

CSE 123b Communications Software

Spring 2003

Lecture 8: Inter-domain routing

Stefan Savage

Some slides courtesy Time Griffin

Last two classes

- Intra-domain routing protocols
 - Also called Interior Gateway Protocols (IGP)
 - Create forwarding tables at each router so packets take **best** path from any source to destination
 - » Best path: shortest path in terms of metric (hops, cost, etc)
 - Distance Vector
 - » Local exchange of global topology information
 - » Bellman-Ford distributed shortest path algorithm
 - » Problems during failures: count-to-infinity
 - Link-state
 - » Global exchange of local topology information
 - » Dijkstra's algorithm
 - » Complicated to implement; esp during failure

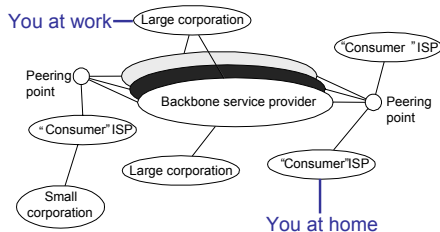
April 28, 2003

CSE 123b -- Lecture 8 -- Inter-domain Routing

2

Key issue: The Internet is not just one network...

- Inter-domain versus intra-domain routing



April 28, 2003

CSE 123b -- Lecture 8 -- Inter-domain Routing

3

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?

April 28, 2003

CSE 123b -- Lecture 8 -- Inter-domain Routing

4

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
- 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)

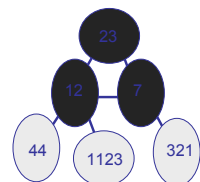
April 28, 2003

CSE 123b -- Lecture 8 -- Inter-domain Routing

5

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



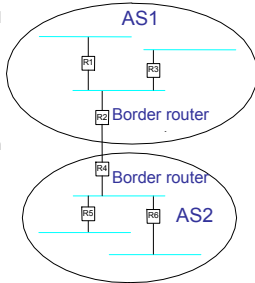
April 28, 2003

CSE 123b -- Lecture 8 -- Inter-domain Routing

6

Inter-Domain Routing

- Border routers summarize and advertise internal routes to external neighbors and vice-versa
- Border routers apply policy
- Internal routers can use notion of default routes
- Core is "default-free"; routers must have a route to all networks in the world



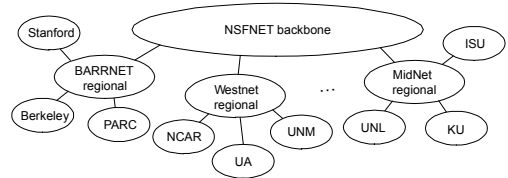
April 28, 2003

CSE 123b -- Lecture 8 -- Inter-domain Routing

7

Exterior Gateway Protocol

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



April 28, 2003

CSE 123b -- Lecture 8 -- Inter-domain Routing

8

Problems with EGP

- In 1995 NSFnet got out of the backbone business
 - Many backbones (MCI, Sprint, AT&T...)
 - Multiconnected regional networks
 - Meshed topology, loops...
- A tree-based structure didn't work anymore
- Need a new protocol...

April 28, 2003

CSE 123b -- Lecture 8 -- Inter-domain Routing

9

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?

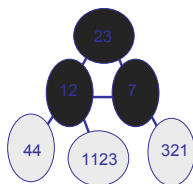
April 28, 2003

CSE 123b -- Lecture 8 -- Inter-domain Routing

10

Path Vectors

- Similar to distance vector, except send entire paths
 - e.g. 321 hears [7,12,44]
 - stronger avoidance of loops
 - supports policies (later)
- Modulo policy, shorter paths are chosen in preference to longer ones
- Reachability only – no metrics



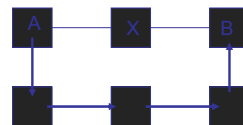
April 28, 2003

CSE 123b -- Lecture 8 -- Inter-domain Routing

11

Policies

- Choice of routes may depend on owner, cost, AUP, ...
 - 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

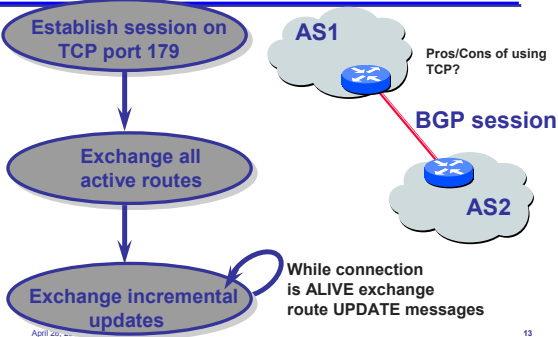


April 28, 2003

CSE 123b -- Lecture 8 -- Inter-domain Routing

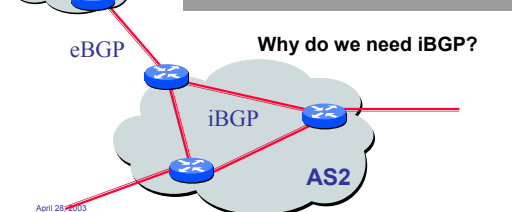
12

How BGP operates (roughly)

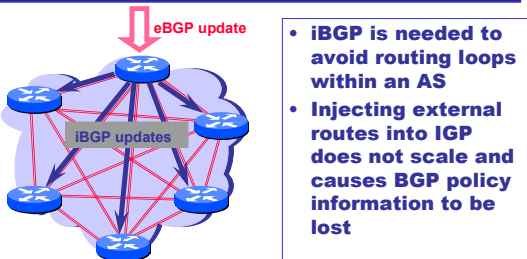


Two types of BGP neighbor relationships

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



iBGP keeps eBGP consistent



iBGP neighbors do not announce routes received via iBGP to other iBGP neighbors.

April 28, 2003

15

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
- **Community:** opaque data used for inter-ISP policy
- **Next-hop:** where the route was heard from

April 28, 2003

CSE 123b -- Lecture 8 -- Inter-domain Routing

16

BGP Decision process

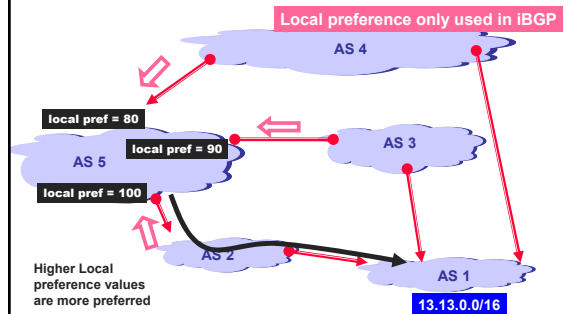
- Default decision for route selection
 - Highest local pref, shortest AS path, lowest MED, prefer eBGP over iBGP, lowest IGP cost, router id
- Many policies built on default decision process, but...
 - Possible to create arbitrary policies
 - » Any criteria: BGP attributes, source address, port # is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

April 28, 2003

CSE 123b -- Lecture 8 -- Inter-domain Routing

17

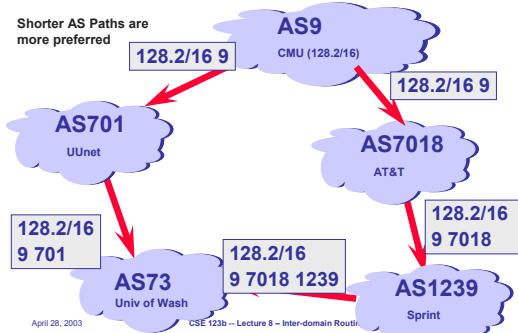
Example: local pref



18

Example: AS Path

Shorter AS Paths are more preferred

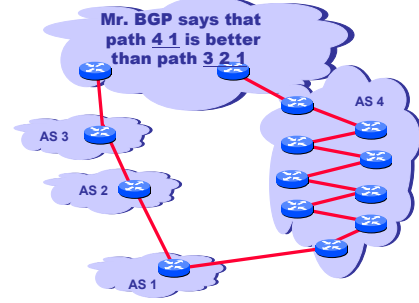


April 28, 2003

CSE 123b -- Lecture 8 -- Inter-domain Routing

19

Shortest AS path doesn't mean best path

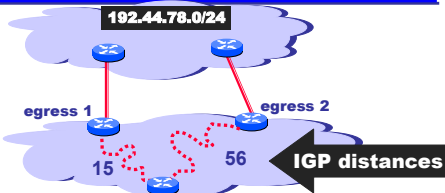


April 28, 2003

CSE 123b -- Lecture 8 -- Inter-domain Routing

20

Example: Using IGP cost for Hot potato routing

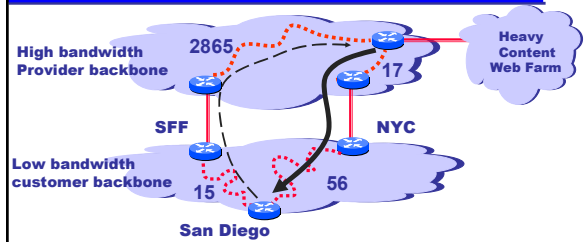


This Router has two BGP routes to 192.44.78.0/24.

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

21

Problems with hot potato



Many customers want their provider to carry the bits!

--- tiny http request
 ——— huge http reply

April 28, 2003

22

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...

April 28, 2003

CSE 123b -- Lecture 8 -- Inter-domain Routing

23

Routing policy

- So far we've discussed mechanism...
- How and why are basic routing policies decided?

April 28, 2003

CSE 123b -- Lecture 8 -- Inter-domain Routing

24

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”

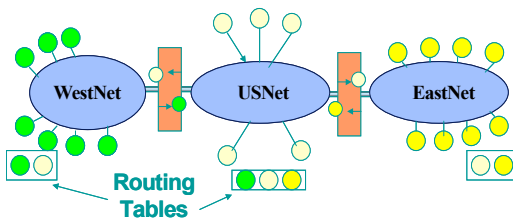
On the Internet...

- 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

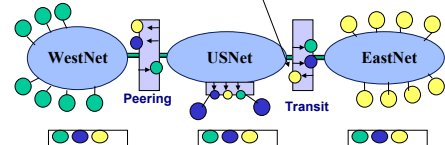
- Peering
 - Two ISPs provide connectivity to each others customers (traditionally for free)
 - Non-transitive relationship
- Transit
 - One ISP provides connectivity to every place it knows about (usually for money)

Example: peering

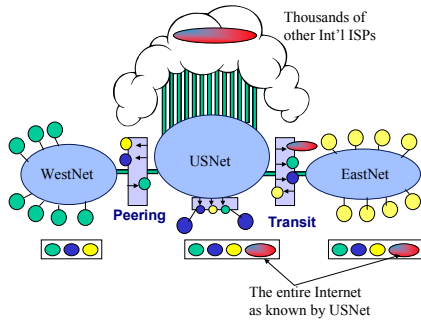


Example: transit

By EastNet purchasing transit, Eastnet is announced by USNet to USNet peering and transit interconnections alike.



Example: transit (2)



April

31

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

April 28, 2003

CSE 123b -- Lecture 8 -- Inter-domain Routing

32

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?

April 28, 2003

CSE 123b -- Lecture 8 -- Inter-domain Routing

33

Summary

- Link-state intra-domain routing
 - Tell everyone about your neighbors
 - Low message overhead, good convergence
 - Must maintain lots of state
- 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

April 28, 2003

CSE 123b -- Lecture 8 -- Inter-domain Routing

34

For next time...

- Multicast routing...
- Chapter 4.2.5 and 4.4

April 28, 2003

CSE 123b -- Lecture 8 -- Inter-domain Routing

35