Data Center Congestion Control – Where's the best fit in IETF/IRTF?

Paul Congdon (Tallac Networks)

IETF-106, Singapore, November 2019

IETF Note Well

https://www.ietf.org/about/note-well/

This is a reminder of IETF policies in effect on various topics such as patents or code of conduct. It is only meant to point you in the right direction. Exceptions may apply. The IETF's patent policy and the definition of an IETF "contribution" and "participation" are set forth in BCP 79; please read it carefully.

As a reminder:
By participating in the IETF, you agree to follow IETF processes and policies.

If you are aware that any IETF contribution is covered by patents or patent applications that are owned or controlled by you or your sponsor, you must disclose that fact, or not participate in the discussion.

As a participant in or attendee to any IETF activity you acknowledge that written, audio, video, and photographic records of meetings may be made public.

Personal information that you provide to IETF will be handled in accordance with the IETF Privacy Statement.

As a participant or attendee, you agree to work respectfully with other participants; please contact the ombudsteam (https://www.ietf.org/contact/ombudsteam/) if you have questions or concerns about this.

Definitive information is in the documents listed below and other IETF BCPs. For advice, please talk to WG chairs or ADs:

- BCP 9 (Internet Standards Process)
- BCP 25 (Working Group processes)
- BCP 25 (Anti-Harassment Procedures)
- BCP 54 (Code of Conduct)
- BCP 78 (Copyright)
- BCP 79 (Patents, Participation)
- https://www.ietf.org/privacy-policy/ (Privacy Policy)

Some History

- IETF-101
  - Introduced TSVWG and ICCRG to IEEE P802.1Qcz on Congestion Isolation
- IETF-103
  - Joint IETF / IEEE 802 workshop on Data Center Networking including topics on congestion control
- IETF-104
  - Side meeting on Hyperscale HPC/RDMA – 9 attendees – All discussion
- IETF-105
  - Side meeting on Large Scale Data Center HPC/RDMA – 35 attendees
    - Ideas explored/discussed for future research:
      - A new UDP based RDMA transport with a reliability/CC shim
      - Injecting more detailed feedback in packets from switches
      - Distinguishing in-network from incast congestion
      - Speeding up congestion notifications from the network
      - Local fast-response congestion mechanisms in switches
    - Drafts discussed;
      - https://tools.ietf.org/html/draft-zhh-tsvwg-open-architecture-00
      - https://tools.ietf.org/html/draft-yueven-tsvwg-dccm-requirements-00

Where to consider DCN CC Research/New-Work

- ICCRG Charter can be interpreted to include DCN
  - "…The ICCRG may also consider congestion and protocol performance problems in general IP networks, i.e., not only on the global Internet. One example of such IP networks are multi-tenant, heterogeneous datacenters,…"
- Congestion control work is on-going in TSVWG
  - However, nothing particularly DCN focused
- Perhaps a new IRTF group is appropriate
- Let's discuss this and status of contributions in our side-meeting

IETF-106 Questions on Congestion Control
in the HPC/RDMA/AI DataCenter Network

- What is needed from NICs for better CC?
  - An open framework to negotiate capabilities and algorithms – OpenCC
    - https://datatracker.ietf.org/doc/draft-zhuang-tsvwg-open-cc-architecture/
- How can the Network participate?
  - An AI model for parameter tuning
    - https://datatracker.ietf.org/doc/draft-zhuang-tsvwg-ai-ecn-for-dcn
  - Fast feedback from the network
    - https://tools.ietf.org/html/draft-even-iccrg-dc-fast-congestion-00
- Other interesting topics
  - Performance metrics for HPC/RDMA/AI networks – like the KPIs discussed by Neal Cardwell in ICCRG yesterday.

Join us for further discussion

- Non-WG IETF Mailing list rdma-cc-interest@ietf.org
  - Subscribe at: https://www.ietf.org/mailman/listinfo/rdma-cc-interest

- Side Meeting: Tuesday 8:30AM – 9:45AM – VIP-A
  - NOTE on side meetings:
    - Open to all
    - Meeting minutes will be posted to rdma-cc-interest@ietf.org
    - Not under NDA of any form

Agenda

- Welcome – Paul Congdon – 10 mins
- Fast Congestion management for Data Centers – Roni Even – 20 mins
  - https://tools.ietf.org/html/draft-even-iccrg-dc-fast-congestion-00
- An Open Congestion Control Architecture for high performance fabrics - Yan Zhuang – 15 mins
  - https://datatracker.ietf.org/doc/draft-zhuang-tsvwg-open-cc-architecture/
- Artificial Intelligence (AI) based ECN adaptive reconfiguration for datacenter networks - Yan Zhuang – 15 mins
  - https://datatracker.ietf.org/doc/draft-zhuang-tsvwg-ai-ecn-for-dcn
- The impact of mixing TCP and RoCEv2 – Yolanda Yu - 10 mins
- How to move forward – All - 5 mins