Tech Topic #3

January 27, 2010
Today’s Objectives

- Multicast

- Course announcement
  - Over the next two days will be updating the reading list
Where the Replication Happens

• At the source
  – Then it is unicast

• At routers in the network
  – “Native” multicast

• At network access points using replication boxes
  – CDNs, or
  – Some kind of hierarchical replication

• At end points
  – Application layer multicast
Reasons to Study Multicast

• Within the context of where replication can occur, it is one of the possible options

• An interesting academic effort to solve a problem, over and over and over again

• If widespread multicast deployment has failed, why?
  – What is the relationship between routing algorithms and what is adopted?

• Touches on a greater tension between support in the network and functionality only at the edges
Multicast Origins

• Original proposal was to use the options field and put multiple unicast addresses in the header

• The first real proposal for multicast was mostly a LAN-based multicast and limited bridging between LANs
  – Fairly straightforward since most LANs easily support broadcast
  – Challenge was getting LAN entities to pay attention to transmission
  – Solved by using special MAC addresses and dynamically assuming multiple/different MAC address identities
  – Bridging had one member of local LAN communicate multicast frames across multiple hops to remote LAN
    • Two end points formed a tunnel and used IP encapsulation
  – Wanted to apply the same concept at Layer 3
Next Steps

• Expanding to the rest of the Internet was based on a similar concept

• The idea was to have locally-enabled multicast clouds that were connected together by tunnels
  – Consider the network topology of such a deployment
  – Consider the kind of daemon necessary to connect tunnel end-points
  – Consider what functionality was necessary

• Eventually there would be support in routers to perform the same functions
  – Consider why such functionality did not instantly exist
Basic Protocol Mechanisms

• Addressing Basics
  – Use the same kind of “dynamic assumption of identity” as for MAC addresses (or now: DHCP)
  – Remember that a host can have multiple IP addrs

• IP Multicast Addrs
  – Class D range: 224.0.0.0-239.255.255.255 (224/4)
  – Every “multicast-capable” entity (router, replicator, host) knows about Class D addresses and treats packets differently

• Routing and Forwarding
  – Takes on slightly different meaning in multicast
Routing and Forwarding: Unicast

- Routing: process of learning all of the possible paths between sources and destinations
  - Routing Information Base (RIB) holds set of possible routes

- Choosing the best next-hop to a particular destination
  - Forms the entries in the Forwarding Information Base (FIB)

- When packets arrive, FIB is checked, outgoing interface is selected
Routing and Forwarding: Multicast

• Routing: process of learning all possible paths from receivers to sources
  – Basically the same as unicast
  – RIB hold sets of possible routes (may be special protocol or may just use the existing unicast RIB)

• NEW: when receivers join a group, they send a request towards the source(s)
  – Lets network know host has taken on new identity
  – Forwarding state is created based the interface on which the request came in and the next hop towards the source
  – A reverse path is created

• When packets arrive, reverse path is first checked
  – multicast have come in on the interface that a packet sent to the source would have gone out on
  – Then FIB is used to select the outgoing interface
Routing and Forwarding: Multicast

• The process just described skips a few evolutionary steps

• First was DVMRP (Distance Vector Multicast Routing Protocol)
  – It was a “broadcast-and-prune”: transmit everywhere and then have tunnels say they weren’t interested in traffic
  – VERY unscalable
  – A few others proposed along the way

• Other was PIM (Protocol Independent Multicast)
  – “Independent” because it relied on unicast RIB
  – Two types (well, now three types)
    • “dense mode”: does broadcast-and-prune (assumes dense interest)
    • “sparse mode”: rendezvous point (RPs) for receivers to learn about sources
    • “source specific mode”: basically what was just described
The Details

• The details can get messy!

• A separate protocol for hosts to communicate to routers
  – Why?
  – Internet Group Management Protocol (IGMP)
    • Three versions
  – Version for IPv6: Multicast Listener Discovery (MLD)
    • Two versions
  – Also a challenge of dealing with switches

• Lots of different ways of doing multicast routing
  – Most are one of the three types
Broadcast-and-Prune
Step #1: Broadcast-and-Prune
Step #2: Pruning

[Diagram showing the process of pruning with nodes labeled 's', 'g', and arrows indicating the flow and directions.]
Steady State
Grafting on New Receivers
Sparse Mode
Shared Trees
RP-Based
Any Source Multicast
PIM Sparse Mode: RPs

Shared tree after R1, R2, R3 join

Join message toward RP
Sources Send to RP

unicast encapsulated data packet to RP

RP decapsulates, forwards down

Shared tree

R1

R2

R3

R4
Build Tree Back to Sources
Traffic Flows: Switch to SPT

- Join messages toward S2
- Shared tree
Steady State
Source Specific Multicast
Single Source Multicast
Traffic Flows: Switch to SPT

Join messages to known source
Steady State

Distribution tree
Inter-Domain Multicast

• So far, most of what we’ve talked about is how multicast works within a domain
  – Inter-domain requires modifications to BGP
    • Luckily already existed as BGP-4+ (multiprotocol extensions: MBGP)
  – Basic idea: use “prefix descriptor” that identifies whether advertised route is for unicast, multicast, or both
  – Remember, what is the role of an advertised route?

• Notes
  – Multicast was originally run as a flat overlay network
    • DVMRP didn’t distinguish between domains
  – “Sparse mode” required a particularly ugly kludge (MSDP)
  – Some throw-out-the-kitchen-sink alternatives
    • BGMP was the most popular
  – Simple is always, always better when talking about the core
Native Multicast Weaknesses

• All native multicast is UDP
  – Can’t run standard TCP
    • Reliable multicast is hard
  – Congestion control is hard too
    • Not having it is worse
  – A lot of UDP is blocked

• Having “source discovery” in the network was bad
  – It was the dominate way to do multicast for a long time (PIM-SM and MSDP)

• Multicast address allocation was never solved
Native Multicast Weaknesses

• Limited deployment
  – Plan was to support incremental deployment
  – Islands of connectivity connected by tunnels
  – Over time islands would grow in size

• Deployment was sloppy
  – See “Multicast Routing Instabilities” Paper

• When we talk about adoption and deployment, motivation to deploy becomes an issue
  – Little incentive for ISPs to deploy multicast
  – Limited economic model to deploy multicast
Full Circle

• If deployment is a challenge, implement multicast without requiring any interior network changes

• Deploy all functionality at the edges
  – Hence, application layer multicast
  – Builds overlay network

• But this technique has some weaknesses
  – They become important metrics
  – Stress: copies of packets on a link
  – Stretch: quality of path between overlay nodes
  – Overhead: communicating info
Possible Differences
Full Circle

• If deployment is a challenge, implement multicast without requiring any interior network changes

• Deploy all functionality at the edges
  – Hence, application layer multicast
  – Builds overlay network

• But this technique has some weaknesses
  – They become important metrics
  – Stress: copies of packets on a link
  – Stretch: quality of path between overlay nodes
  – Overhead: communicating info
ALM Algorithms

- Mesh-First
- Tree-First
- Implicit
ALM Protocols

• Protocol performance depends heavily on parameters
  – How many nodes are sources
  – How large the streams are
  – How dynamic network conditions are

• No single protocol (or class of protocols) performs best in all situations
  – Leads to runaway number of papers on the topic
  – This paper was an attempt to bring some organization

• Can create an endless supply of papers that:
  – Suggest one set of parameters is more important
  – Develop a protocol that does better than another protocol for that set of parameters
    • Not necessarily the “best” other protocol
    • Not necessarily offering a protocol with the “best” performance
Reasons to Study Multicast

• Within the context of where replication can occur, it is one of the possible options

• An interesting academic effort to solve a problem, over and over and over again

• If widespread multicast deployment has failed, why?
  – What is the relationship between routing algorithms and what is adopted?

• Touches on a greater tension between support in the network and functionality only at the edges