

Inter-domain Routing

CSE 561 Lecture 6, Spring 2002.
David Wetherall

Overview

- Inter-domain routing
 - BGP mechanics of route selection
 - ISP policy considerations
- Traffic engineering
 - Intra-ISP:
 - ARPANET dynamic metrics (Khanna & Zinky)
 - Traffic demand models and static costs
 - Circuits and MPLS
 - Inter-ISP
 - AS pre-pending, MEDs, community signaling (complex)
- Credit and thanks to Tim Griffin for slide material.

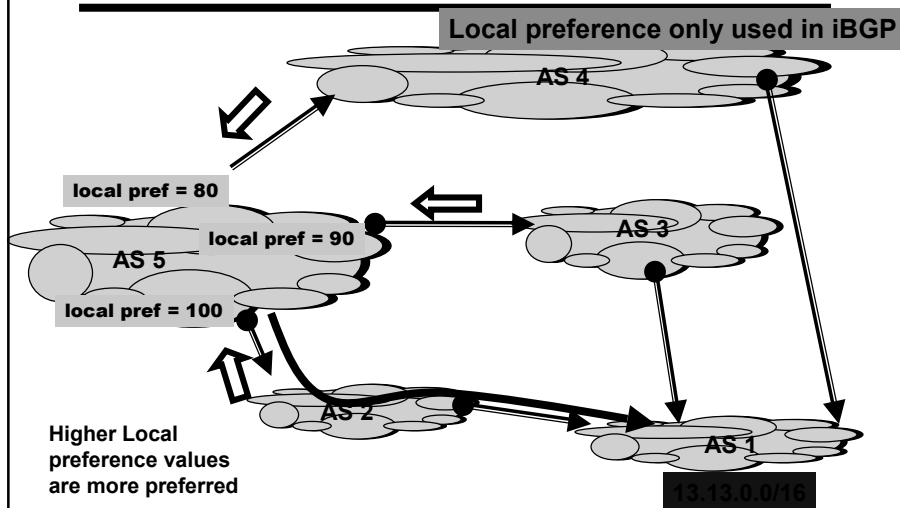
Mechanism: Important BGP attributes

- Associated with an announcement and used as part of the route selection process
- **Local pref:** Statically configured ranking of routes within AS
- **AS path:** ASNs 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

djw // CSE 561, Spring 2002, with credit to savage

L6.3

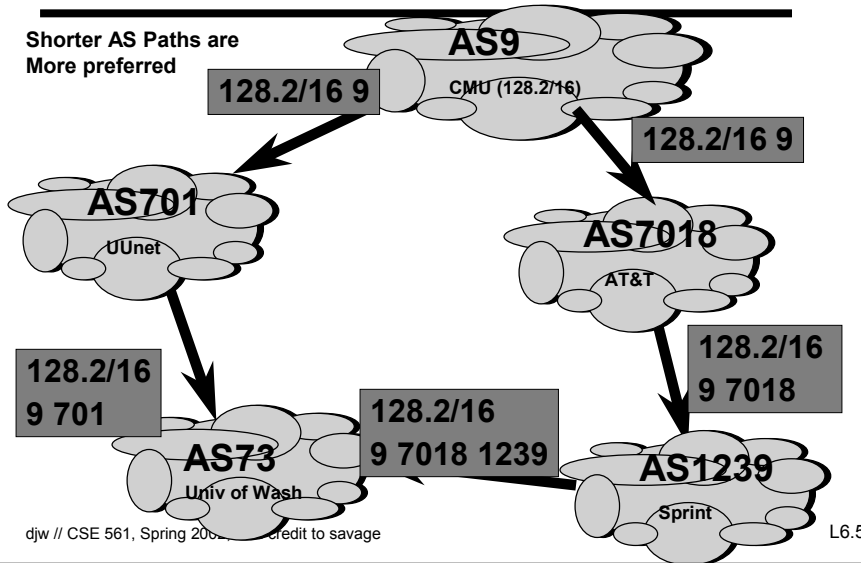
Example: local pref



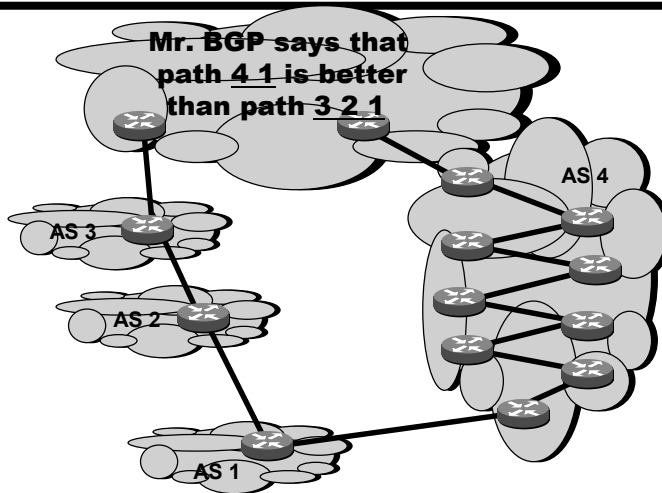
L6.4

Example: AS Path

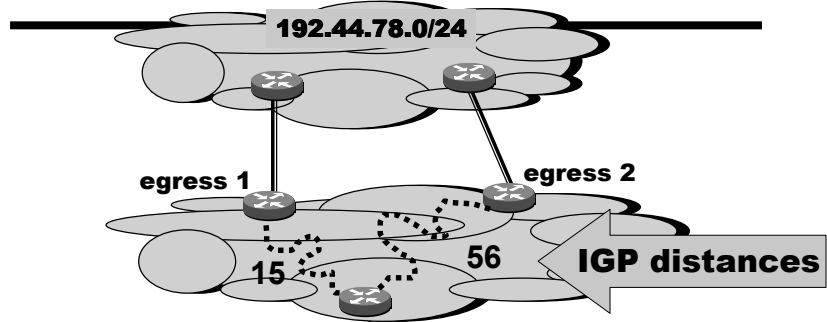
Shorter AS Paths are
More preferred



Shortest AS path doesn't mean best path



Example: Using IGP cost for Hot potato (early-exit) routing

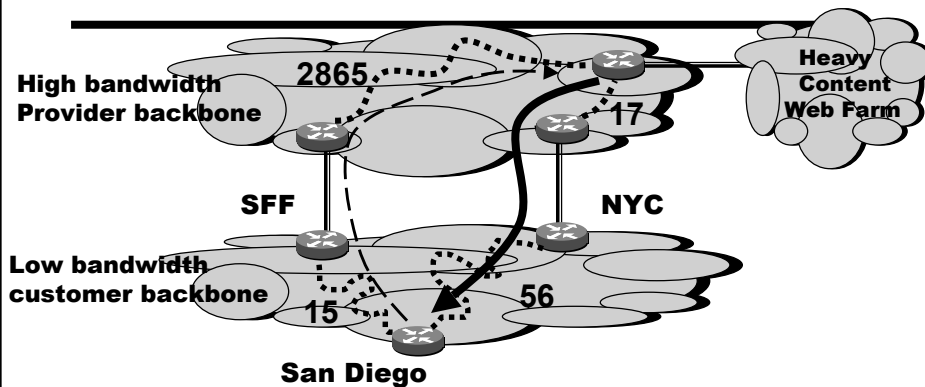


This Router has two BGP routes to 192.44.78.0/24.

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

L6.7

Problems with hot potato



Many customers want their provider to carry the bits!

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

L6.8

BGP Decision process

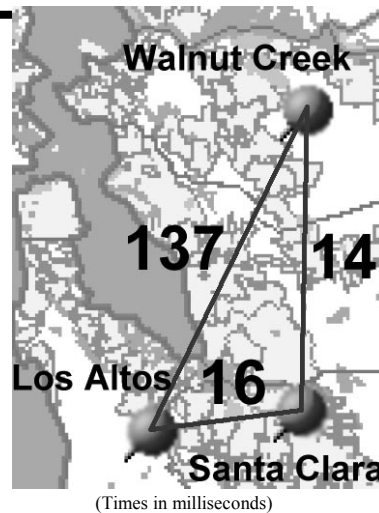
- Default decision for route selection
 - Highest local pref, shortest AS path, lowest MED, prefer eBGP over iBGP, lowest IGP cost, router id
- Many policies built on default decision process, but...
 - Possible to create arbitrary policies
 - Any criteria: BGP attributes, source address, port # is prime,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

djw // CSE 561, Spring 2002, with credit to savage

L6.9

BGP+policy is not shortest path

- Measured round-trip times between sites
- Pythagoras would have wept



djw // CSE 561, Spring 2002, with credit to savage

L6.10

General 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**
 - ~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.

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...

djw // CSE 561, Spring 2002, with credit to savage

L6.13

Background – Settlement

- The telephone world
 - LECs (local exchange carriers)
 - IXC (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”

djw // CSE 561, Spring 2002, with credit to savage

L6.14

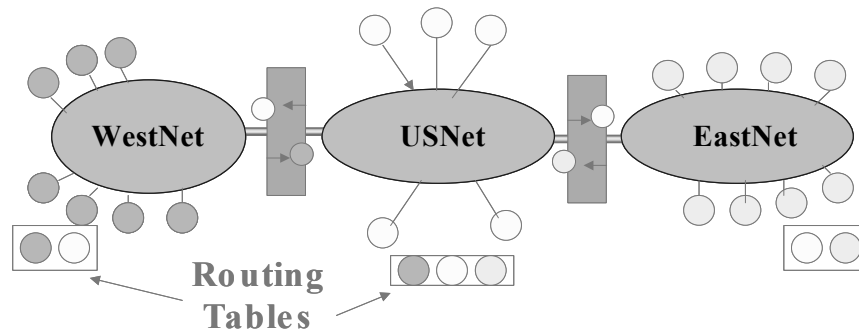
On the Internet...

- 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

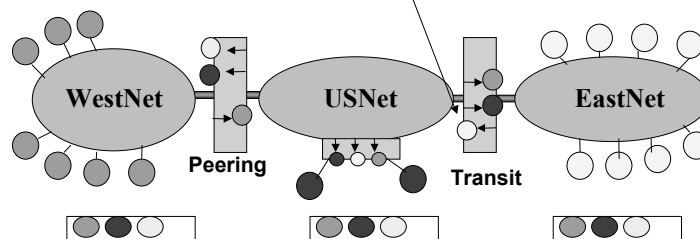


djw // CSE 561, Spring 2002, with credit to savage

L6.17

Example: transit

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



djw // CSE 561, Spring 2002, with credit to savage

L6.18

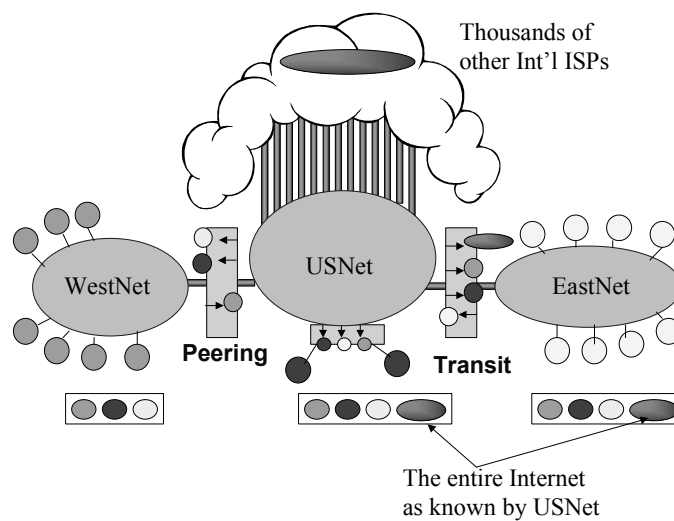
The value of transit

- 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

djw // CSE 561, Spring 2002, with credit to savage

L6.19

The value of transit (2)



djw // C

L6.20

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?

Terminology 101: What's a Tier-1 ISP?

- Simplistic definition:
 - ISP big enough that they don't have to buy transit
 - AT&T, Sprint, Uunet, Genuity, etc.
- Tier-2 buy transit from Tier-1, etc.

- Increasingly worthless terms
 - Everyone claims to be Tier-1
 - More complicated forms of settlement
 - Leverage depends on business model

Terminology 101: Public vs private peering

- **Public peering**
 - Connection via shared switch or network at “public” exchange point (place anyone can be if they pay money)
 - Still negotiated bilaterally
- **Private peering**
 - Private point-to-point link between peers

djw // CSE 561, Spring 2002, with credit to savage

L6.23

Why peer?

- **Transit is very expensive**
 - Was \$150,000 for an OC3 (155Mbps) transit link
- **Peering with other ISPs can reduce the amount of traffic sent on transit link**
 - Also lower latency?
- **Communication patterns aren't uniform**
 - More of your traffic is exchanged with some networks than others
 - Try to peer with other ISPs whose customers exchange traffic frequently with your customers...

djw // CSE 561, Spring 2002, with credit to savage

L6.24

Why not peer?

- Traffic asymmetry
 - More traffic goes one way than the other
 - Peer who carries more traffic feels cheated
- Hassle
- Top tier (big) ISPs have no interest in helping lower tier ISPs compete
 - The “Big Boys” all peer with each other at no/little cost
- Harder to deal with problems without strong financial incentive

djw // CSE 561, Spring 2002, with credit to savage

L6.25

How to interconnect?

- Direct connection
 - Cost of circuit lease (\$\$\$)
- Exchange-based interconnect
 - Exchange: place that houses equipment from multiple networks to exchange traffic
 - If you both already have equipment in the same building somewhere, then just run a cable between your machines (cheap)
 - Neutral exchanges vs affiliated exchanges

djw // CSE 561, Spring 2002, with credit to savage

L6.26

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?)
 - Very hard to be a small ISP
- **Business issues can have serious operational/performance impact on the Internet**

Discussion

- **Competition and incentives**
 - What policy knobs do we need?
- **Implicit trust issues in transit routes**
 - Will X really get my packets to Y who isn't X's customer?
 - What if someone lies?