

PFC Headroom Project PAR&CSD Discussion

Lily Lv, Paul Congdon (Huawei)

Mick Seaman (Independent)

Brad Booth (Microsoft)

Individually contributed PAR proposals(1/3)

5.1 Approximate number of people expected to be actively involved in the development of this project:

20

5.2.b. Scope of the project:

Use of the existing Precision Time Protocol (PTP) and enhancements to the Data Center Bridging Capability Exchange protocol (DCBX) with procedures and managed objects to support automated Priority Flow Control (PFC) headroom calculation and Media Access Control security (MACsec) protection of PFC frames, with a particular emphasis on the low latency and lossless requirements of large-scale and geographically dispersed data centers.

Support of the existing PFC MAC Control interface, with addition of an option to transmit PFC frames using the Internal Sublayer Service interface (IEEE Std 802.1AC) supported by MACsec (IEEE Std 802.1AE).

This project also addresses errors and omissions in the description of existing IEEE Std 802.1Q functionality.

Individually contributed PAR proposals(2/3)

5.3 Is the completion of this standard dependent upon the completion of another standard: No

5.4 Purpose: Bridges, as specified by this standard, allow the compatible interconnection of information technology equipment attached to separate individual LANs.

5.5 Need for the Project: PFC is used to avoid packet loss in low latency, high reliability Ethernet data centers and data center interconnects. For PFC to function properly and without wasting memory, the amount of headroom buffer must be calculated. Deployment in large scale data center networks and long-distance interconnects is currently problematic requires manual configuration. Separately, there are customer requirements for the integrity and confidentiality protection of all frames transmitted between geographically distributed data centers. The current specification is inconsistent and incomplete regarding the operation of PFC and MACSec together.

Individually contributed PAR proposals(3/3)

5.6 Stakeholders for the Standard: Developers and users of networking for data center environments including integrated circuit developers, switch and end-node adaptor vendors, network operators and users.

Individually contributed CSD proposals(1/4)

1.2.1 Broad market potential

Each proposed IEEE 802 LMSC standard shall have broad market potential. At a minimum, address the following areas:

a) Broad sets of applicability.

The data center market continues to grow very fast. Networks with tens of thousands of nodes are common in cloud data centers. There is a trend to converge computing and storage on Ethernet in these networks, supporting cloud services and high-performance applications such as Artificial Intelligence (AI), Machine Learning (ML), and High-Performance Computing (HPC). Remote Direct Memory Access over Converged Ethernet (RoCEv2) is widely deployed, both within data centers and across data center interconnects. RoCEv2 requires lossless operation on Ethernet to avoid wasteful retransmissions. PFC enhancements make Ethernet technology more applicable and appealing for data center environments.

Individually contributed CSD proposals(1/4)

1.2.1 Broad market potential

Each proposed IEEE 802 LMSC standard shall have broad market potential. At a minimum, address the following areas:

b) Multiple Vendors and Numerous Users.

Most cloud data center providers are using RoCEv2 within their networks, require lossless operation, and currently need to configure PFC manually. RoCEv2 is also used by large enterprises, financial institutions and other high-performance computing environments. Further there is strong interest in accessing new high-speed solid-state data storage technologies over Ethernet networks using RoCEv2.

There are public cloud vendor requirements for MACsec integrity and confidentiality protection of all frames transmitted between geographically distributed data centers.

Individually contributed CSD proposals(2/4)

1.2.2 Compatibility

Each proposed IEEE 802 LMSC standard should be in conformance with IEEE Std 802, IEEE 802.1AC, and IEEE 802.1Q. If any variances in conformance emerge, they shall be thoroughly disclosed and reviewed with IEEE 802.1 WG prior to submitting a PAR to the Sponsor.

a) Will the proposed standard comply with IEEE Std 802, IEEE Std 802.1AC and IEEE Std 802.1Q?

Yes

b) If the answer to a) is no, supply the response from the IEEE 802.1 WG.

The review and response is not required if the proposed standard is an amendment or revision to an existing standard for which it has been previously determined that compliance with the above IEEE 802 standards is not possible. In this case, the CSD statement shall state that this is the case.

The project will be in conformance with IEEE Std 802, IEEE Std 802.1AC, and the existing provisions of IEEE 802.1Q.

Individually contributed CSD proposals(3/4)

1.2.3 Distinct Identity

Each proposed IEEE 802 LMSC standard shall provide evidence of a distinct identity. Identify standards and standards projects with similar scopes and for each one describe why the proposed project is substantially different.

There are no other IEEE 802 standards or projects that specify automatic configuration of PFC headroom, and the current standard does not adequately specify MACsec protection of PFC frames. Interoperability issues have been encountered.

Individually contributed CSD proposals(4/4)

1.2.4 Technical Feasibility

Each proposed IEEE 802 LMSC standard shall provide evidence that the project is technically feasible within the time frame of the project. At a minimum, address the following items to demonstrate technical feasibility:

a) Demonstrated system feasibility.

The proposed project incorporates techniques for peer-to-peer link delay measurement and information exchange mechanisms that are currently available in many production bridges and end-stations. PFC operation with MACsec support has been deployed, although existing implementations do not necessarily interoperate.

b) Proven similar technology via testing, modeling, simulation, etc.

The proposed project enables peer nodes to advertise the new capability through the DCBX mechanism which is widely deployed today using LLDP. Roundtrip delay measurements for participating systems are based on the existing PTP delay measurement mechanism.

Backup – change tracking

Individually contributed PAR proposals(1/3)

5.1 Approximate number of people expected to be actively involved in the development of this project: 20

5.2.b. Scope of the project:

Use of the existing Precision Time Protocol (PTP) protocol and enhancements to the Data Center Bridging Capability Exchange protocol (DCBX) protocol with procedures and managed objects to support automated Priority Flow Control (PFC) headroom calculation and Media Access Control security (MACsec) protection of PFC frames, with a particular emphasis on the low latency and lossless requirements of large-scale and geographically dispersed data centers.

Support of the existing PFC MAC Control interface, with addition of an option to transmit PFC frames using the Internal Sublayer Service interface (IEEE Std 802.1AC) supported by MACsec (IEEE Std 802.1AE).

~~This amendment specifies protocols, procedures, and managed objects that support the automatic configuration of PFC headroom, and MACSec protection of PFC frames in order to simplify PFC deployments in large-scale data centers and long-distance data center interconnects, enabling low latency and high reliability Ethernet networking for data centers.~~

~~This amendment defines a shim layer between the existing PFC function and the MAC Service and (optional) MAC Control layers in a bridge. The shim preserves the existing interface to the PFC function and allows the generation and consumption of PFC frames above the MAC Service layer, enabling MACSec protection if desired.~~

~~This amendment automatically calculates and configures PFC headroom based on internal processing delays and roundtrip delay measurements using existing PTP protocols between participating systems. Participating systems are discovered, and parameters are shared through enhancements to the DCBX protocol.~~

This project also addresses errors and omissions in the description of existing IEEE Std 802.1Q functionality.

Individually contributed PAR proposals(2/3)

5.3 Is the completion of this standard dependent upon the completion of another standard:

No

5.4 Purpose: Bridges, as specified by this standard, allow the compatible interconnection of information technology equipment attached to separate individual LANs.

5.5 Need for the Project: PFC is used to avoid packet loss in low latency, high reliability Ethernet data centers and data center interconnects. For PFC to function properly and without wasting memory, the amount of headroom buffer must be calculated. Deployment in large scale data center networks and long-distance interconnects is currently problematic requires manual configuration. **Separately, there are customer requirements for the integrity and confidentiality protection of all frames transmitted between geographically distributed data centers. The current specification is inconsistent and incomplete regarding the operation of PFC and MACSec together.** ~~Separately, it is desirable to protect PFC frames traversing the data-center interconnect. The current specification is inconsistent and unclear regarding the operation of PFC and MACSec together.~~

Individually contributed CSD proposals(1/4)

1.2.1 Broad market potential

Each proposed IEEE 802 LMSC standard shall have broad market potential. At a minimum, address the following areas:

b) Multiple Vendors and Numerous Users.

Most cloud data center providers are using RoCEv2 within their networks, require lossless operation, and currently need to configure PFC manually. RoCEv2 is also used by large enterprises, financial institutions and other high-performance computing environments. Further there is strong interest in accessing new high-speed solid-state data storage technologies over Ethernet networks using RoCEv2.

There are public cloud vendor requirements for MACsec integrity and confidentiality protection of all frames transmitted between geographically distributed data centers.
~~Public cloud vendors that are interconnecting data centers over long distance have expressed interest in have PFC frames protected by MACSec.~~

Individually contributed CSD proposals(4/4)

1.2.4 Technical Feasibility

Each proposed IEEE 802 LMSC standard shall provide evidence that the project is technically feasible within the time frame of the project. At a minimum, address the following items to demonstrate technical feasibility:

a) Demonstrated system feasibility.

The proposed project incorporates techniques for peer-to-peer link delay measurement and information exchange mechanisms that are currently available in many production bridges and end-stations. **PFC operation with MACsec support has been deployed, although existing implementations do not necessarily interoperate.** ~~The proposed shim layer does not define new functionality, but rather further specifies how to map PFC primitives to MAC Service primitives or transparently pass through PFC primitives to MAC Control.~~

b) Proven similar technology via testing, modeling, simulation, etc.

The proposed project enables peer nodes to advertise the new capability through the DCBX mechanism which is widely deployed today using LLDP. Roundtrip delay measurements for participating systems are based on the existing PTP delay measurement mechanism. ~~The proposed shim layer consolidates the existing interfaces for PFC over MACs with and without a MAC Control sublayer.~~