ECMP for SPBM 802.1Qbo

Proposal for PAR and 5 Criteria

Version 1

Peter Ashwood-Smith – Huawei Technologies
Ali Sajassi – Cisco Systems

Project Authorization Request

Title

- PAR for an amendment to existing Standard 802.1Q-2005

- P802.1Qbo

- IEEE Standard for Local and Metropolitan Area Networks---Virtual Bridged Local Area Networks :
  Amendment:
  SPBM - Equal Cost Multiple Paths

Scope

- This standard specifies protocols, procedures and managed objects to support multiple possible next hop choices for SPBM encapsulated frames.

- This is accomplished by defining a family of **standard** hash functions with different L2/L3 input parameters and using the result to pick from multiple next hop choices for a given B-VID/B-DA.

- This is used for known unicast frames only.

- A given B-VID would be configured with one of a small number of known hash functions (consistency would be assured through 802.1aq's ECT-ALGORITHM mechanism).

- There are impacts on 802.1ag as a result of this work so this scope must include CFM enhancements to support ECMP.

- 802.1aq does not require changes to support this new mode of forwarding as the ECT-ALGORITHM framework is already generic and can support any kind of new forwarding modes and their consistency/migration requirements.

Purpose

- .1Qbo provides the capability to use many more ECMP paths than 802.1aq's ECT mechanism due to its constant scale compared to the ECT mechanisms' non linear scale with respect to network diameter.

- .1Qbo would be used in conjunction with an 802.1aq control plane and would use a specific B-VID(s) and new ECT-ALGORITHM(s) as advertised/controlled by 802.1aq IS-IS.

- .1Qbo would use TTL for providing loop mitigation in lieu of loop prevention and a corresponding new ETHERTYPE would identify the a frame format (various proposals being discussed).

- It is expected that both the current 802.1aq ECT and 802.1Qbo ECMP would be used at the same time in the same network (for different traffic/service categories) but differentiated by B-VID and hence ECT-ALGORITHM. Further routing isolation may also be provided by 802.1aq's IS-IS Multi Topology mechanisms.

Need

- The deployment of 802.1aq SPBM mode in networks with considerably more shortest paths than available SPBM ECT Algorithms requires new hash based ECMP forwarding. This will permit some traffic classes to obtain broader traffic spreading in such networks.

- ECT load spreading scales as O(# adjacencies ^ diameter) while ECMP scales as a C, a constant. There are therefore networks with larger diameter and / or numbers of adjacencies that cannot use all links with ECT while ECMP can get some traffic on all links. It is desirable to be able to support both mechanisms. We recognized that neither is considered adequate by itself and neither guarantees 'perfect' fill alone but the sum of the two allows broad use of all links (ECMP) and finer control over a subset of links (ECT).

- TTL allows for simple implementation (and explanation) of ECMP loop prevention for SPBM.

- Competitive pressure from other SDOs protocols, a desire to provide a single solution to avoid yet another encapsulation and a single OA&M in the NGDC.

- Direct customer requests to provide a single protocol and a single data path and OA&M.

- We anticipate that the growing interest in OpenFlow will likely also be able to make use of .1Qbo.

Stakeholders

- Vendors, users, administrators, designers, customers, and owners of Provider Backbone Bridged Networks or of future larger Ethernet networks such as next generation Data Center (NGDC).

- Believe also that stakes are even broader than this due to competitive pressures and potential broad applicability of competing non IEEE solutions.

Other Standards with a Similar Scope

- There are no standards solving this problem for IEEE however LAG comes close. LAG however only operates on a per link basis.

- The IETF however has a TRILL/Rbridge protocol which encapsulates Ethernet in a similar manner to 802.1aq but which supports hashed forwarding to multiple next hops but without an ISID and much smaller addresses.

- The TRILL protocol currently has no OA&M but will evolve to cover OA&M similar to .1ag etc. TRILL will likely also evolve to have some concept of an I-SID putting 802.1aq and TRILL squarely in the same territory.

Five Criteria

Broad Market Potential

– Broad sets of applicability.

  • *The commercial provision of Ethernet services across a Data Center, metropolitan or larger networks is large and growing business. Provider Backbone Networks are a significant part of this market and a required component of the NGDC. .1Qbo will permit the even greater use of these mesh topologies.*

– Multiple vendors and numerous users.

  • *There are other standards bodies trying to address all of these requirements simultaneously instead of simply adding ECMP to Ethernet. This covers most all vendors looking at NGDC solutions and Ethernet Exchange points. Non Ethernet solutions are already appearing in ASIC's and proprietary NPU based switches.*

– Balanced costs (LAN versus attached stations).

  • *This project does not materially alter the existing cost structure of bridged networks. Attached stations would not be aware of the encapsulation and hashing/TTL operations by transit bridges.*

Compatibility

– IEEE 802 defines a family of standards. All standards shall be in conformance with the IEEE 802.1 Architecture, Management and Inter-working documents as follows: 802- Overview and Architecture, 802.1D, 802.1Q and parts of 802.1f. If any variances in conformance emerge, they shall be thoroughly disclosed and reviewed with 802.

- .1Qbo would use the 802.1aq SBPM ECT-Algorithm framework for such forwarding compatibility guarantees.

- .1Qbo however does require CFM (802.1ag) extensions to support ECMP. Providing for well **known hash** algorithms are possible methods to mitigate the CFM issues of determinism however further study is required.

- .1Qbo adds one more reason for a frame to be dropped in transit. As such it does not create by itself any incompatibilities with existing protocols. However it is anticipated that .1Qbo would require a new ETHERTYPE and therefore end to end agreement as to the use of that EHERTYPE would be required. The 802.1aq SPBM ECT-ALGORITHM framework is designed for such compatibility guarantees at the B-VID level and easily can operate without change at the ETHERTYPE/B-VID level… i.e. the ALGORITH specifies behavior and ETHERTYPE.

Compatibility – Cont.

- Each standard in the IEEE 802 family of standards shall include a definition of managed objects that are compatible with systems management standards.

  - *Such a definition will be included.*

Distinct Identity

- Substantially different from other IEEE 802 standards.

  - There currently is only LAG, but it operates on a single link, not network wide.

- One unique solution per problem (not two solutions to a problem).

  - There is currently no general on-data-path solution for loop prevention with ECMP forwarded frames. However there is active research in this area.

- Easy for the document reader to select the relevant specification.

  - *This project will amend only the IEEE 802 standard defining shortest Path Bridging – MAC and Connectivity Fault Management.*

Technical Feasibility

– Demonstrated system feasibility.

- *Hash based ECMP is widely deployed in IP and is well known. The main issue is one of OA&M and we will standardize Ethernet solutions to OA&M issues raised by this new behavior.*

– Proven technology, reasonable testing.

- *Widely deployed and tested with IP. The main issue will be testing OA&M behaviors over all the possible paths.*

– Confidence in reliability.

- *It has wide spread use today with known acceptable reliability.*

Economic Feasibility

- Known cost factors, reliable data.

  - *Minimally this will require either a software upgrade to NPU based Ethernet switches, or in the case of ASIC based devices a new B-VID behavior that mirrors existing 802.1ah with the exception of a hash based choice of possible next hops. There would therefore be a cost upgrade for ASIC based switches.*

- Reasonable cost for performance.

  - *Either the Ethernet ASIC/NPU's will require upgrading, or a completely new encapsulation will be required to address full use of large NGDC meshes. The alternatives will also require all new OA&M. This solution is the least expensive of the two choices.*

- Consideration of installation costs.

  - *Either the Ethernet switches will require upgrading, or a completely new encapsulation will be required with new OA&M. This solution is the least expensive of the two choices.*