CSE 123A
Computer Networks
Winter 2005
Lecture 11
Internet Routing:
Inter-domain routing
Homework #2 + Misc

- Will be up on the Web page by the end of today
- Due next Tuesday (at the beginning of class)
- Midterm solutions on the Web page
  - We’re grading as fast as we can
Last few classes....

- Routing: how to get packets to their destination

- Intra-domain routing protocols
  - Single network
  - Calculate shortest path between all nodes (for some metric)
    » Distance Vector: Local exchange of global info
    » Link-State: Global exchange of local info
But the Internet is not just one network...

- Inter-domain versus intra-domain routing

![Diagram showing Internet routing concepts]

You at work — Large corporation — Backbone service provider — “Consumer” ISP

You at home — Small corporation — Peering point

“Consumer” ISP

But the Internet is not just one network...
Historic context

- Original ARPAnet had single routing protocol
  - Dynamic DV scheme, replaced with static metric LS algorithm
- New networks came on the scene
  - NSFnet, CSnet, DDN, etc...
  - The total number of nodes was growing exponentially
  - With their own routing protocols (RIP, Hello, ISIS)
  - And their own rules (e.g. NSF AUP)
- **Scalability**: Routing tables with millions of entries?
- **Heterogeneity**: Network A uses hop count as a metric, Network B uses measured delay, Network C uses link capacity; what if networks use different routing protocols?
- **Policy**: Network A connects to Networks B and C. Network B is only allowed to carry network C’s traffic?
Solution: Inter-domain routing

- Separate routing inside a domain from routing between domains
  - Inside a domain use traditional interior gateway protocols (RIP, OSPF, etc)
  - Between domains use Exterior Gateway Protocols (EGPs)
    » Only exchange reachability information (no metrics)
    » Decide what to do based on local policy

- What is a domain?
**Terminology**

- **Autonomous Systems (ASs)**
  - Unit of abstraction in interdomain routing; another word for domain
  - Roughly, a network with common administrative control, a coherent internal routing policy, and presenting a **consistent** external view of connectivity
  - Represented by a 16-bit number
    - Example: UUnet (701), Sprint (1239), UCSD (7377)
Inter-Domain Routing

- Network comprised of many Autonomous Systems (ASes) or domains
- To scale, use hierarchy: separate inter-domain and intra-domain routing
- Also called interior vs exterior gateway protocols (IGP/EGP)
  - IGP = RIP, OSPF
  - EGP = EGP, BGP
Inter-Domain Routing

- Border routers summarize and advertise internal routes to external neighbors and vice-versa.
- Border routers apply polic
- Internal routers can use notion of default routes.
- Core is “default-free”; routers must have a route to all networks in the world.
Exterior Gateway Protocol

- First major inter-domain routing protocol
- Spanning tree: no loops
Problems with EGP

- In 1995 NSFnet got out of the backbone business
  - Many backbones (MCI, Sprint, AT&T…)
  - Multi-connected regional networks
  - Meshed topology, loops…

- A tree-based structure didn’t work anymore
- Need a new protocol…
What kind of protocol?

- Link state?
  - Too much state
    - Currently 11,000 ASs and > 100,000 networks
  - Relies on global metric & policy

- Distance vector?
  - May not converge; loops
  - Relies on global metric and policy

- Solution: path vector
  - Reachability protocol, no metrics
  - Route selection based on local policy
  - Route advertisements carry list of ASs
    - “I can reach UCSD through this path: AS73, AS703, AS1”
    - Automatic loop detection. Why? How?
Path Vectors

- Similar to distance vector, except send entire paths
  - e.g. AS 321 gets route [7, 12, 44] from AS 7 (similar DV route to 44 with cost 2)
  - Strong loop avoidance
  - Supports policies (later)
- Modulo policy, shorter paths are chosen in preference to longer ones
- Reachability only – no metrics
Policies

- Choice of routes may depend on owner, cost, AUP (acceptable use policy), …
  - Business considerations (more on this later)
- Local policy dictates what route will be chosen and what routes will be advertised!
  - e.g., X doesn’t provide transit for B, or A prefers not to use X
How BGP operates (roughly)

Establish session on TCP port 179

Exchange all active routes

Exchange incremental updates

While connection is ALIVE exchange route UPDATE messages

AS1
BGP session
AS2
Two types of BGP neighbor relationships

- External Neighbor (eBGP) in a different Autonomous Systems
- Internal Neighbor (iBGP) in the same Autonomous System

Why do we need iBGP?
iBGP keeps eBGP consistent

- iBGP is needed to avoid routing loops within an AS
- Need all routers to agree on routing policy for external routes
- Existing IGPs (like OSPF) can’t handle scale of all Internet routes

iBGP neighbors do not announce routes received via iBGP to other iBGP neighbors.
Important BGP attributes

- **Local pref**: Statically configured ranking of routes within AS
- **AS path**: ASs the announcement traversed
- **Origin**: Route came from IGP or EGP
- **Multi Exit Discriminator**: preference for where to exit network
- **Community**: opaque data used for inter-ISP policy
- **Next-hop**: where the route was heard from
**BGP Decision process**

- Default decision for route selection
  - Highest local pref, shortest AS path, lowest MED, prefer eBGP over iBGP, lowest IGP cost, router id
- Many policies built on default decision process, but…
  - Possible to create arbitrary policies in principal
    - Any criteria: BGP attributes, source address, prime number of bytes in message, …
    - Can have separate policy for inbound routes, installed routes and outbound routes
  - Limited only by power of vendor-specific routing language
Example: local pref

Higher Local preference values are more preferred

Local preference only used in iBGP

local pref = 80
local pref = 90
local pref = 100
Example: AS Path

Shorter AS Paths are more preferred

128.2/16 9

AS701
UUnet

128.2/16 9 701

AS73
Univ of Wash

AS9
CMU (128.2/16)

128.2/16 9

AS7018
AT&T

128.2/16 9 7018

AS1239
Sprint

128.2/16 9 7018 1239
Shortest AS path doesn’t mean best path

Mr. BGP says that path 4 1 is better than path 3 2 1.
**Example: Using IGP cost for Hot potato routing**

This Router has two BGP routes to 192.44.78.0/24.

**Hot potato: get traffic off of your network as Soon as possible. Go for egress 1!**
Problems with hot potato

Many customers want their provider to carry the bits!

High bandwidth Provider backbone

Low bandwidth customer backbone

SFF

NYC

San Diego

Heavy Content Web Farm

28
17
15
56

... tiny http request

... huge http reply
Ongoing Problems w/BGP

- Instability
  - Route flapping
  - Long AS-path decision criteria defaults to DV-like behavior (bouncing)
  - Not guaranteed to converge, NP-hard to tell if it does
- Scalability still a problem
  - ~100,000 network prefixes in default-free table today
  - Tension: Want to manage traffic to very specific networks (eg. multihomed content providers) but also want to aggregate information.
- Performance
  - Non-optimal, doesn’t balance load across paths
- Security…
Routing policy

- So far we’ve discussed mechanism…
- How and why are basic routing policies decided?
History

- First policies for political reasons
  - NSFnet AUP (even today Internet2)
- Emergence of commercial policies
  - 1994-1995 NSFnet transition
    - NSF ceases to run Internet backbone
    - Commercial carrier (MCI, Sprint, ANS) start selling IP backbone service
    - Interconnected with each other and regional networks at several public NAPs
    - Everyone talks to everyone
  - Then five years went by…
Background – Settlement

- The telephone world
  - LECs (local exchange carriers)
  - IXCs (inter-exchange carriers)
- LECs MUST provide IXCs access to customers; regulation
- When a call goes from one phone company to another:
  - Call billed to the caller
  - The money is split up among the phone systems – this is called “settlement”
On the Internet...

- No regulation
  - One ISP doesn’t have to talk to another

- Founded on “shared goodwill”
  - Pay for connectivity, not per packet
  - Not clear who should pay anyway

- No standard settlement
Peering vs Transit

- Peering
  - Two ISPs provide connectivity to each others customers (traditionally for free)
  - Non-transitive relationship

- Transit
  - One ISP provides connectivity to every place it knows about (usually for money)
Example: peering
Example: transit

By EastNet purchasing transit, Eastnet is announced by USNet to USNet peering and transit interconnections alike.
Example: transit (2)

Thousands of other Int’l ISPs

WestNet

USNet

EastNet

Peering

Transit

The entire Internet as known by USNet
The value of transit

- Not just paying for the fiber, but the connectivity
  - Remember, there is no single “backbone”
  - If you’re an ISP, how do your customers get to yahoo.com?

- Means big ISPs have more value to offer small ISPs than vice-versa
Aside...

- Peering and transit are really two popular points on a continuum
- Some places sell “partial transit”
- Other places sell “usage-based” peering
- Principle issue is:
  - Which routes do you give away and which do you sell? To whom? Under what conditions?
Summary

- Interdomain-routing
  - Exchange reachability information (plus hints)
  - Local policy to decide which path to follow

- Traffic exchange policies are a big issue $$$
  - Complicated by lack of compelling economic model (who creates value?)
  - Can have significant impact on performance
For next time...

- Multicast and multicast routing...
- Chapter 4.4