Anniversary Retrospective: Where *Multicast* Has Been & Where It's Headed

<*dino@procket.com*>

Agenda

- Origins of Multicast
  - Dating back to late 80s
- Requirements from the early 1990s
- Protocol Generation Evolution
- Where we are now
- Where we can go
- Closing Summary

Origins of Multicast

- Link-local multicast addresses date back to 1989
  - RFC 1112 spec'ed out IGMP, 224.0.0.{1,2} were born (but not used at that time)
  - OSPF from day-1 used 224.0.0.{5,6} (used since 1989)
  - ST-II used 224.0.0.{7,8} (used since 1990)

Origins of Multicast

- IGMPv1 implementations started appearing in host stacks in the early 1990s
- Router vendors came next with IGMPv2
- Early routing protocols – 1993 timeframe
  - DVMRP and MOSPF
- Next generation wave of protocols – 1994
  - CBT, PIM-DM, PIM-SM, IGMPv3
- Inter-domain related protocols – 1997
  - BGMP and MSDP
- Auxiliary protocols – 1999 to present
  - MZAP, MSNIP, RMT protocols

Origins of Multicast

- MBONE was a great experiment - 1995
  - Consolidated number of protocols needed in practice
    - IGMPv2, PIM-SM, DVMRP
  - Tunneling was getting out of hand
- Effort by ISPs to go native - 1996
  - Run sparse-mode protocols only
    - IGMPv2, PIM-SM, MSDP
- Brokerage firms and enterprise turn on multicast - 1997
  - Brokerage – Stock trade/quote distribution
  - Enterprise – Desktop conferencing and distance learning

Requirements from the early 90s

- Goals - Applications
  - Desktop Conferencing
  - Distance Learning
  - Brokerage Applications
- Non-goals
  - Resource Discovery
  - Cache Coherency
  - Mother's Day Problem

Requirements from the early 90s

- vat/vic/wb being used in early 90s to distribute IETF working groups
- Content providers interested in reaching very large audience
- Brokerage firms were using UDP broadcast

Requirements from the early 90s

- When IETF was developing PIM and CBT
  - Router state seem to be a technical goal
  - CBT helped with shared-trees only – but could not give low-delay paths
  - PIM had both shared- and source-trees to deal with the low-delay/more-state versus less-delay/less-state tradeoff
  - Lessons learned over time
    - CBT didn't have enough functionality
    - PIM shared-tree to source-tree switch-over too dynamic
- Bursty source issue wasn't known to be a problem or a design goal at the time

Requirements from the early 90s

- Customers didn't want to rev their unicast routing protocols
- Multicast protocols had to be augmented to their configurations
- Needed to work over AS boundaries and IGP redistribution boundaries
- Customers wanted a broadcast mode variant to minimize control message overhead (i.e. dense-mode)

Requirements from the early 90s

- Routing domains didn't want the interworking issues they had with unicast routing protocols
  - No mIGP/mEGP split
  - Single protocol which builds distribution trees across domains and routing protocols
- Transition would be incremental so a unicast protocol that reflected a different "multicast" topology was required

Protocol Generation Evolution

- Started with flood-and-prune protocols
  - DVMRP and PIM-DM
- Couldn't run these across the Internet
  - PIM-SM and CBT could work
- Using shared-trees caused RP distribution issues across the Internet
  - MSDP and BGMP could work

Protocol Generation Evolution

- DVMRP dissolved when workstation based routers were replaced with commercial routers
- PIM-DM was limited to broadcast applications (brokerage firms) but generally not scaling
- PIM-SM worked out because we only needed one tree-distribution protocol
- BGMP was too complex on top of PIM-SM running in a domain
- MSDP was used for source discovery
- MBGP was used for topology non-congruency and multicast-specific policy

Where we are now

- Domains run PIM-SM and manage their own RPs
- Domains discover sources in other domains by running MSDP between their RPs and RPs in MBGP peering domains
- MBGP is used in parallel with BGP to find paths to multicast sources
- This has been coined Any Source Multicast (ASM)

Where we can go

- IGMPv3 implementations are appearing in hosts
- Hosts can join/leave “channels” by specifying (S,G) (“Finding Nemo” at Disney)
- Routers can support source-trees only
- Sources are learned at the application level
- This is coined Single Source Multicast (SSM)

Where we can go

- Hosts can join groups like in ASM
- Routers can forward packets on a bi-directional shared-tree
- Bidir-tree is setup when
  - RP is learned for a group range (sender branches)
  - When group is joined (receiver branches)
- Low delay paths are used from sources to receivers
- This is coined Bidir Multicast

Closing Summary

- We have gone from dynamic switching of tree types to using a single mode per group range
- PIM-SM is now tri-modal
  - 232.0.0.0/8 are SSM groups using source-trees only
  - Bidir-RPs are learned to select which group ranges run in Bidir mode using shared-trees only
  - All other RPs learned run in ASM mode using original sparse-mode PIM definition (building both tree types)
- All 3 modes can run intra-domain or inter-domain

Closing Summary

- For IPv4 Multicast
  - Tri-modal are the options
  - MBGP still a necessity
- For IPv6 Multicast – Dual-modal
  - SSM in intra- and inter-domain
  - Bidir intra-domain and possibly inter-domain
  - MBGP still a necessity
  - Don't need MSDP
  - We finally split control-plane from data-plane
- Multicast gets simpler and therefore more reliable

"10,000 users on 1,500 networks in 30 countries spanning 22 timezones will be terribly disappointed to learn that all the Mbone meetings, seminars, and social events are a mass delusion."

— *Van Jacobson, 1994*