WIKIPEDIA

Data center bridging

Data center bridging (**DCB**) is a set of enhancements to the Ethernet local area network communication protocol for use in data center environments, in particular for use with clustering and storage area networks.

Contents

Motivation

Ethernet is the primary network protocol in data centers for computer-to-computer communications. However, Ethernet is designed to be a best-effort network that may experience packet loss when the network or devices are busy.

In IP networks, transport reliability under the end-to-end principle is the responsibility of the transport protocols, such as the Transmission Control Protocol (TCP). One area of evolution for Ethernet is to add extensions to the existing protocol suite to provide reliability without requiring the complexity of TCP. With the move to 10 Gbit/s and faster transmission rates, there is also a desire for finer granularity in control of bandwidth allocation and to ensure it is used more effectively. These enhancements are particularly important to make Ethernet a more viable transport for storage and server cluster traffic. A primary motivation is the sensitivity of Fibre Channel over Ethernet to frame loss. The higher level goal is to use a single set of Ethernet physical devices or adapters for computers to talk to a Storage Area Network, Local Area network and InfiniBand fabric.[1]

Approach

DCB aims, for selected traffic, to eliminate loss due to queue overflow (sometimes called **lossless Ethernet**) and to be able to allocate bandwidth on links. Essentially, DCB enables, to some extent, the treatment of different priorities as if they were different pipes. To meet these goals new standards are being (or have been) developed that either extend the existing set of Ethernet protocols or emulate the connectivity offered by Ethernet protocols. They are being (or have been) developed respectively by two separate standards bodies:

- The Institute of Electrical and Electronics Engineers (IEEE) Data Center Bridging Task Group of the IEEE 802.1 Working Group
- Internet Engineering Task Force (IETF).

Enabling DCB broadly on arbitrary networks with irregular topologies and without special routing may cause deadlocks, large buffering delays, unfairness and head of line blocking. It was suggested to use DCB to eliminate TCP slow start using approach of **TCP-Bolt**.[2]

Terminology

Different terms have been used to market products based on data center bridging standards:

- **Data Center Ethernet** (DCE) was a term trademarked by Brocade Communications Systems in 2007 but abandoned by request in 2008.[3] DCE referred to Ethernet enhancements for the Data Center Bridging standards, and also including a Layer 2 Multipathing implementation based on the IETF's Transparent Interconnection of Lots of Links (TRILL) standard.[4]
- **Convergence Enhanced Ethernet** or **Converged Enhanced Ethernet** (CEE) was defined from 2008 through January 2009 by group of including Broadcom, Brocade Communications Systems, Cisco Systems, Emulex, HP, IBM, Juniper Networks, QLogic.[5] The ad-hoc group formed to create proposals for enhancements that enable networking protocol convergence over Ethernet, specially Fibre Channel. Proposed specifications to IEEE 802.1 working groups initially included:
    - The Priority-based Flow Control (PFC) Version 0 Specification (http://www.ieee802.org/1/files/public/docs2008/bb-pelissier-pfc-proposal-0508.pdf) was submitted for use in the IEEE 802.1Qbb (http://www.ieee802.org/1/pages/802.1bb.html) project, under the DCB task group of the IEEE 802.1 working group.
    - The Enhanced Transmission Selection (ETS) Version 0 Specification (http://www.ieee802.org/1/files/public/docs2008/az-wadekar-ets-proposal-0608-v1.01.pdf) was submitted for use in the IEEE 802.1Qaz (http://www.ieee802.org/1/pages/802.1az.html) project, under the DCB task group of the IEEE 802.1 working group.
    - The Data Center Bridging eXchange (DCBX) Version 0 Specification (http://www.ieee802.org/1/files/public/docs2008/az-wadekar-dcbx-capability-exchange-discovery-protocol-1108-v1.01.pdf) was also submitted for use in the IEEE 802.1Qaz (http://www.ieee802.org/1/pages/802.1az.html) project.

IEEE task group

The following have been adopted as IEEE standards:

- Priority-based Flow Control (PFC): IEEE 802.1Qbb (http://www.ieee802.org/1/pages/802.1bb.html) provides a link level flow control mechanism that can be controlled independently for each frame priority. The goal of this mechanism is to ensure zero loss under congestion in DCB networks.
- Enhanced Transmission Selection (ETS): IEEE 802.1Qaz (http://www.ieee802.org/1/pages/802.1az.html) provides a common management framework for assignment of bandwidth to frame priorities.
- Congestion Notification: IEEE 802.1Qau (http://www.ieee802.org/1/pages/802.1au.html) provides end to end congestion management for protocols that are capable of transmission rate limiting to avoid frame loss. It is expected to benefit protocols such as TCP that do have native congestion management as it reacts to congestion in a more timely manner.
- Data Center Bridging Capabilities Exchange Protocol (DCBX): a discovery and capability exchange protocol that is used for conveying capabilities and configuration of the above features between neighbors to ensure consistent configuration across the network. This

protocol leverages functionality provided by IEEE 802.1AB (http://www.ieee802.org/1/pages/802.1ab.html) (LLDP). It is actually included in the 802.1az standard.

Other groups

- The IETF TRILL (Transparent Interconnection of Lots of Links) standard provides least cost pair-wise data forwarding without configuration in multi-hop networks with arbitrary topology, safe forwarding even during periods of temporary loops, and support for multipathing of both unicast and multicast traffic. TRILL accomplishes this by using IS-IS (Intermediate System to Intermediate System) link-state routing and by encapsulating traffic using a header that includes a hop count. TRILL supports VLANs and frame priorities. Devices that implement TRILL are called RBridges. RBridges can incrementally replace IEEE 802.1 customer bridges. TRILL Working Group Charter (http://www.ietf.org/dyn/wg/charter/trill-charter.html)
- IEEE 802.1aq specifies shortest path bridging of unicast and multicast Ethernet frames, to calculate multiple active topologies (virtual LANs) that can share learned station location information. Two modes of operation are described, depending on whether the source Bridge is 802.1ad (QinQ) which is known as SPBV or 802.1ah (MACinMAC), which is known as SPBM. SPBV supports a VLAN using a VLAN Identifier (VID) per node to identify the shortest path tree (SPT) associated with that node. SPBM supports a VLAN by using one or more Backbone MAC addresses to identify each node and its associated SPT, and it can support multiple forwarding topologies for load sharing across equal cost trees using a single B-VID per forwarding topology. Both SPBV and SPBM use link-state routing technology. SPBM by virtue of its MACinMAC encapsulation is more suitable for a large data centre than SPBV. 802.1aq defines 16 tunable multipath options as part of the base protocol, with an extensible multipathing mechanism to allow many more multipath variations in the future. 802.1aq supports the dynamic creation of virtual LAN's that interconnect all members with symmetric shortest path routes. The virtual LANs can be deterministically assigned to the different multi paths providing a degree of traffic engineering in addition to multipathing and can grow or shrink with simple membership changes. 802.1aq is fully backward compatible with all 802.1 protocols. 802.1aq became an IEEE standard in April 2012.
- Fibre Channel over Ethernet: T11 FCoE (http://www.t11.org/fcoe) This project utilizes existing Fibre Channel protocols to run on Ethernet to enable servers to have access to Fibre Channel storage via Ethernet. As noted above, one of the drivers behind enhancing Ethernet is to support storage traffic. While iSCSI was available, it depends on TCP/IP and there was a desire to support storage traffic at layer 2. This gave rise to the development of the FCoE protocol, which needed reliable Ethernet transport. The standard was finalized in June 2009 by the ANSI T11 committee.
- IEEE 802.1p/Q provides 8 traffic classes for priority based forwarding.
- IEEE 802.3bd provided a mechanism for link-level per priority pause flow control.

These new protocols required new hardware and software in both the network and the network interface controller. Products were being developed by companies such as Avaya, Brocade, Cisco, Dell, EMC, Emulex, HP, Huawei, IBM, and Qlogic.

References

1. Silvano Gai, *Data Center Networks and Fibre Channel over Ethernet (FCoE)* (Nuova Systems, 2008)
2. Stephens, B.; Cox, A. L.; Singla, A.; Carter, J.; Dixon, C.; Felter, W. (2014-04-01). *Practical DCB for improved data center networks*. *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*. pp. 1824–1832. CiteSeerX 10.1.1.713.2937 (https://citeseerx.is

t.psu.edu/viewdoc/summary?doi=10.1.1.713.2937). doi:10.1109/INFOCOM.2014.6848121 (https://doi.org/10.1109%2FINFOCOM.2014.6848121). ISBN 978-1-4799-3360-0.

3. "Data Center Ethernet" (http://tsdr.uspto.gov/#caseNumber=77287410&caseType=SERIAL_NO). *Trademark serial number 77287410*. US Patent and Trademark Office. Retrieved July 18, 2013.
4. Radia Perlman; et al. (July 2011). *Routing Bridges (RBridges): Base Protocol Specification*. IETF. RFC 6325 (https://tools.ietf.org/html/rfc6325).
5. "cee-authors" (http://tech.groups.yahoo.com/group/cee-authors/). *Yahoo Groups archive*. January 2008 – January 2009. Retrieved October 6, 2011.

Retrieved from "https://en.wikipedia.org/w/index.php?title=Data_center_bridging&oldid=1011013301"

This page was last edited on 8 March 2021, at 15:34 (UTC).