1 # EECS 122, Lecture 14

Kevin Fall

kfall@cs.berkeley.edu

2 # Wide Area Multicast Delivery

- Simple approaches (flooding and ST modifications) don t scale so well
- Two Types of Distribution Techniques:
  - Source-Based Tree Approaches
    - build a distribution tree rooted at each sender
  - Shared Tree Approaches
    - one shared tree per group that hosts must attach to

3 # Source-Based (“shortest path ”) Trees

- Build spanning trees rooted at each sender of each group
- Reverse Path Broadcasting (RPB)
- Truncated Reverse Path Broadcasting (TRPB)
- Reverse Path Multicasting (RPM)

4 # Reverse Path Broadcasting

- Build a *simplex* spanning tree for each potential source [really source subnet]
- Given multiple senders per group, implies a different delivery tree for each source
- RPB Operation
  - for each each received packet, if it was received on the link on the shortest path back to the source, forward to all but the receiving link

5 # Components

6 Reverse Path Forwarding

- Uses both source and destination addresses
- Algorithm:
  - look up source address in routing table
  - compare route entry with receiving interface
  - if wrong interface, drop
  - for each outgoing interface with group members downstream, forward packet

7 Implications of RPF Algorithm

- different tree for each source (source address identifies tree)
- tree is shortest path from source to destination (fastest delivery)
- packets spread over multiple links, leading to better network utilization

8 RPB Operation

9 RPB Operation

10 RPB Operation

11 RPB Operation

12 RPB Operation

13 RPB Operation

14 RPB Operation

15 Observations

- Packet duplication due to flooding

- No use of global topology information or dynamic group membership
- Specialized forwarding algorithm:
  - RPF (reverse-path forwarding)

16 Extending RPB

- Improvement to RPB ("extended RPB"):
  - use routing protocol to detect which of neighbors links are parent links
  - can cut down some packet duplication
- Leaf subnets with no members still see traffic
- How to take advantage of group info?

17 Truncated RPB

- Use IGMP information to determine which leaf subnets contain members
- Routers don t deliver to subnets with no members
- Results:
  - saves some bandwidth on LANs, but does not address duplication across branches of distribution tree

18 Reverse Path Multicasting (RPM)

- Enhance RPB (TRPB) so that tree branches are activated only when necessary
- More precisely, tree branches only serve:
  - subnets with group members
  - routers and subnets along the shortest path to those subnets

19 Pruning the Distribution Tree

- RPM supports notion of "pruning" where routers can send messages indicating they should not receive traffic

- So, use TRPB algorithm for first packet to (source, group) pair
- Leaf routers with no group members send "prune" message on parent link ("broadcast and prune" approach)

20 Router Prune Processing

- (Upstream) routers receiving prune messages store them ("prune state")
- Routers with no local group members and that have received prunes on each child may, in turn, send prune on their parent links
- Cascade of prunes results in only "live" tree being used (to active receivers)

21 Reacting to Change

- Both topology and group members may change over time
- Thus, prune state contains a kill timer, and must be refreshed periodically (example of *soft state*)
- If kill timer expires but more traffic arrives, it is treated anew (using TRPB initially)

22 How Long to Age Prunes?

- Too long: join time unreasonable (prune state keeps data from flowing)
- Too short: back toward TRPB (extra traffic overhead)
- Solution, use longer values [default 2 hours] with prune cancellation messages (*grafts*)

23 ❐ ## Grafts

- Routers discovering new members for groups they have pruned may send graft messages to cancel existing prune state
- Grafts are reliably delivered
  - end nodes can t easily determined if a graft was lost or the sender stopped sending

24 ❐ ## Scaling Issues

- Multicast packets still periodically broadcast to all routers
- Each router must maintain (S,G) routing or prune state
- Biggest problem is when there is sparse membership across big internet (intermediate routers hold state)

25 ❐ ## Alternative Approaches

- approaches to handle multicast groups with "sparse" membership (widely distributed)
- keep state for receivers present rather than the reverse (assume non-membership rather than membership)
- do not keep state for each source

26 ❐ ## Shared Trees

- Idea is to construct a single tree (per group) that each source and receiver attaches to [spanning tree per group]
- Routers maintain only (*,G) state, not (S,G) state, leading to better scaling especially with many

senders

- Do not require periodic flooding

27 ## Limitations

- Leads to traffic concentrations near core set of routers (routers on shared tree)
- May result in suboptimal routing to source

28 ## Perspective

- Dense Mode Protocols
  - bandwidth plentiful, receivers are densely distributed
  - assume membership, correct for mistakes
- Sparse Mode Protocols
  - bandwidth expensive, receivers are sparsely distributed
  - assume not members, require joins

29 ## Routing Protocols (briefly)

- Distance Vector
  - carries (cost, direction) information
  - each node has partial information
  - computation is distributed, relies on intermediates
- Link State
  - multicasts topology information to all routers
  - each router computes shortest paths

30 ## Dense Mode Protocols

- DVMRP (distance vector approach)
  - source-based trees using RPM
  - supports tunnels (IP/IP across unicast nets)
- PIM-DM (distance vector approach)
  - uses unicast routing tables
  - source-based trees using RPM

– does not compute child interfaces

31 ## Dense Mode Protocols

- MOSPF (link-state approach)
  - no tunnels
  - SP trees built on-demand using entire topology database
  - no initial flooding, uses explicit join
  - not RPF based

32 ## Sparse Mode Protocols

- PIM-SM (PIM Sparse Mode)
  - uses unicast routing protocol information
  - routers must explicitly join shared tree
  - use rendezvous points (RP) for sources to meet receivers
  - routers must know RP set for region
  - joins are unicast toward RP
  - can switch dynamically to source-based tree

33 ## Sparse Mode Protocols

- CBT (Core Based Trees)
  - uses unicast routing protocol information
  - never switches to source based trees
  - tree branches are bi-directional, no special unicast encapsulation

34 ## Protocol Characteristics

- Use of unicast routing tables
  - PIM-SM and PIM-DM, CBT
- Link state or distance-vector
  - DV: DVMRP, LS: MOSPF
- Soft (DVMRP, PIM) or hard (CBT) state

- Sparse (PIM-SM, CBT), Dense (DVMRP)

35 Multicast Scope Control

- With multicasting, very easy to send traffic all over
- Would like to limit using scope control:
  - TTL scope: use TTL value in IP packets to limit number of hops traversed
  - Admin scope: allocate certain IP address ranges, and do not forward them

36 TTL Scope Control

- Assign TTL thresholds to each link:
  - if (packet TTL < threshold) drop packet
  - recall no ICMP time exceeded for multicast
- TTL Threshold Conventions:
  - 0: same host, 1: same subnet
  - 15: same site, 63: same region (west coast)
  - 127: worldwide, 191: worldwide, limited bw
  - 255: unlimited scope

37 Expanding Ring Search

- Can use TTL as a basis for searches
- Expanding ring search:
  - 1> start with TTL=1
  - 2> multicast query, await response
  - 3> if no response, increase TTL
  - 4> if TTL reaches max, failure
  - 5> go to step 2

38 Limitation of Expanding Ring Search

- TTL scoping requires "successive containment" property and will not work with overlapping regions

- In addition, routers which discard TTL-expired packets may not be capable of pruning sources, leading to excessive bandwidth consumption

39

Administrative Scoping

- Associate scope limitations with special address ranges (the 239/8 address block)
- Key properties [RFC 2365]:
  - packets addressed to admin scoped addresses do not cross admin boundaries
  - admin scoped multicast addresses are locally assigned; may be re-used in different administrative regions

40

Supporting Admin Scoping

- Routers support per-interface scoped IP multicast boundaries
- Does not forward packets in either direction across boundary
- Senders use admin scoped destination addresses, limiting overall distribution