Discussion of Congestion Isolation Changes to 802.1Q

Paul Congdon (Huawei),

IEEE 802.1 DCB

Geneva, Switzerland

January 2018

High Level Questions

- **Do we support and define end-station behavior?**
  - yes
- **Should we create independent and parallel congestion isolation clauses, or find a way to stuff new things into a significantly modified Congestion Notification set of clauses? For example, change Congestion Notification clause to Congestion Management clauses and insert Congestion Isolation sub-clauses where appropriate? (NOTE: remainder of slides assume separate independent clauses).**
  - Separate independent
- **Should we define a new (optional) traffic selection algorithm for congested traffic to avoid out-of-order packets and starvation, or simply reference an example in an informative annex?**
  - yes
- **Where will feature specific YANG model additions be added? – TBD**
  - Consider creating UML instead of clause 12 and consider what level of diagnostics are required
- **Should we follow the Qau lead and create 4 main clauses (principles of operation, entity operation, protocol, PDU encoding)? Can't we consolidate a bit?**
  - Yes in principle, but likely only 2 clauses or subclauses

Scope

- **Scope 1.1**
  - Insert a paragraph describing scope of CI. **Question**: Does it need to go at the end of the list, or can it be inserted just after Congestion Notification text for relevancy and to aid the reader?
    - This Standard specifies protocols, procedures and managed objects that support the isolation of congested data flows within networks of limited bandwidth delay product. This is achieved by enabling bridges to individually identify flows creating congestion, adjust transmission selection for packets of those flows and signal to the upstream peer. This mechanism avoids head-of-line blocking for uncongested flows sharing a traffic class in lossless networks. Congestion Isolation is intended to be used with higher layer protocols that utilize end-to-end congestion control in order to reduce packet loss and latency. To this end, it:

      bd) Defines a means for VLAN-aware Bridges that support congestion isolation for identifying flows that are creating congestion.

      be) Defines a means for adjusting transmission selection for packets of congested flows

      bf) Provides for a means for discovering peer VLAN-aware Bridges and stations that support congestion isolation

      bg) Defines a means for signaling congestion isolation to supporting upstream peer Bridges and stations.

Definitions

- **Questions:**
  - Should we modify some existing terms or define all new ones (e.g Congestion Point)?
  - Do we want to support end-stations or only bridge components?
- **Some Proposed Needed Terms:**
  - **Congestion Isolation Aware System**: A bridge component conforming to the congestion isolation provisions of this standard.
  - **Congested Flow:** A sequence of frames the end-to-end congestion controlled higher-layer protocol treats as belonging to a single flow that is experiencing congestion within a Congestion Isolation Aware System.
  - **Congestion Isolation Message (CIM):** A message transmitted by a Congestion Isolation Aware System, conveying congestion Congested Flow information used by the upstream peer Congestion Isolation Aware System.
  - **Congestion Isolation Point (CIP):** A Congestion Isolation Aware System that monitors a set of queues for Congested Flows, and can generate Congestion Isolation Messages.

Abbreviations

- **Some Possible Abbreviations:**
  - CF Congested Flow
  - CI Congestion Isolation
  - CIM Congestion Isolation Message
  - CIP Congestion Isolation Point

Conformance

- **Questions: Assume we need to put these on the end of clause 5.4 and not insert after Congestion Notification (5.4.3) in order to avoid renumbering others?**
- **5.4.7 VLAN Bridge requirements for congestion isolation**
  - A VLAN-aware Bridge implementation that conforms to the provisions of this standard for congestion isolation (46, 47, 48, 49) shall:
    a) Support, on one or more Ports, the creation of at least one Congestion Isolation Point (xx.x.x);
    b) Support, at each Congestion Isolation Point, the generation of Congestion Isolation Messages (xx.x);
    c) Support the ability to configure the variables controlling the operation of each Congestion Isolation Point (xx.x.x);
    d) Conform to the required capabilities of the LLDP of 5.2 of IEEE Std 802.1AB-2009;
    e) Support the use of the Congestion Isolation TLV in LLDP (xx.x.x).
  - A VLAN Bridge implementation that conforms to the provisions of this standard for congestion isolation may:
    f) Support the creation of up to four CIPs on a Bridge Port (xx.x.x)
    g) Support Congested Traffic Enhanced Traffic Selection (yy.y.y)
    h) Support the Congestion Isolation YANG model (xx.x.x).

8.6.6 Queuing of frames

- **We need to include text (similar to Congestion Notification) about CIM frame generation out of the Forwarding process. This could be done as follows:**
  - In a congestion aware Bridge (Clause 30) or a congestion isolation aware Bridge (Clause XX), the act of queuing a frame for transmission on a Bridge Port can result in the Forwarding Process generating a CNM or a CIM. The CNM ~~is~~ or CIM are injected back into the Forwarding Process (8.6.1) as if it had been received on that Bridge Port.

- **Changes to schedule to avoid starvation and out-of-order packets should be documented in the main clauses, not clause 8**

Clause 12 Management

- **Insert Congestion Isolation objects into the existing list of managed objects in 12.1.1**

  j) The ability to create and delete the functional elements of congestion isolation and to control their operation

- **Insert Congestion Isolation in the list of VLAN Bridge Objects of 12.2**

  p) The congestion isolation entities (12.xx).

- **New clause 12.xx Congestion Isolation managed objects with potential objects:**

  a) CI component managed object

  b) Congestion Isolation Point (CIP) component managed object

- **NOTE: No need for the Congested Flow Table to be managed because it would be programmed by the CIP.**

CI Managed Object

- **A per-bridge object that includes:**
  - ciMasterEnable
    - Boolean that turns on or off the functionality
  - ciCimTransmitPriority
    - Integer between 0-7 that determines the traffic class used to send CIM messages upstream. Default 7
  - ciInactivityTimeout
    - The inactivity timeout for entries placed in the congested flow table by the CIP when congested flows are detected.

Congestion Isolation Point (CIP) Managed Object

- **A per CIP object that includes:**
  - cipMonitoredQueues
    - A bit mask of traffic classes that are to be monitored for congestion
  - cipCongestedQueue
    - An integer specifying which traffic class will be used as the congested flow queue for the monitored queues
  - cipMinHeaderOctets
    - The minimum number of octets that the CIP is to return in theEncapsulated MSDU field (xx.x.x) of each CIM it generates (xx.xx.x). Default value 64.
  - cipTransmittedCims
    - A counter of the number of CIMs sent

NOTE: we may need to replicate the cpXX variables that are used to detect congestion

Principles of Operation

- **Introduce the concepts essential to congestion isolation, including:**
  - Reference diagram
  - Problems being solved; preventing head-of-line blocking in a lossless environment while providing time for the end-to-end congestion control loop to react.
  - Requirements and objectives for the solution
  - Methods for identifying congested flows
  - Operation of adjusting traffic selection for congested flows
  - How signaling to upstream peer mitigates head-of-line blocking
  - Relationship and comparison with Congestion Notification

- **Start off with a reference diagram showing the problem/incast**

Congestion Isolation Entity Operation

- **Describes the architecture of the CIP in the Forwarding Process, including:**
  - Congestion Isolation aware Bridge Forwarding Process diagram (see next slide)
  - Congestion Isolation Point (CIP)
  - Congestion Isolation Input Multiplexor – how CIMs are decoded and received.
  - Congested Flow Identification and Table

Architecture Fit Considerations/Thoughts

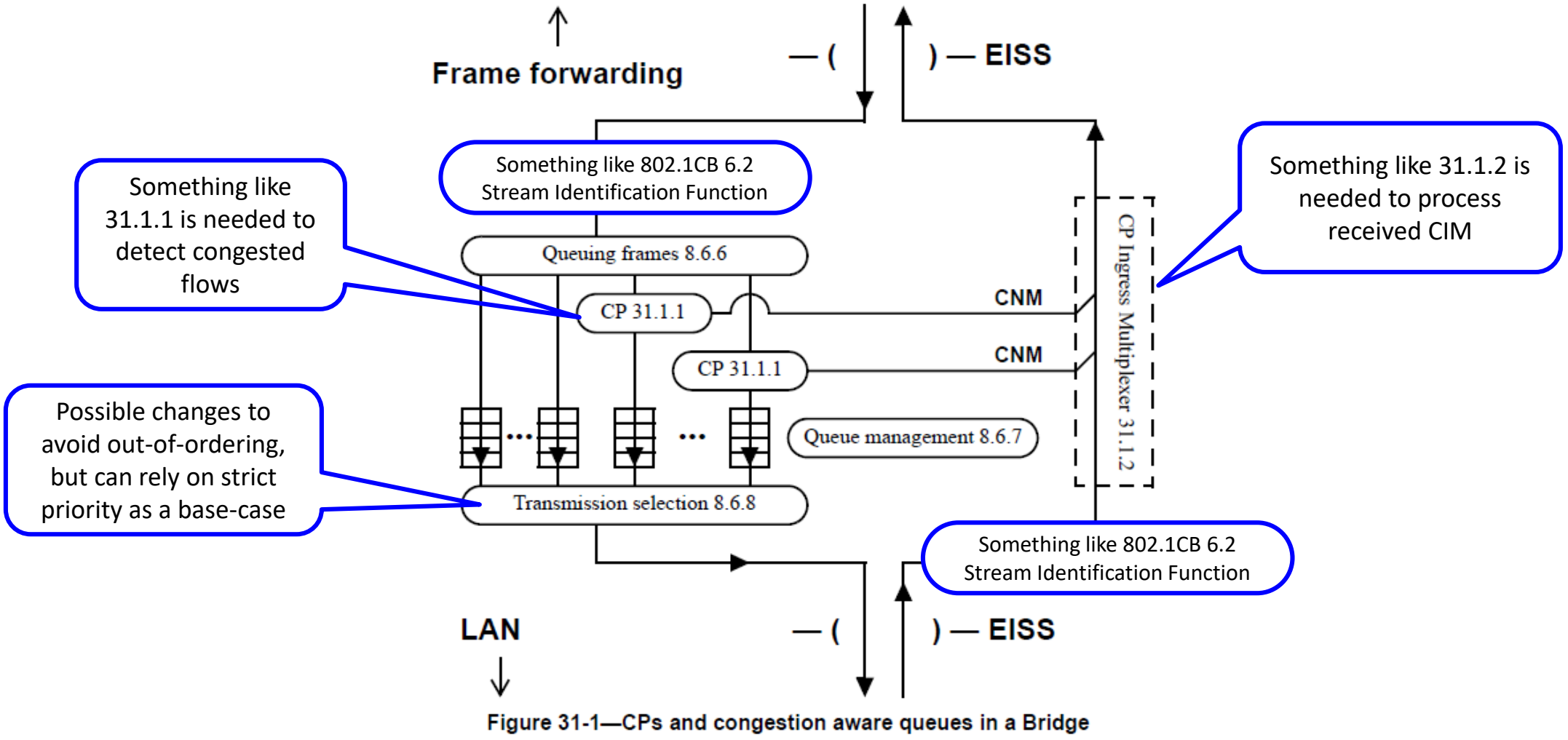

Figure 31-1—CPs and congestion aware queues in a Bridge

Congestion Isolation Protocol

- **Specifies how Congestion Isolation Aware systems participate in the congestion isolation protocol and procedures, including:**
  - Variables controlling operation
  - State machine and associated variables and procedures that control the CIP and congested flow table
  - The Congestion Isolation Protocol
    - Generating CIMs
    - Creating entries in the Congested Flow Table
    - Queuing frames via EM_UNITDATA.request
    - Processing received CIMs
  - Encoding of CIM PDUs
  - Congestion Isolation LLDP TLV definition

YANG Model

- **How and were to include this?**