Flow Control Extension for DLEP
draft-dalela-dlep-flow-control-00.txt

Status of this Memo

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

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79. This document may not be modified, and derivative works of it may not be created, and it may not be published except as an Internet-Draft.

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79. This document may not be modified, and derivative works of it may not be created, except to publish it as an RFC and to translate it into languages other than English.

This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.

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

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."
Abstract

There is a need for a mechanism to flow control the data plane traffic between the router and the modem. This draft extends DLEP protocol to provide Ethernet based flow control schemes.

Table of Contents

1. Introduction........................................................................................................3
2. Conventions used in this document.................................................................3
3. Terminology.........................................................................................................4
4. Overview.............................................................................................................4
5. Priority-based Flow Control..........................................................5
   5.1. Operation..............................................................6
   5.2. PFC Data Item Definitions........................................6
   5.3. Limitations.........................................................7
1. Introduction

The Dynamic Link Exchange protocol [DLEP] runs between a router and its attached modem devices, allowing the modem to communicate radio link characteristics, and radio network convergence events as they change. Flow control in DLEP is an optional extension. One possible extension is [CREDIT].

This draft describes an extension to DLEP protocol, allowing for PFC based flow control schemes to be implemented on a destination-by-destination basis.

1. Priority Flow Control (PFC) that is, [IEEE802.1Qbb]

2. PFC in conjunction with MAC address

3. PFC in conjunction with VLAN tagging.

This draft describes an extension to DLEP protocol and doesn’t attempt to suggest or make any updates/modifications to any other protocol or standard mentioned in the draft.

2. Conventions used in this document

In examples, "C:" and "S:" indicate lines sent by the client and server respectively.

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].

In this document, these words will appear with that interpretation only when in ALL CAPS. Lower case uses of these words are not to be interpreted as carrying significance described in RFC 2119.
In this document, the characters ">>" preceding an indented line(s) indicates a statement using the key words listed above. This convention aids reviewers in quickly identifying or finding the portions of this RFC covered by these keywords.

3. Terminology

Peer: A router and the modem with which it has established a DLEP session are referred to as peers.

Neighbor: A router which is accessible via a peer’s RF link.

CoS: This is a 3-bit 802.1p Class of Service value in an Ethernet frame defined in the IEEE 802.1q standard. Up to 8 different traffic classes are supported.

Traffic Class: A stream of traffic characterized by 802.1p Class-of-Service or DSCP.

HWM/LWM: Higher and lower water marks are the thresholds set against the buffer occupancy levels at which traffic should be stopped/started via IEEE 802.1Qbb pause frames.

4. Overview

This protocol extension to DLEP provides a PFC based flow control scheme. Here, we present an overview of PFC based flow control mechanisms which offloads the flow control into the Ethernet layer, rather than keeping them within DLEP.

In the PFC based schemes, data plane traffic flowing between two devices is being controlled by the buffer occupancy level and the thresholds set against the receive buffers per traffic class level. On sensing potential of congestion for CoS based traffic class it sends a PFC PAUSE frame with a bitmap of all traffic classes to the link partner, indicating which traffic classes are to be paused. TRANSMIT-OFF/TRANSMIT-ON pair can be sent for each individual priority based flow, while all other flows can continue transmitting and receiving data. As the congestion alleviates, another PFC PAUSE frame is sent so as to enable link partner to start transmitting the traffic for the given traffic class.

In PFC, the sender of the PFC PAUSE frame assigns separate logical queues for each traffic class. The sender SHOULD manage and account buffer occupancy levels and HWM, LWM thresholds for each logical queue.
This schemes uses IEEE 802.1Qbb PAUSE semantics and uses it in conjunction with L2 constructs to provide flow control schemes for scale. Depending upon the number of neighbors and traffic classes per neighbor support requirement, we propose three flow control schemes based on PFC.

1. Priority-based Flow Control (PFC) using CoS: This mechanism is limited to use 8 priorities, which must be distributed over traffic classes and neighbors. It can be used if there are few neighbors or traffic classes (their product not exceeding 8).

2. PFC in conjunction with MAC address: The PFC mechanism is augmented with the Source MAC value in the PFC pause frame. The CoS values of 802.1p field defined in [IEEE 802.1q] classify the traffic destined to a neighbor into 8 traffic classes, and the source mac address value is used to map it to the neighbor whose traffic class is under congestion.

3. PFC in conjunction with VLAN tagging: The PFC mechanism is used in conjunction with the 802.1q VLAN ID. The VLAN ID represents traffic to a particular neighbor, and 8 traffic classes to each neighbor are supported.

The traffic flow control schemes needs to cater to following traffic flow scenarios:

1. Router to Modem traffic - When a DLEP router sends data to the modem and the modem’s RF link is experiencing congestion, the modem MUST send a PFC PAUSE frame to the router.

2. Modem to Router traffic - When a DLEP modem is sending data to the router and the router experiences congestion on its egress port it MUST send a PFC PAUSE frame to the modem.

The PFC pause frame sent by the modem MUST have information about the DLEP neighbor that is experiencing congestion in addition to the specific traffic class to be flow controlled.

The flow control mechanisms proposed here will be working within DLEP peers scope, it doesn’t define or propose any flow control scheme for Modem to Modem traffic over RF link.

5. Priority-based Flow Control

In this scheme, traffic class per neighbor is characterized by the 802.1p CoS values. Hence it allows max 8 neighbors with single traffic class per neighbor. Conversely a single neighbor is
supported if there are 8 traffic classes per neighbor. For example, if there are 2 neighbors, we can support up to 4 CoS values per neighbor.

5.1. Operation

DLEP peers supporting this method of flow control MUST include a DLEP 'Extensions Supported’ data item, including the value TBD representing this extension in the "DLEP Destination UP Response" message. As part of DLEP Destination UP Response message, the router will use the "CoS bitmap Data Item" to assign CoS values to the individual DLEP peers. Refer section 5.2 for details.

Traffic destined to modem is mapped to a unique CoS value based on the ingressing packets traffic class and the neighbor it is destined to. CoS value characterizes the traffic class per neighbor.

The logical queue in a router or a modem implementing this scheme is a function of Class of Service value.

The router SHALL decide at initialization time how many neighbors and CoS values per neighbor to support.

The modem MUST assign the CoS value, obtained from the router, to all incoming packets from the RF link and then send it to the router.

5.2. PFC Data Item Definitions

1. CoS bitmap

<table>
<thead>
<tr>
<th>Data Item Type</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>CoS bitmap</td>
<td></td>
</tr>
</tbody>
</table>

---

Figure 1 Assignment of CoS values to Neighbor
Data Item Type: TBD
Length: 4
CoS bitmap: The "Cos Values" allocated to traffic classes for a neighbor.

5.3. Limitations

All neighbors and their traffic classes need to be allocated from the total available 8 CoS values, which greatly reduces the number of neighbors and the number of traffic classes per neighbor that can be supported.

6. PFC using CoS and MAC for DLEP flow control

When the router or modem is generating the pause frame, it MUST set the value of the source MAC address to the MAC address of the DLEP neighbor corresponding to which logical queues threshold has been hit. On receiving the pause frame, the DLEP peer MUST use SMAC value in addition to the traffic class to identify the queue to start/stop traffic.

The logical queue in a router or a modem implementing this scheme is a function of Class of Service value and the source MAC address.

6.1. PFC PAUSE frame

IEEE 802.1Qbb standards PAUSE frame is re-interpreted to support this scheme using both Class of Service and MAC. The SMAC denotes the MAC of the DLEP neighbor instead of its own MAC.
The PFC PAUSE frame has the following fields:

1. Destination MAC: 6 octets, is reserved: 01:80:C2:00:00:01.

2. Source MAC: 6 octets, The SMAC is set to the MAC of the DLEP neighbor.

3. EtherType: 2 octets, the ethertype is set to 0x8808 to indicate flow control.

4. OpCode: 2 octets, The opcode field is set to 0x0101 to indicate PFC.

5. Class-enable vector: 2 octets, The 16 bits of the 2 octets are set to either to 0 or 1; generate pause(1) or ignore(0) for each of the 8 CoS values.

6. Time/Class n: 2 octets per CoS totaling 16 octets for 8 different CoS values indicating the time to pause the particular CoS value. Each field has a range of 0-0xFFFF. A value of zero has the meaning of unpausing the link.
7. PAD: 28 octets of padding.

8. CRC - The Cyclic redundancy check.

7. PFC using the VLAN ID for DLEP flow control

In this scheme, traffic class per neighbor is characterized by the CoS values and Vlan is used to identify the neighbor. Hence it allows up to 4K neighbors with 8 traffic classes per neighbor; that is, a modem can be paired with 4K other modems over RF. To support such a scheme, Vlan-id needs to be carried in the PFC PAUSE frame generated by modem to distinguish PAUSE frames corresponding to different neighbors. We propose to add IEEE 802.1Q tag to the PFC PAUSE frame generated by a DLEP node to carry the Vlan-ID indicating the neighbor for which the PAUSE frame has been generated.

7.1. Operation

DLEP peers supporting this method of flow control MUST include a DLEP 'Extensions Supported' data item, including the value TBD representing this extension in the appropriate "DLEP Destination UP Response" message. As part of "DLEP Destination UP Response" message, the router will use the "VLAN-ID Data Item" to assign Vlans to the individual DLEP neighbors. Refer section 7.3 for details.

Traffic destined to modem SHOULD be mapped to a specific Vlan-ID as a function of the neighbor the traffic is destined to.

When the router or modem is generating the PFC PAUSE frame, it MUST tag the pause frame with 802.1Q tag with the vlan-id indicating the neighbor experiencing congestion. On receiving PAUSE frame, the DLEP peer MUST use vlan-id in addition to traffic class to identify the queue to start/stop traffic.

The logical queue in a router or a modem implementing this scheme is a function of the Class of Service value and the VLAN-ID of the neighbor.

7.2. Tagged PFC PAUSE frame

We propose to use the PFC PAUSE frame defined in IEEE 802.1Qbb with an IEEE 802.1Q tag.
The tagged PFC PAUSE frame has the following fields:

1. Destination MAC: 6 octets, is reserved: 01:80:C2:00:00:01.

2. Source MAC: 6 octets, The MAC address of the sender of the PFC PAUSE frame.

3. Tag Protocol Identifier: 2 octets, The TPID is set to 0x8100 to indicate VLAN tagging.

4. PCP/DEI/VLAN ID: 2 octets, The 12-bit VID field is used to indicate the number of PFC "virtual links".

5. EtherType: 2 octets, is set to 0x8808 to indicate flow control.

6. OpCode: 2 octets, is set to 0x0101 to indicate PFC.
7. Time/Class n: 2 octets per CoS totaling 16 octets for 8 different CoS values indicating the time to pause the particular CoS value. Each field has a range of 0-0xFFFF. A value of zero has the meaning of unpausing the link.


9. CRC - The Cyclic redundancy check.

7.3. Data Item Definitions

1. VLAN-ID

Data Item Type: TBD
Length: 4
DSCP: The "Vlan-ID" allocated to DLEP neighbor

The "VLAN-ID" data item is sent from a DLEP router to the peer as part of "DLEP Destination Up Response" message.

8. Security Considerations

Security considerations similar to [CREDIT] extension are applicable here. This extension doesn’t bring in any new security considerations.

9. IANA Considerations

This section specifies requests to IANA.

9.1. Registrations

This specification defines new data items for DLEP. Assignments from the DLEP data item registry are requested for:

Dalela Expires September 4, 2016 [Page 11]
1. CoS bitmap

2. VLAN-ID

10. References

10.1. Normative References


[CREDIT] Ratliff, S., "Credit Windowing extension for DLEP",


[802.1Q] IEEE 802.1Q, "Virtual LANs",
http://www.ieee802.org/1/pages/802.1Q.html

[802.1Qbb] IEEE 802.1Qbb or PFC, "Priority-based Flow Control",
http://www.ieee802.org/1/pages/802.1bb.html

11. Acknowledgments

This document was prepared using 2-Word-v2.0.template.dot.

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

Redistribution and use in source and binary forms, with or without modification, is permitted pursuant to, and subject to the license terms contained in, the Simplified BSD License set forth in Section 4.c of the IETF Trust’s Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info).

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

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

Dalela          Expires September 4, 2016      [Page 12]
o Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

o Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

o Neither the name of Internet Society, IETF or IETF Trust, nor the names of specific contributors, may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.
Authors’ Addresses

Ashish Dalela
Cisco Systems,
Cessna Business Park,
Bangalore 560103
India
Email: adalela@cisco.com

Arun Parashar
Cisco Systems,
Cessna Business Park,
Bangalore 560103
India
Email: arparash@cisco.com

S L Ananth
Cisco Systems,
Cessna Business Park,
Bangalore 560103
India
Email: alaxmina@cisco.com