Lecture 11 (Feb 24, 2004)

Outline
Overlay Multicast

IP Multicast

- No duplicate packets
- Highly efficient bandwidth usage
Key Architectural Decision: Add support for multicast in IP layer

Key Concerns with IP Multicast
- Scalability with number of groups
  - Routers maintain per-group state
  - Aggregation of multicast addresses is complicated
  - Reliability and congestion control is hard
- Deployment is difficult and slow
  - ISP’s reluctant to turn on IP Multicast

Supporting Multicast on the Internet

At which layer should multicast be implemented?
Potential Benefits

- **Scalability**
  - Routers do not maintain per-group state
  - End systems do, but they participate in very few groups
- **Easier to deploy**
- **Potentially simplifies support for higher level functionality**
  - Leverage computation and storage of end systems
  - For example, for buffering packets, transcoding, ACK aggregation
  - Leverage solutions for unicast congestion control and reliability

What is an efficient overlay tree?

- The delay between the source and receivers is small
- Ideally,
  - The number of redundant packets on any physical link is low
  - Every member in the tree has a small degree
  - Degree chosen to reflect bandwidth of connection to Internet

Why is self-organization hard?

- **Dynamic changes in group membership**
  - Members may join and leave dynamically
  - Members may die
- **Limited knowledge of network conditions**
  - Members do not know delay to each other when they join
  - Members probe each other to learn network related information
  - Overlay must self-improve as more information available
- **Dynamic changes in network conditions**
  - Delay between members may vary over time due to congestion

Overlay Multicast Metrics

- **Relative Delay Penalty (or Stretch) at members**
  - Ratio of overlay path length to direct unicast path length
  - Traditional Unicast or IP Multicast: RDP = 1
- **Stress on links**
  - Number of duplicate copies of same data carried by a link
  - Traditional Unicast
    - Stress at source = N (number of group members)
  - IP Multicast: Stress = 1
- **Objective in overlay multicast**
  - Low stress and low RDP
  - Low control overheads
  - Quick recovery from overlay node failures (i.e., when a host suddenly leaves the group, the data delivery path should repair quickly)

End System Multicast

- Each joining host contacts a Rendezvous Point (RP)
- RP informs host of a few other existing group members
- New host chooses a few other such members (hosts) and becomes their peers
  - This is called a mesh
  - Mesh is a virtual overlay graph
- **Multicast routing on the mesh**
  - Run some routing protocol (e.g., Distance Vector) on each host
  - Compute multicast tree distributedly using RPF check
End System Multicast (Example)

**Step 1**
- **Mesh**: Richer overlay that may have cycles and includes all group members
  - Members have low degrees
  - Shortest path delay between any pair of members along mesh is small
- Source rooted shortest delay spanning trees of mesh
- Constructed using Distance Vector and RPF
- **Objectives**
  - Members have low degrees
  - Small delay from source to receivers

**Step 2**

Optimizing Mesh Quality

- Path from Gatech1 to Gatech2 goes via Berk1
  - Because Gatech1-Gatech2 not in mesh
  - Would want the Gatech1-Gatech2 edge to be on the mesh
- ESM has techniques that improve mesh

Recovering from partitions

- Each member in the group sends a periodic heartbeat on the mesh that reaches all other members
- When a member doesn’t get any heartbeat from a set of other members, it can assume that mesh is partitioned
- Partition can happen when one or more members suddenly leave the group
- Recovery: each member from one partition tries to contact a few other members and tries create an edge on the mesh

NICE Application-layer Multicast

Consider a set of hosts (members)

NICE Hierarchy (simplified)

Clusters
- Some specific properties, e.g. members in a cluster are close to each other

NICE Hierarchy (simplified)

One member in each cluster becomes leader, goes to next layer
NICE Hierarchy (simplified)

Layer 1

Leaders form the higher layer

NICE Multicast Data Path

NICE Multicast Data Path