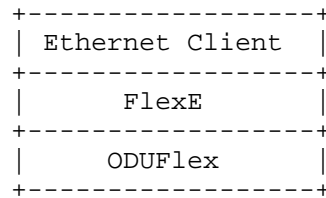


Figure 9: Mapping Hierarchy on OTN Link

3. GMPLS Applicability

The goal of this section is to provide an insight into the application of GMPLS as a control mechanism in FlexE networks. Specific control-plane requirements for the support of FlexE networks are covered in Section 5. This framework is aimed at controlling the FlexE shim layer in different network scenario based on the

capability of FlexE described in OIF Flex Ethernet (FlexE) Implementation Agreement.

3.1. General Considerations

The GMPLS control of the FlexE layer deals with the establishment of FlexE connections that are transferred in FlexE capable nodes. GMPLS labels are used to locally represent the FlexE connections and its associated slots transferring.

3.2. Consideration of LSPs in FlexE

The FlexE LSP is a control-plane representation of a FlexE Connection and MUST be carried by Ethernet PHYs LSP or ODU LSP in the network.

Figure 4 depicts a scenario that the FlexE LSP is carried over Ethernet PHYs LSP between node B and node C, node D and node E. When the Ethernet client signal arrives at node B, node B first check if there are enough Ethernet PHYs available for setting up FlexE LSP. If no, node B will first set up Ethernet PHYs LSP from node B to node C, and then set up the FlexE LSP over the Ethernet PHYs LSP. This process involves twice signalling procedures, one is to set up Ethernet PHYs, and the other is used to set up FlexE LSP over the Ethernet PHYs. The set-up signalling of FlexE LSP includes the allocation of resource for Ethernet client.

Figure 7 depicts a scenario that the FlexE LSP is carried over ODU LSP between node C and node D. This scenario is different, and is used to support cases where the Ethernet PHY rate is be greater than the wavelength rate, the wavelength rate is not an integral multiple of the PHY rate. Node C and node D MUST support the partial-rate ability. When the FlexE LSP over Ethernet PHYs arrives at node C, node C first check if there is enough resource for carrying the FlexE LSP signal across the transport network. If no, node C will check if there is enough resource for carrying FlexE LSP signal after discarding the unavailable slots. If yes, node C will first set up the ODUFlex LSP to node D, and then continue the signalling process of FlexE LSP across the transport network.

3.3. Control-Plane Modelling of FlexE Network Elements

FlexE is a new kinds of transport technology, which brings new constraints. These constraints are listed below:

Unavailable slots: this is different from "unused" slot, in that it is known, due to transport network constraints, that not all of the calendar slots generated from the FlexE mux will reach the FlexE demux and therefore no FlexE client should be assigned to

those slots. As defined in the Flex Ethernet Implementation Agreement, unavailable slots are always at the end of the sub-calendar configuration for the respective PHY.

Unused slots: unused slots can be allocated to Ethernet client as available resource.

Partial-rate capability: the partial-rate capability is usually supported by an OTN access equipment. If an equipment supports partial-rate, it means this equipment has the capability of discarding unavailable slots and transfers the left slots across OTN transport network.

Slot granularity: currently, only one kinds of 5G slot granularity is defined in OIF Flex Ethernet (FlexE) Implementation Agreement.

3.4. FlexE Layer Resource Allocation Considerations

FlexE LSP transfers based on the slot information, so it SHOULD be able to expose the unused slot resource information towards the client layer. Besides the slot information, there are also many other attributes that need to be specified when allocating resource. In GMPLS-controlled system, these information should be taken into consideration as a label when transferring.

FlexE group number: a bunch of Ethernet PHYs can be bounded together and used as a whole by the FlexE LSP. FlexE LSPs between the same source and destination equipment SHOULD NOT have the same FlexE group number. Source equipment and destination equipment SHOULD be aware of the existing of FlexE group and which Ethernet PHYs are in which FlexE group.

PHY Number: it's a dynamic and logical number that is assigned through control plane or management plane, which is unique within the context of (source, destination), and has a one-to-one correlation with physical port. This information will also be carried in the FlexE overhead. Source equipment and destination equipment SHOULD negotiate a value for every Ethernet PHYs within one FlexE group.

Slot Assignment information: the FlexE LSP transfers based on the slot positions, so the equipment SHOULD be able to tell which slot is assigned to which client.

Partial-rate: during the process of resource allocation, where the partial-rate would happen should be indicated.

Granularity: currently, only one kinds of 5G slot granularity is defined in OIF Flex Ethernet (FlexE) Implementation Agreement.

3.5. Neighbour Discovery and Link Property Correlation

There are potential interworking problems between different FlexE capable equipment. Devices or equipments might not be able to support the interworking of every slot due to the constraints of transport network or other constraints. In this case, two directly connected FlexE capable equipments SHOULD run the neighbour discovery process and correlate the link property to make sure which slots are unavailable, which slots can be used by the client.

3.6. Routing and Topology Dissemination

The topology and routing information is used by the path computation entity to compute an end-to-end path. Besides the basic interconnected information, there are also some FlexE specific attributes that should be taken into consideration.

Partial-rate: partial-rate capability is a special feature which allows an equipment to discard unavailable slots and transfers the left slots across OTN transport network. Path computation entity is more likely to compute a feasible path if this capability is taken into consideration when computing path.

Unavailable slot information: this information is used to indicate certain slots SHOULD not be considered when computing an end-to-end path. The unavailable slots can not be used to transfer signal because of the transport constraints.

Unused slot information: unused slot can be allocated to the path as available resource.

4. Control-Plane Requirements

The control of FlexE networks places additional requirements on the GMPLS protocols. This section summarizes those requirements for signalling and routing.

4.1. Support for Signalling of FlexE

The signalling procedures shall be able to assign an FlexE group number for a FlexE LSP, so a number of Ethernet PHYs can be bonded together and uniquely indicated.

The signalling procedures shall be able to assign an unique PHY number for each bonded Ethernet PHY, and a correlation relationship

SHOULD also be indicated between the assigned PHY number and real physical port number when signalling.

The signalling procedures shall be able to configure the slots information allocated for a FlexE LSP.

The Signalling procedures shall be able to indicate the place where partial-rate mapping happens.

4.2. Support for Routing of FlexE

The routing protocol will support all functions described in [RFC4202] and extend them to a FlexE data plane.

The routing protocol SHALL distribute sufficient information to compute paths to enable the signalling procedure to establish LSPs as described in the previous sections.

The routing protocol SHALL update its advertisements of available resources and capabilities to include the partial-rate support information and unused slot information on each Ethernet PHY port.

4.3. Support for Neighbour Discovery and Link Property and Link Correlation

The control plane MAY include support for neighbour discovery such that a FlexE network can be constructed in a "plug-and-play" manner.

The control plane SHOULD allow the nodes at opposite ends of a link to correlate the properties that they will apply to the link. Such a correlation SHOULD include at least the identities of the nodes and the identities that they apply to the link. Other FlexE specific properties, such as the link characteristics of unavailable slot information, SHOULD also be correlated. Such neighbour discovery and link property correlation, if provided, MUST be able to operate in both an out-of-band manner.

5. Security Considerations

TBD

6. Manageability Considerations

TBD

7. References

7.1. Normative References

- [FlexE] Stephen, J. and David. R. Stauffer, "FlexE Implementation Agreement", 2016.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<http://www.rfc-editor.org/info/rfc2119>>.
- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, DOI 10.17487/RFC3209, December 2001, <<http://www.rfc-editor.org/info/rfc3209>>.
- [RFC3471] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, DOI 10.17487/RFC3471, January 2003, <<http://www.rfc-editor.org/info/rfc3471>>.
- [RFC3473] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, DOI 10.17487/RFC3473, January 2003, <<http://www.rfc-editor.org/info/rfc3473>>.
- [RFC3603] Marshall, W., Ed. and F. Andreassen, Ed., "Private Session Initiation Protocol (SIP) Proxy-to-Proxy Extensions for Supporting the PacketCable Distributed Call Signaling Architecture", RFC 3603, DOI 10.17487/RFC3603, October 2003, <<http://www.rfc-editor.org/info/rfc3603>>.
- [RFC4202] Kompella, K., Ed. and Y. Rekhter, Ed., "Routing Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4202, DOI 10.17487/RFC4202, October 2005, <<http://www.rfc-editor.org/info/rfc4202>>.
- [RFC4203] Kompella, K., Ed. and Y. Rekhter, Ed., "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, DOI 10.17487/RFC4203, October 2005, <<http://www.rfc-editor.org/info/rfc4203>>.
- [RFC4204] Lang, J., Ed., "Link Management Protocol (LMP)", RFC 4204, DOI 10.17487/RFC4204, October 2005, <<http://www.rfc-editor.org/info/rfc4204>>.

7.2. Informative References

[RFC3945] Mannie, E., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Architecture", RFC 3945, DOI 10.17487/RFC3945, October 2004, <<http://www.rfc-editor.org/info/rfc3945>>.

Author's Address

Qilei Wang (editor)
ZTE
Nanjing
CN

Email: wang.qilei@zte.com.cn