Network Working Group
Internet Draft
Intended status: Informational

Y. Jiang
F. Yang
I. Busi
Huawei
J. Wang
Fiberhome

Expires: May 2020 November 4, 2019

Problem Statements of FlexE Interface Management
draft-jiang-ccamp-flexe-ifmps-00

Abstract

This document outlines the problem statements for FlexE interface management; it also gives an analysis of configuration requirements for Flex Ethernet (FlexE) interface management. Requirements on FlexE interface management are summarized in the end.

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 https://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 April 4, 2020.

Table of Contents

1. Introduction

The Flex Ethernet (FlexE) 2.0 Implementation Agreement [FLEXE] defined by the OIF provides the support of 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.

Besides 100GBASE-R PHYs as supported in FlexE 1.0, FlexE 2.0 supports the bonding of 200GBASE-R PHYs or 400GBASE-R PHYs respectively. FlexE 2.1 further supports the bonding of 50GBASE-R PHYs.

According to [FLEXE], FlexE supports the following features:

- Bonding of multiple ETH PHYs (1~254)

- Sub-rates of ETH PHY (minimum of 5G to maximum capacity of bandwidth in a PHY)

- Channelization within a PHY or a group of bonded PHYs (5G ~ k*5G, combination of multiple slots, where each slot can be allocated from any PHY)

In the FlexE, multiple Ethernet PHYs (each PHY can further consist of one or more FlexE Instances) are bonded into a FlexE Group, and the total capacity of the FlexE Group is represented as a collection of slots (e.g., each slot has a granularity of 5Gbps or 25Gbps). Based on their bandwidth needs, FlexE Clients are each allocated with one or more slots in a FlexE group. The FlexE mechanism operates by using a calendar consisting of these slots.

This calendar is partitioned into sub-calendars for each PHY (earlier than FlexE 2.0) or sub-calendars for each FlexE instance (FlexE 2.0 and above). For example, the calendar for a FlexE Group composed of n 100G PHYs is partitioned into 20n slots (each slot representing 5Gbps of bandwidth when the slot granularity is 5Gbps).

Some FlexE use cases are introduced in details in [flexe-usecases].

This document describes the problem statements for FlexE interface management to support the transport of FlexE clients over a FlexE Group between two FlexE nodes. The equipment can be routers or optical transport products, which can support FlexE interfaces. Multiple hops of FlexE aware transport in OTN or MTN is out of the scope of this document.

1.1. Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

1.2. Terminology

Terminologies used in this document are extracted from [FLEXE].

FlexE: Flex Ethernet.

FlexE Client: An Ethernet flow based on a MAC data rate that may or may not correspond to any Ethernet PHY rate, the format of the FlexE Client is simply a logically serial stream of 66B blocks at a given rate.

FlexE Group: A FlexE Group is composed of from 1 to n Ethernet PHYs.

FlexE Instance: A FlexE Instance is a unit of information consisting of 100G of capacity able to carry FlexE Client data, together with its associated overhead.

Ethernet PHY: an entity representing Ethernet Physical Coding Sublayer (PCS), Physical Media Attachment (PMA), and Physical Media Dependent (PMD) layers. Each PHY is consisted of one or more FlexE Instance (e.g., a 400GBASE-R PHY has four FlexE Instances).

FlexE Calendar: The total capacity of a FlexE Group is represented as a collection of slots. The calendar for a FlexE Group composed of n PHYs is represented in each PHY as an array of slots (e.g., each representing 5Gbps of bandwidth), i.e., calendar-slot-list.

CCA: Calendar Configuration A

CCB: Calendar Configuration B

2. Problem statements

2.1. Overview of FlexE management

Figure 1 depicts the overview diagram of FlexE management for a FlexE Group between PE1 and PE2, where PE1 and PE2 are network equipments such as routers or OTN products. SDN/NMS may control or manage the FlexE Group between PE1 and PE2 by interactions with PE1 and PE2 separately (by using Netconf or Restconf to connect with a management or control agent on each PE node).

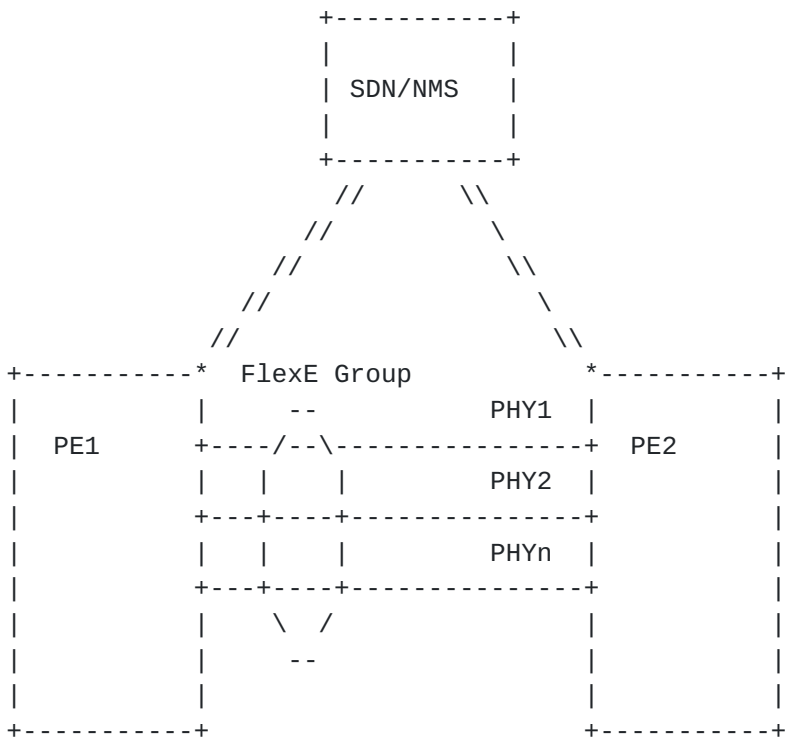

Figure 1: FlexE management overview

2.2. Considerations on parameters in FlexE Overhead

The OIF specifies how the configuration and verification of a FlexE Group can be realized in Section 7.3 of [FLEXE]. A FlexE Overhead Frame is defined in [FLEXE] to convey FlexE group specific information from PE1 to PE2, including configuration information (FlexE Group Number, FlexE Map, FlexE PHY/Instance Number, CCA and CCB), status information (RPF) and signaling information (CC, CR and CA).

The configuration information in the overhead frame is used by the receiving side to verify that both ends are properly configured with the same set of values in a FlexE Group. If PE2 finds out that the configuration

information in the overhead sent from PE1 does not match its own configuration, a mismatch alarm should be raised.

Note: Two calendar configurations are used in the FlexE data plane to facilitate reconfiguration, i.e., CCA and CCB. They are actually two lists (e.g., each list is 20*2Bytes for a PHY of 100GBASE-R; or 4*20*2Bytes for a PHY of 400GBASE-R) wherein each list item indicates the client number carried in a calendar slot. At any given time, only one of the calendar configurations is active and used for mapping the FlexE Clients into the FlexE Group and demapping the FlexE Clients from the FlexE Group. When a switch of calendar configurations adds/removes/resizes FlexE Clients in a FlexE Group, the switching does not affect existing clients whose size and calendar slot assignments are not changed.

Status information indicates status of the bonded PHYs, currently only RPF (Remote PHY Fault) is defined in [FLEXE], which informs the far-end of a locally detected failure of the PHY if set to one.

The signaling information can be used to coordinate the switching from the active calendar configuration (e.g., CCA) to the backup calendar configuration (e.g., CCB) between PE1 and PE2. As described in 7.3.4 of [FLEXE], CC, CR and CA are used to coordinate the switching of calendar A to calendar B or vice versa between the FlexE mux and FlexE demux (i.e., the source and the sink of a FlexE group). The protocol is optional to implement.

2.2.1. Static or configurable

If the FlexE is static, a FlexE Group is composed of a fixed number of FlexE PHYs, e.g., simple bonding for a single fixed client; fixed calendar configuration. That means, there is no need to configure a device or it is impossible to configure the device. Some simple implementations only support static configuration.

If the FlexE is configurable, the FlexE parameters can be controlled by a SDN controller or be configured by a network management system (NMS).

A FlexE static PE usually interconnect with another PE in FlexE static (it is required that both PEs implement a FlexE group with exact the same set of fixed parameter

values). However, sometimes there is a need to interconnect one FlexE static PE with another PE in FlexE configurable if the latter is properly configured with the same set of fixed parameter values.

2.2.2. Negotiation enable or not

[FLEXE] uses two calendar configurations (i.e., CCA and CCB) to facilitate client reconfiguration. Furthermore, Section 7.3 of [FLEXE] specifies a dynamic negotiation protocol (by signaling in the FlexE overhead) for the automatic switching of calendars in a FlexE Group. If this negotiation protocol is enabled (if one PE enables negotiation, the other PE MUST enable negotiation too), a receiving PE (i.e., slave) can further extract the configuration information (particularly CCA and CCB) from the FlexE overhead and multi-frame sent from a sending PE (i.e., master).

It seems that the slave does not need to configure any FlexE parameters if negotiation protocol is enabled. However, from the viewpoint of bidirectional transport, a receiving PE in one direction is also acting as a sending PE in the other direction. Furthermore, FlexE group and its bonding PHYs must be configured firstly so that FlexE overhead channel can be set up for the signaling protocol to work properly. Therefore, FlexE configuration is needed on both side of PEs even if negotiation is enabled.

Since the dynamic signaling of CC, CR and CA is done automatically in the data plane (especially, CR and CA are request and acknowledgement exchanged dynamically over the FlexE overhead, CC decides whether CCA or CCB is active), the mechanism works on the FlexE data plane independent from the management plane. Exposing all these signaling information to the management plane is not only unnecessary, but also greatly complicates the operations of FlexE. Thus, it is not needed to configure or retrieve these ephemeral signaling states.

2.2.3. Management of FlexE clients

The following management of FlexE clients needs to be supported:

-Add a client or clients

-Delete a client or clients

-resize a client or clients

-adjust slot locations for a client or clients

If the negotiation protocol is not enabled, management of FlexE clients (add/delete/resize/adjust) usually is a sequential action to the current calendar of each FlexE PE, and retrieval of the calendar configuration values is also based on the active calendar. Thus, synchronous switching from the active calendar configuration to the backup calendar configuration is not needed. However, some client traffic may be lost during the reconfiguration.

If the negotiation protocol is enabled, management of FlexE clients (add/delete/resize/adjust) usually is a sequential action to the backup calendar of each FlexE PE. Then dynamic negotiation as described in Section 2.2.2 controls peer PEs to switch the backup calendar configuration into the active calendar configuration synchronously. The switching is hitless since the client traffic is not lost during the reconfiguration, thus it is recommended to be the default working mode. Moreover, retrieval of the calendar configuration values SHOULD be based on the new active calendar after protocol convergence (the convergence time is expected to be around 10ms calculated according to [FLEXE]).

In either cases, the management plane only needs to deal with a single calendar, and there is no need to monitor whether the calendar is CCA or CCB from the SDN/NMS point of view.

2.3. Considerations on bidirectional transport

OIF only discusses the configuration of a unidirectional client.

In fact, the overhead signaling of CR and CA relies on a bidirectional channel in the same FlexE Group.

Furthermore, the FlexE links (including each of the bonding PHYs) are always bidirectional, and FlexE clients

are usually reserved with the same number of slots (or bandwidth) in both directions over the same FlexE link. For a FlexE client, the expected value of FlexE parameters to be received will be the same as the values of those parameters configured in the transmit direction on the same PE, thus the expected parameters are not needed to be configured explicitly. If the received parameters are not the same values as those parameters configured locally, a PE should report the mismatch to the SDN controller/NMS. Examples of mismatch may include: FlexE group number mismatch, FlexE PHY number mismatch, Calendar configuration Mismatch, and etc.).

3. Requirements

This section summarizes the management requirements of FlexE interfaces.

a). Support of a flexible FlexE group bonding with one to 254 Ethernet PHYs, the Ethernet PHY types may include 50 GBASE-R, 100GBASE-R, 200GBASE-R and 400GBASE-R.

b). Support add/remove Ethernet PHYs to/from a FlexE group, the range of FlexE PHY number still follows a).

c). Support of flexible FlexE client management in a FlexE group, and the total clients number can be in a range of from 0 to a value equal to the maximum number of slots in the FlexE group (that is, each client is allocated a single slot).

d). Support add/delete/resize FlexE clients in a FlexE group without impacts on the traffic of any existing FlexE clients in the same FlexE group, the range of FlexE client number still follows c).

e). Support FlexE static or FlexE configurable operations.

f). Support coordination of calendar updates and switchover by enabling FlexE negotiation between peer FlexE PEs.

g). Support a client with bidirectional slot allocation while its bandwidth can be inferred from the allocated slot number and slot granularity.

h). Support retrieval status of a FlexE group, a FlexE PHY, or a FlexE client.

i). Management shall be compatible as much as possible with all OIF FlexE versions, including 1.0, 1.1, 2.0 and 2.1.

4. Security Considerations

This document gives the problem statements for FlexE management, and summarizes the requirements. As no solution is discussed in this document, no security concerns are raised.

5. IANA Considerations

There are no IANA actions required by this document.

6. References

6.1. Normative References

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

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, May 2017

6.2. Informative References

[FLEXE] OIF, "Flex Ethernet 2.0 Implementation Agreement", FlexE 2.0, June 2018

[flexe-usecases] Hussain, I., Valiveti, R., Pithewan, K., "FlexE Usecases", draft-hussain-ccamp-flexe-usecases-01, work in progress

7. Acknowledgments

The authors would like to thank Jinfeng Chen and Yalei Han for their contribution to this document.

Authors' Addresses

Yuanlong Jiang
Huawei Technologies Co., Ltd.
Bantian, Longgang district
Shenzhen 518129, China
Email: jiangyuanlong@huawei.com

Fan Yang
Huawei Technologies Co., Ltd.
Huawei Campus, No. 156 Beiqing Rd.
Beijing 100095
Email: shirley.yangfan@huawei.com

Italo Busi
Huawei Technologies Co., Ltd.
HUAWEI TECHNOLOGIES ITALIA Srl Centro Direzionale Milano 2
Milan, Milan 20090, Italy
Email: Italo.Busi@huawei.com

Junfang Wang
Fiberhome
Email: wjf@fiberhome.com