LLM MoE Multicast Discussions

Jeffrey Zhang, HPE

LLM MoE Multicast Side Meeting, IETF123

Summary

- Good to see the discussion of using multicast
- This does not seem to be LLM MoE specific, though
- PIM/RIFT-Bidir and BIER may work well
  - With assistance from the applications

Multicast Applicability

- Any large-scale, high-rate content replication benefits from multicast
  - Not just LLM MoE
  - But the benefits need to outweigh the cost of establishing/maintaining multicast states
- In the case of LLM MoE (or any AIDC scenario, or any scenario)
  - How large are the receiver sets, and how high is the data rate?
    - The following slides assume that the set size and/or data rate warrant the use of multicast
  - How dynamic/scattered are the receiver sets, and how long do they last?
    - The applications can help in this regard

BIER

- BIER does not maintain per-tree state, so it is perfect for scenarios where the receiver sets that may change dynamically or come/go very fast
- BIER may not work well if a receiver set is scattered across a large number of nodes
  - E.g., receivers scattered across thousands of leaf nodes – many copies of data may need to be sent even if the total number is less than the BitStringLength
- If the application can choose the receivers from a set of nodes in the same BIER set, then only one copy is needed

PIM/RIFT-Bidir

- PIM-Bidir is generally considered much simpler and more scalable than PIM-SM
- RIFT multicast is similar to PIM-Bidir
  - With the enhancement of group-prefix-based trees
- Both require the pre-establishment of trees
- The application may be smart enough to pre-establish enough trees, each with a pre-determined set of receivers, and dynamically choose which tree to use for a particular flow
  - E.g. tree1 for expert(1,2), tree2 for expert(3,4), tree3 for expert(5,6), …

Summary

- Multicast is desired if the benefits outweigh the costs
  - In LLM MoE or in general
- Applications can help reduce the impact of the following
  - Scattered receivers – BIER would work well
  - Burstiness of flows – PIM/RIFT-Bidir would work well (if BIER is not feasible)