CSE 222
Graduate Networking

Winter 2001

Lecture 14: Protocol Independent Multicast

Geoffrey M. Voelker

Topic Overview

Three lectures discussing multicast routing:

1. Multicast as an IP layer service
   - Seminal paper by Deering
2. Multicast++
   - Protocol Independent Multicast (PIM)
   - Single-source multicast
3. Reliable Multicast
   - Tough problem, creative solutions

Wide-Area Multicast Routing

- Problem: Traditional multicast routing algorithms do not scale well to the wide-area
  - After deployment and experience with the MBone, it was found that the original DVMRP would not scale well
- Link state multicast (MOSPF)
  - Membership info is broadcast to all routers
  - CPU cost of calculating shortest path trees for all |S,G|
- Distance vector multicast (DVMRP)
  - Initial packets from source are broadcast to entire internet
  - It does prune back to shortest path, but pruning is periodic
  - Routers not on multicast tree still have to handle periodic broadcast packets and the resulting prune messages

Core Based Tree Multicast

- Idea: Use a single tree for each group (all sources)
  - Tree is rooted at a core router
  - Number of trees is |G| instead of |S|*|G|
  - Aside: Should read paper if interested in multicast
- Problems with CBT
  - Data no longer sent along shortest paths
    - » Increases average delay
  - Traffic concentration: Data from all sources to all members travel along same links
    - » Does not exploit parallelism of different trees
  - Max delay of 2*shortest-path, experimental 1.1-1.2

Protocol Independent Multicast (PIM)

- Idea: Why not combine the two approaches?
  - Use single tree for sparse groups (sparse mode)
  - Use shortest path tree for dense groups (dense mode)
- Protocol Independent Multicast
  - Support for different kinds of trees
  - Trees can be intermixed within same group depending upon senders

PIM Requirements

- Paper outlines several requirements for protocol
  - Efficient sparse group support – avoid link and router processing overhead of broadcast
  - High quality data distribution – avoid delay overhead and traffic concentration of single trees when important
  - Routing protocol independent – just use contents of forwarding table, however generated
    - » Works even with policy routing – relies heavily on symmetry still
  - Interoperability – work together with traditional algorithms
  - Robustness – adaptive to routing changes
    - » Soft state, etc. (E2E)

PIM Protocol

- Routers contain entries either for single sources or aggregates
  - (S,G) – matches single source, used for shortest path tree
  - (*,G) – matches all sources, used for single tree
- Key network component in PIM is the rendezvous point (RP)
  - Connects senders and receivers
  - Replaces use of broadcast in RPM (traditional dense mode) for receivers to establish branches of multicast tree
  - An RP can be any PIM router in the network
    - » Which router is a good one to use (topologically)?

Rendezvous Points (RPs)

- Tree growth revolves around the RP
  - Receivers join trees by sending a membership record towards the RP for the group
  - Senders announce themselves by sending registration records and towards the RP for the group
  - RPs therefore become the root of a single tree for all senders and receivers for the group
  - If you want shortest path, you branch out from this tree
- Problem: Both receivers and senders need to know the addresses of RPs for every group
  - Configured (manual tables) or learned (new IGMP message)

RPs and Shared Trees

- Joining a shared tree multicast group
  - Host R announces membership report for G to local network
  - Designated router (DR) on LAN receives report
  - DR looks to see if G maps to an RP
    - » No – use dense mode, propagate join towards sender
  - DR creates a (*,G) entry, propagates a join message towards the RP
  - Upstream routers adjust forwarding tables to ensure that the downstream receiver R will get multicast packets for G
  - Join stops at RP
  - At this point, the receiver is linked to the shared tree
    - » Join can encounter a branch before reaching RP

RPs and Shortest Path Trees

- Shortest path trees are established in two ways
  - When a router sees a host report for group G and it does not have a mapping of G to some RP
  - As a mutation from the shared tree
    - » Data arrives over shared tree until shortest path tree established
  - Not sure what the common case is expected to be (latter?)
- Router performs two actions
  - Sends join message to the best hop towards source S
    - » To extend shortest path tree to receiver
  - Sends prune message towards RP (just for source S)
    - » To remove itself from the shared tree for S
    - » RP will not forward packets from S to the router – but it will forward packets from other sources in the same group to it

PIM Care and Feeding

- Routers
  - Send periodic messages upstream to next-hop to sources (SP trees) and to RPs (shared trees) to refresh soft state
  - Handles state, topology, and membership changes
- RPs
  - Send periodic messages downstream to routers
  - These keep-alives enable routers to detect when the RP is unreachable (has failed)
    - » Routers need to know about other RPs
  - Sources send to all RPs for a group
  - Receivers only connect to one RP, failing over if necessary

Discussion

- Complicated
  - Overall, I thought the architecture described was extremely complicated and would require vast changes to the routing infrastructure of the Internet.
  - Many details to the protocol (operation and messages)
  - Still many open issues
  - Start to get a sense of the challenge of Internet multicast
- Analysis
  - There are experiments showing CBT can perform poorly
  - Why not run experiments for their algorithm?!?
- Shared vs. Shortest
  - I also didn't quite understand how a router knows when to switch between a shared tree and a SPT. Is this done manually or is there a way it can dynamically know when to switch?

Discussion (2)

- RP selection and placement
  - …it appears that every router in the domain "can" be a rendevous point; this implies that every router would now be modified or configured to do the necessary (and complex) bookkeeping to keep track whether a multicast forwarding entry is of the form (S,G) -- for source-specific tree -- or (*,G) -- for shared tree
  - Note: All PIM routers must be able to handle (S,G) or (*,G) whether or not they serve as an RP
- "Shortest path"
  - Though we know that the shortest path need not be the best path and hop count need not be the best metric for path performance, I have been coming across so many papers(including this paper) where the whole algorithm is based on this metric. Why?
  - They relax "shortest" to be whatever is defined in the router forwarding tables (so policy routing is supported). But still depends on symmetry.

For Next Time…

- May skip EXPRESS paper (will decide Thu/Fri)
- Final project info