Fast Notification for Traffic Engineering and Load Balancing (FANTEL)

Side Meeting@IETF 122

18:30 – 20:00 Tuesday, March 18, 2025

Yingzhen Qu

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)

Background and Previous Activities

Fast Notification for Traffic Engineering and Load Balancing (FANTEL) aims to define mechanisms for real-time notification of network conditions to improve traffic engineering and load balancing in networks.

The Adaptive/Perceptive Routing side meeting held on IETF 121 summarized the observations and existing work:

- Changes of traffic pattern in new scenarios (e.g. AI training)
- Key challenges to the network
- Generalized adaptive/perceptive routing concept and model

This side meeting would be more focusing on the requirements, gaps and possible solutions related to the fast notification mechanisms

Agenda

- **Opening** (5 mins)

Use Case& Requirements

- **Requirements in AI Networks** Pengfei Huo (Byte Dance) (10 mins)
- **FANTEL scenarios in IP MAN** Yongqing Zhu (China Telecom) (10 mins)

Gap Analysis

- **Gap Analysis from the Perspective of Protection and Load Balancing** Rui Zhuang (China Mobile) (10 mins)

Potential Solutions

- **Global Load Balancing (GLB) for AI/ML Fabrics** Jeffery Zhang, Kevin Wang (Juniper) (10 mins)
- **Fast Reroute based on Programmable Data Plane** Shuai Wang (Zhongguancun Lab) (10 mins)
- **Ethereal: Divide and Conquer Network Load Balancing in Large-Scale Distributed Training** Vamsi Addanki (TU-Berlin) (15 mins)

- **Open Discussions** (20 mins)

Meeting Materials

https://github.com/XuesongGeng/IETF-122--Fantel-Side-Meeting

Mailing list: fantel@ietf.org

Problem Statement

- As networks evolve, numerous factors influence how traffic is routed.
- In addition to network topology, other parameters such as node/link congestion, link quality, and cost can significantly affect routing decisions.
- Traditional routing mechanisms may not have real-time awareness of dynamic network conditions.
- Congestion, link degradation, or cost changes can go unnoticed, leading to suboptimal routing decisions and performance degradation.

The Need for Real-Time Visibility/Notifications

- Optimal routing requires more than just topology awareness.
- Real-time, reliable notifications about network conditions are critical for:
  - Detecting congestion hotspots
  - Identifying degraded or high-cost links
  - Supporting adaptive, intelligent path selection
- Desired notifications should be real-time, reliable, scalable and standardized

Call to Action

- Examine existing notification protocols/tools, identify gaps and key information that needs to be notified (e.g., congestion levels, link state changes, node failures).
- Define the scope of notifications—local vs. global, per link/node, real-time event based vs. regular interval notifications.
- Drive standardization of notification mechanisms and develop scalable and real-time notification protocols.