Lecture 6 and 7 (Feb 5 and 10, 2004)

Outline

Exterior Gateway Protocol
- Border Gateway Protocol – BGPv4

EGP: Exterior Gateway Protocol

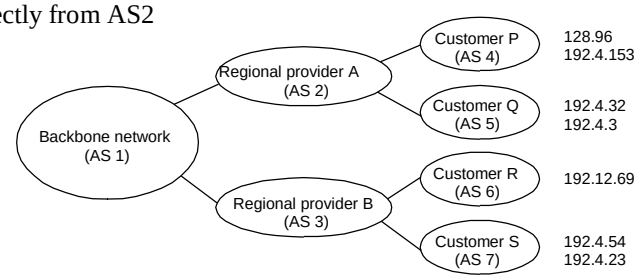
- Overview
  - designed for tree-structured Internet
  - concerned with *reachability*, not optimal routes
- Protocol messages
  - neighbor acquisition: one router requests that another be its peer; peers exchange reachability information
  - neighbor reachability: one router periodically tests if the another is still reachable; exchange HELLO/ACK messages; uses a k-out-of-n rule
  - routing updates: peers periodically exchange their routing tables (distance-vector)

BGP-4: Border Gateway Protocol

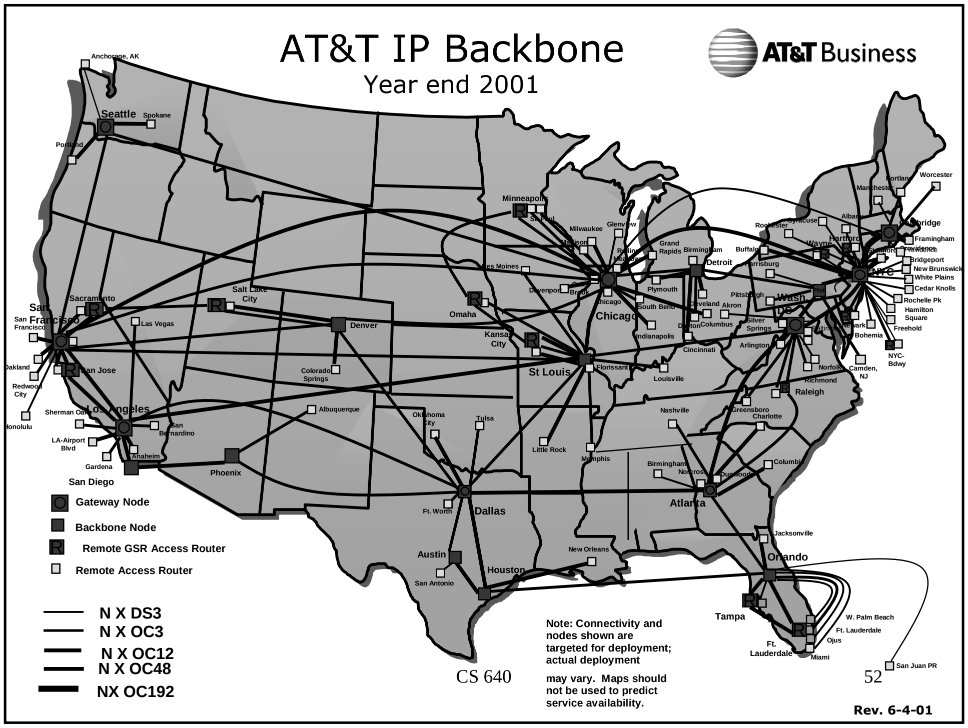
- AS Types
  - stub AS: has a single connection to one other AS
    - carries local traffic only
  - multihomed AS: has connections to more than one AS
    - refuses to carry transit traffic
  - transit AS: has connections to more than one AS
    - carries both transit and local traffic
- Each AS has:
  - one or more border routers
  - one BGP *speaker* that advertises:
    - local networks
    - other reachable networks (transit AS only)
    - provides *path* information

BGP Example

- Speaker for AS2 advertises reachability to P and Q
  - network 128.96, 192.4.153, 192.4.32, and 192.4.3, can be reached directly from AS2

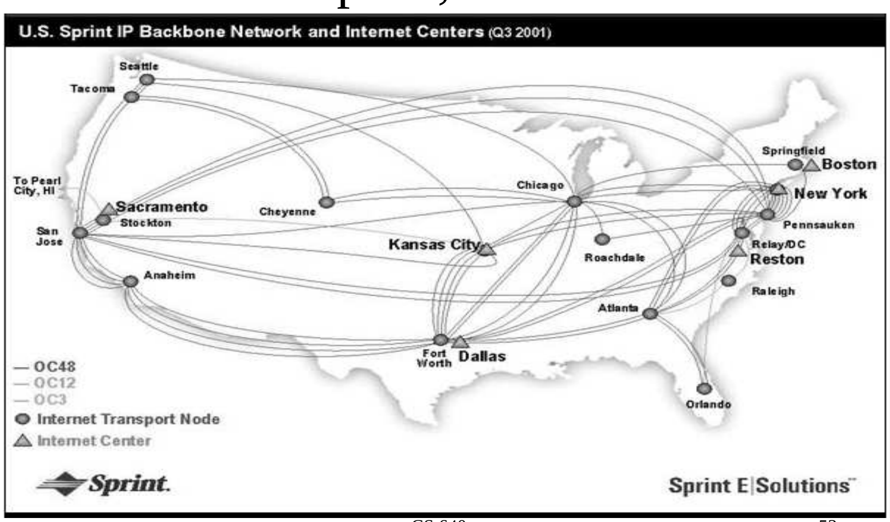


- Speaker for backbone advertises
  - networks 128.96, 192.4.153, 192.4.32, and 192.4.3 can be reached along the path (AS1, AS2).
- Speaker can cancel previously advertised paths



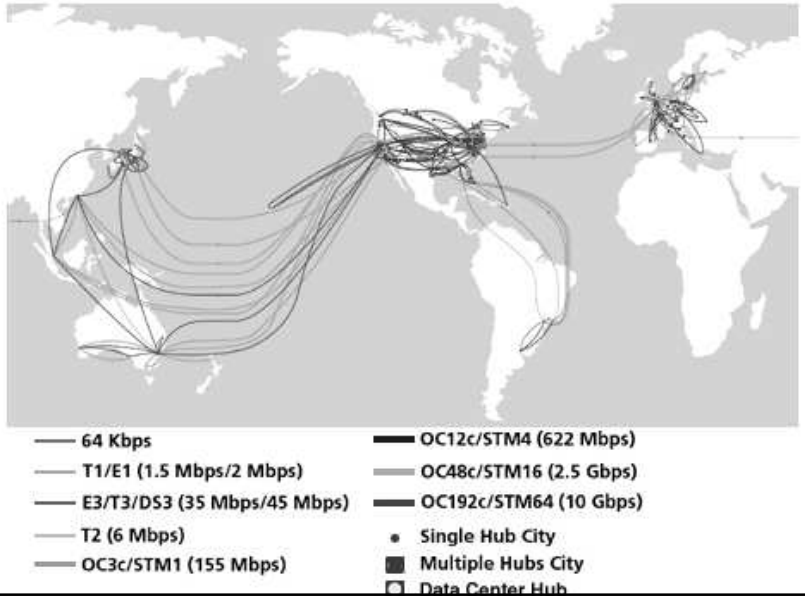
CS 640 52

Sprint, USA

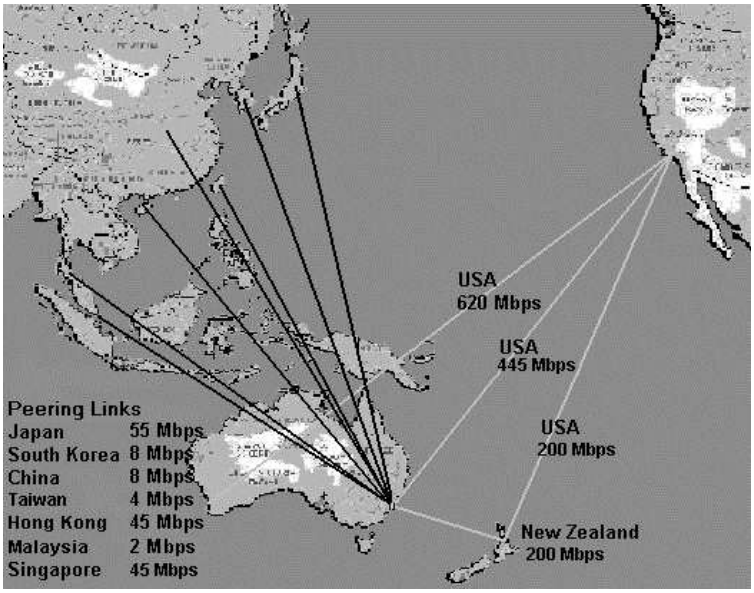


CS 640 53

WorldCom (UUNet)

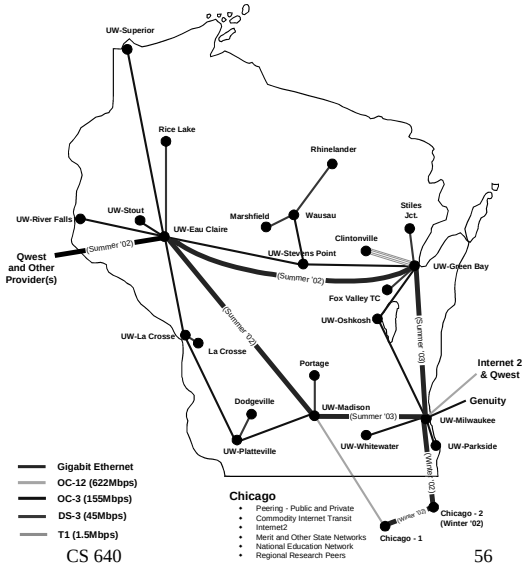


Telstra international

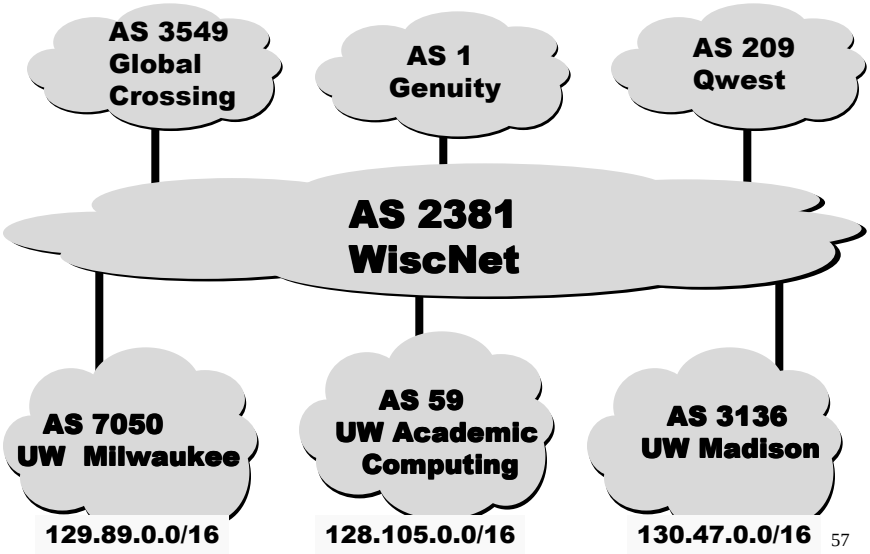


55

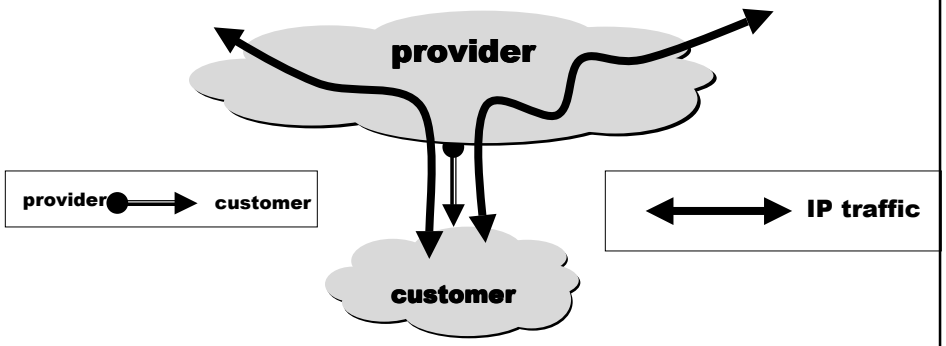
wiscnet.net



Partial View of cs.wisc.edu Neighborhood

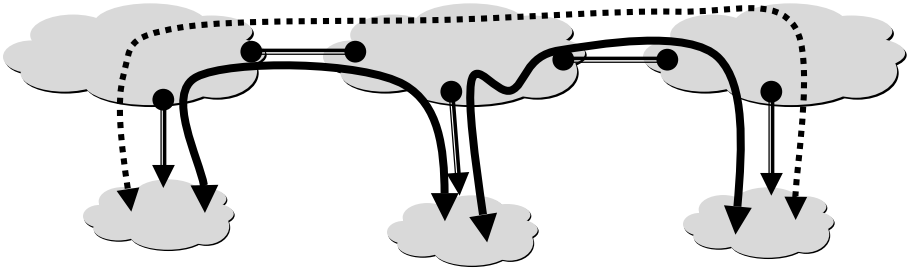


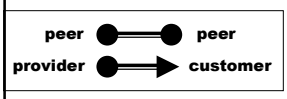
Customers and Providers

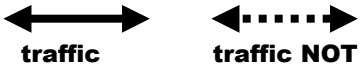


Customer pays provider for access to the Internet

The "Peering" Relationship





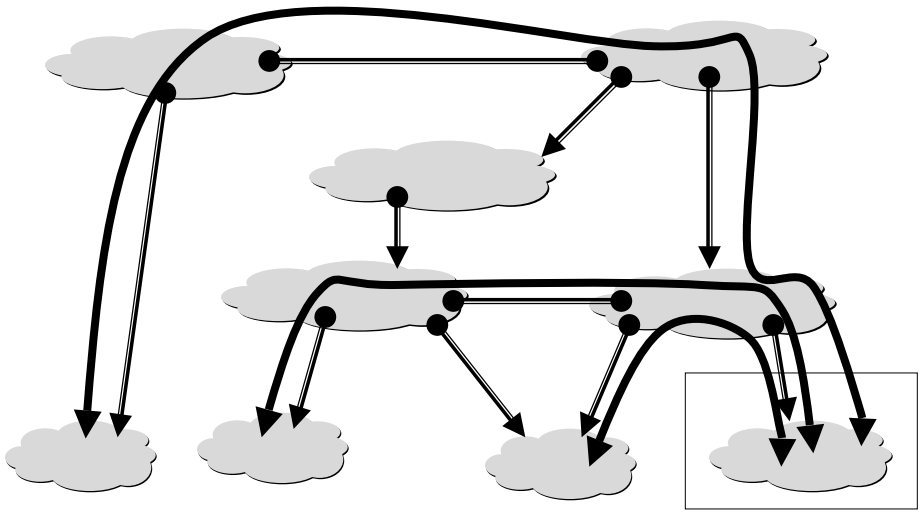


Peers provide transit between their respective customers

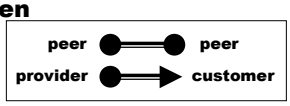
<u>Peers do not provide transit between peers</u>

Peers (often) do not exchange $$$

Examples of Peering



Peering also allows connectivity between the customers of "Tier 1" providers.



To Peer or Not to Peer?

Peer

- Reduces upstream transit costs
- Can increase end-to-end performance
- May be the only way to connect your customers to some part of the Internet ("Tier 1")

Don't Peer

- You would rather have customers
- Peers are usually your competition
- Peering relationships may require periodic renegotiation

Peering struggles are by far the most contentious issues in the ISP world!

Peering agreements are often confidential.

Autonomous Systems (ASes)

An autonomous system is an autonomous routing domain that has been assigned an Autonomous System Number (ASN).

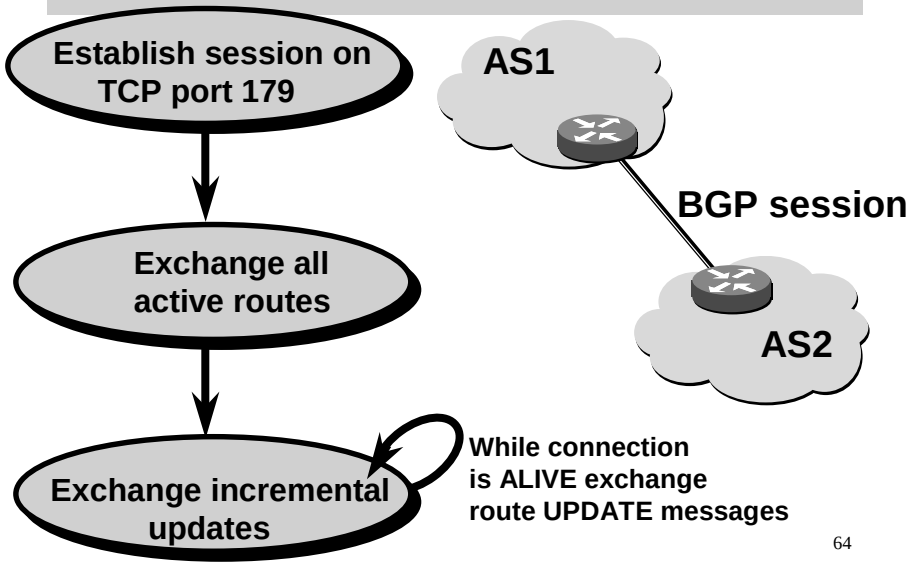
> … the administration of an AS appears to other ASes to have a single coherent interior routing plan and presents a consistent picture of what networks are reachable through it.
>
> RFC 1930: Guidelines for creation, selection, and registration of an Autonomous System

BGP-4

- **BGP** = **B**order **G**ateway **P**rotocol
- Is a **Policy-Based** routing protocol
- Is the **de facto EGP** of today's global Internet
- Relatively simple protocol, but configuration is complex and the entire world can see, and be impacted by, your mistakes.

- **1989 : BGP-1 [RFC 1105]**
  - **Replacement for EGP (1984, RFC 904)**
- **1990 : BGP-2 [RFC 1163]**
- **1991 : BGP-3 [RFC 1267]**
- **1995 : BGP-4 [RFC 1771]**
  - **Support for Classless Interdomain Routing (CIDR)**

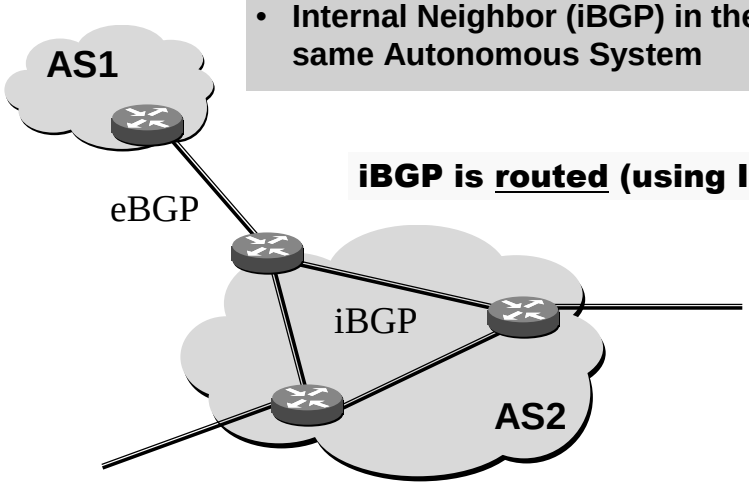
BGP Operations (Simplified)



64

Two Types of BGP Neighbor Relationships

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



iBGP is routed (using IGP!)

65

Four Types of BGP Messages

- **Open** : Establish a peering session.
- **Keep Alive** : Handshake at regular intervals.
- **Notification** : Shuts down a peering session.
- **Update** : Announcing new routes or withdrawing previously announced routes.

announcement
=
prefix + attributes values

BGP Attributes

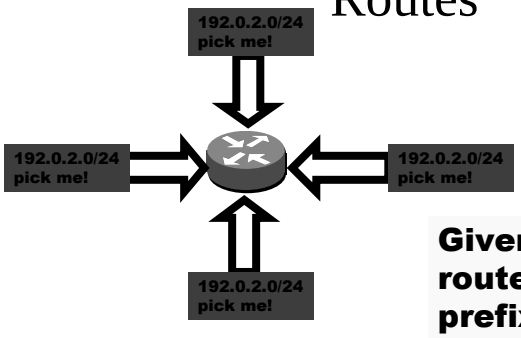
Value	Code	Reference
1	ORIGIN	[RFC1771]
2	AS_PATH	[RFC1771]
3	NEXT_HOP	[RFC1771]
4	MULTI_EXIT_DISC	[RFC1771]
5	LOCAL_PREF	[RFC1771]
6	ATOMIC_AGGREGATE	[RFC1771]
7	AGGREGATOR	[RFC1771]
8	COMMUNITY	[RFC1997]
9	ORIGINATOR_ID	[RFC2796]
10	CLUSTER_LIST	[RFC2796]
11	DPA	[Chen]
12	ADVERTISER	[RFC1863]
13	RCID_PATH / CLUSTER_ID	[RFC1863]
14	MP_REACH_NLRI	[RFC2283]
15	MP_UNREACH_NLRI	[RFC2283]
16	EXTENDED COMMUNITIES	[Rosen]
...		
255	reserved for development	

From IANA: http://www.iana.org/assignments/bgp-parameters

Most important attributes

Not all attributes need to be present in every announcement

Attributes are Used to Select Best Routes



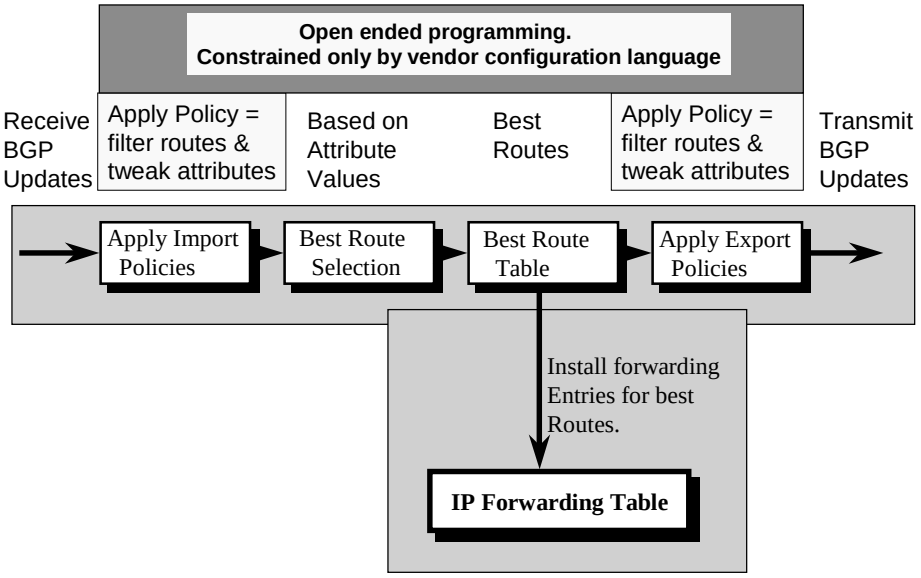
Given multiple routes to the same prefix, a BGP speaker must pick at most <u>one</u> best route

(Note: it could reject them all!)

Route Selection Summary

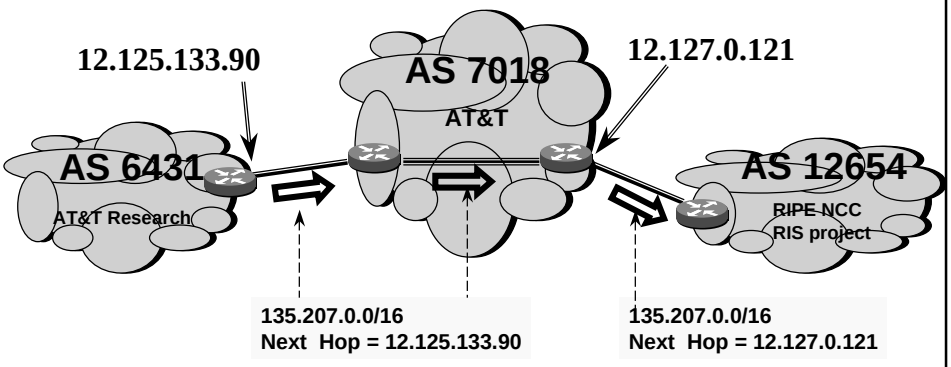
Highest Local Preference	**Enforce relationships**
Shortest ASPATH **Lowest MED** **i-BGP < e-BGP** **Lowest IGP cost to BGP egress**	**traffic engineering**
Lowest router ID	**Throw up hands and break ties**

BGP Route Processing



70

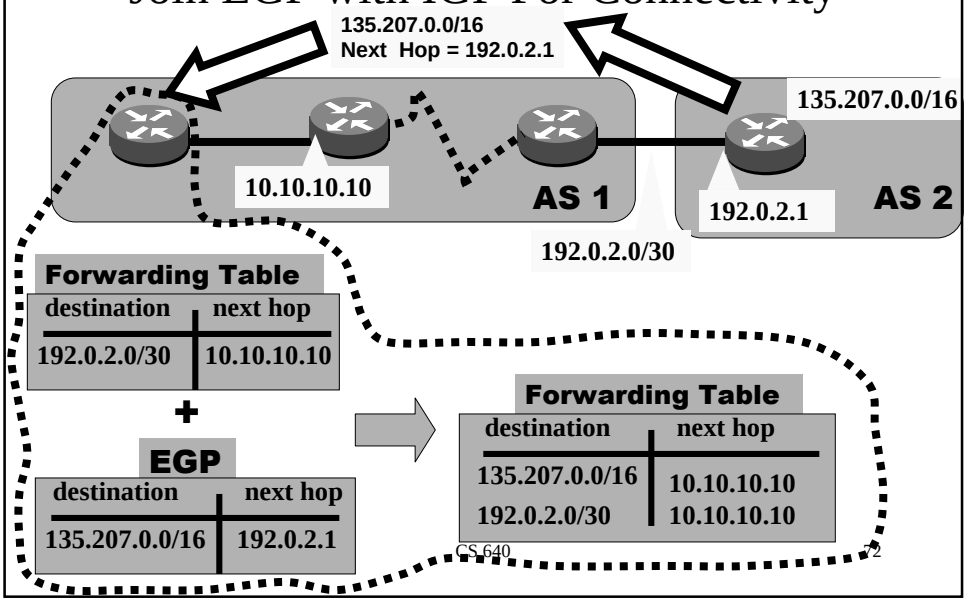
BGP Next Hop Attribute



Every time a route announcement crosses an AS boundary, the Next Hop attribute is changed to the IP address of the border router that announced the route.

71

Join EGP with IGP For Connectivity

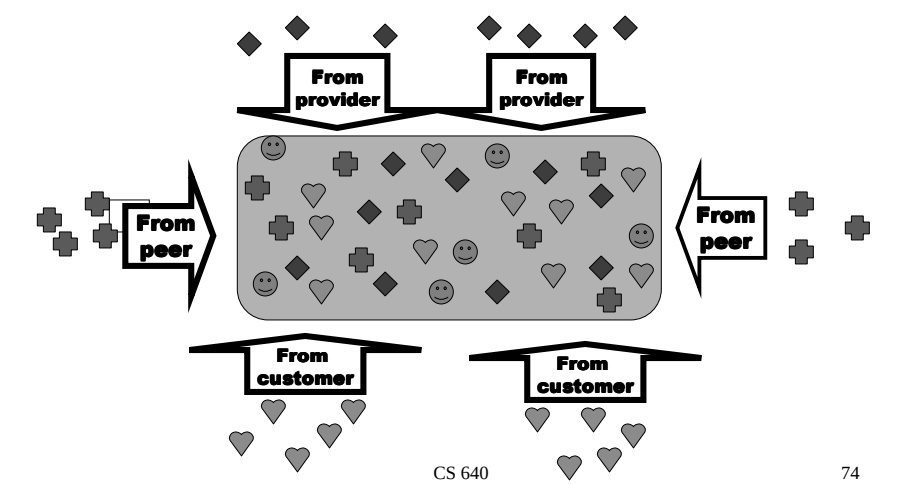


Implementing Customer/Provider and Peer/Peer relationships

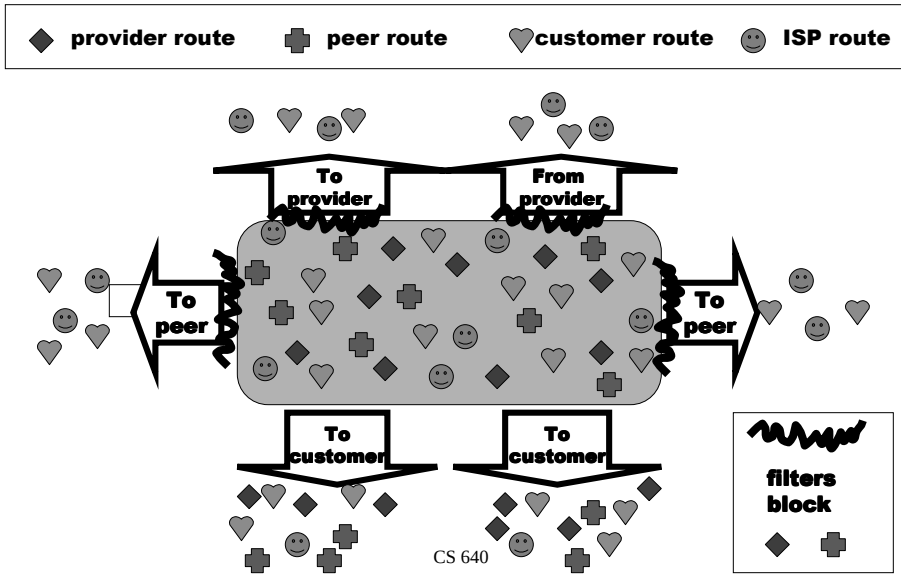
Two parts:

- Enforce transit relationships
  - Outbound route filtering
- Enforce order of route preference
  - provider < peer < customer

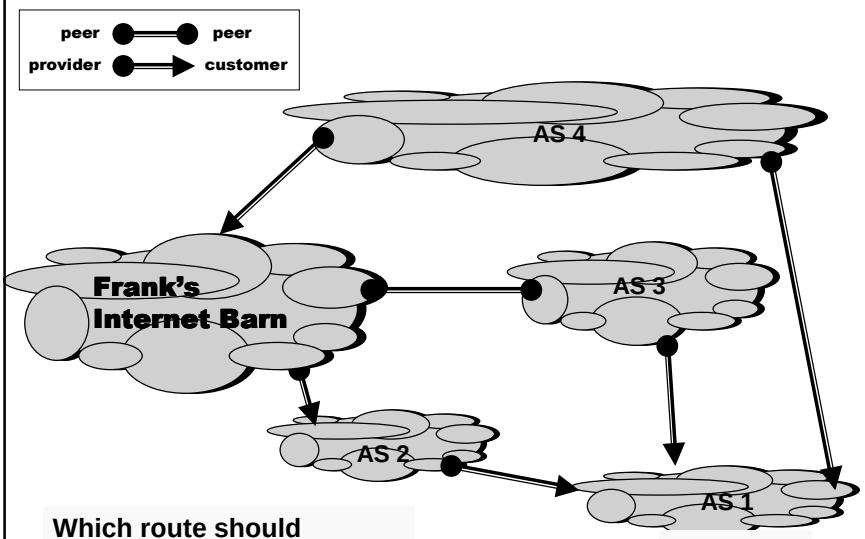
Import Routes



Export Routes



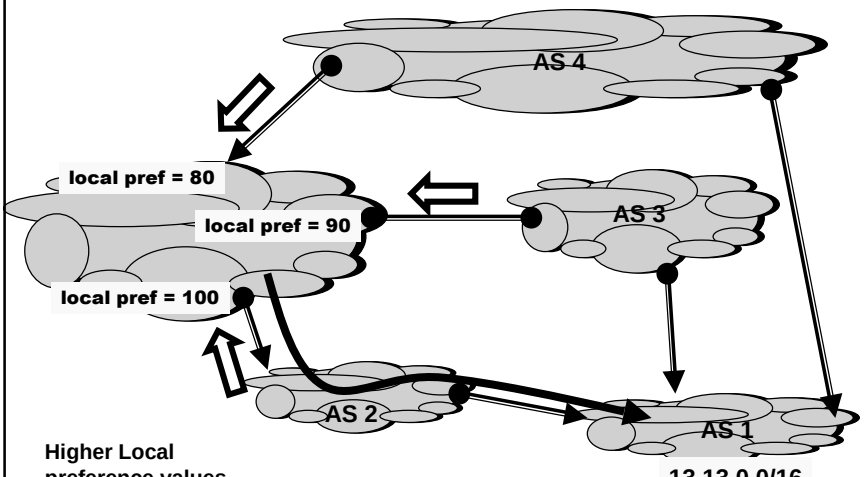
So Many Choices



Which route should Frank pick to 13.13.0.0./16?

76

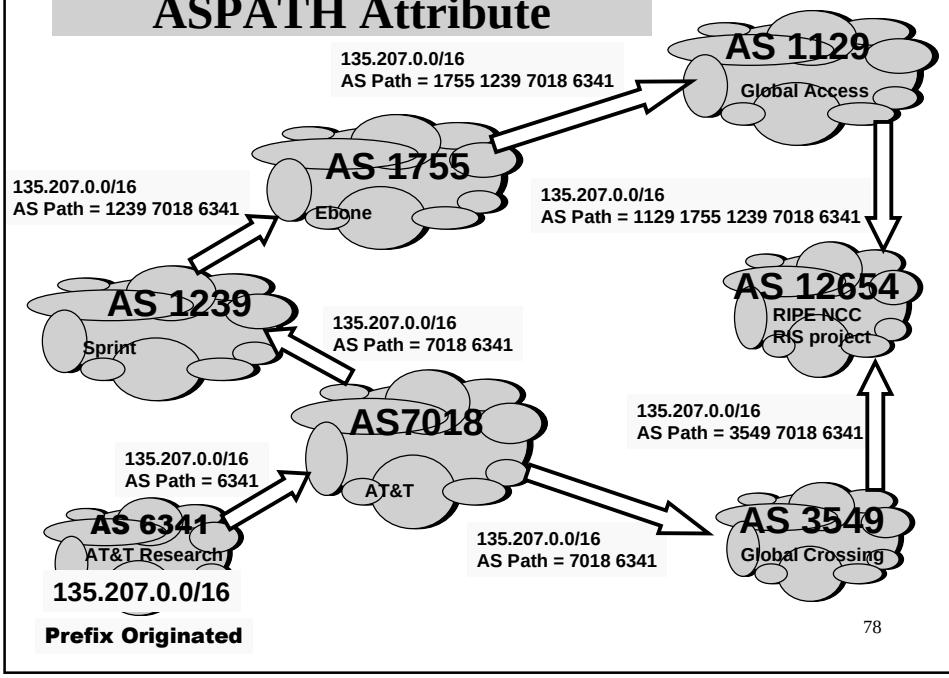
LOCAL PREFERENCE



Higher Local preference values are more preferred

77

ASPATH Attribute

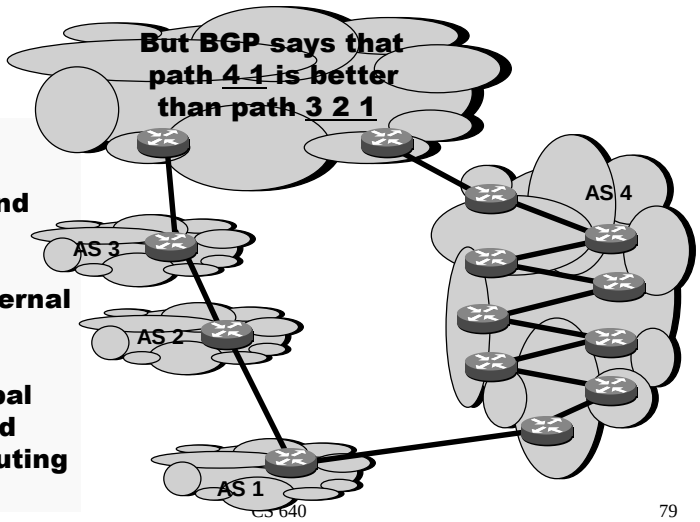


78

Shorter Doesn't Always Mean Shorter

In fairness: could you do this "right" and still scale?

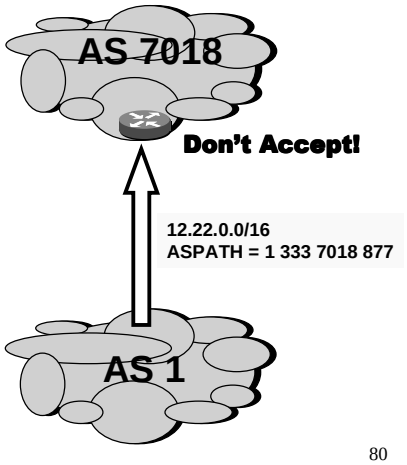
Exporting internal state would dramatically increase global instability and amount of routing state



79

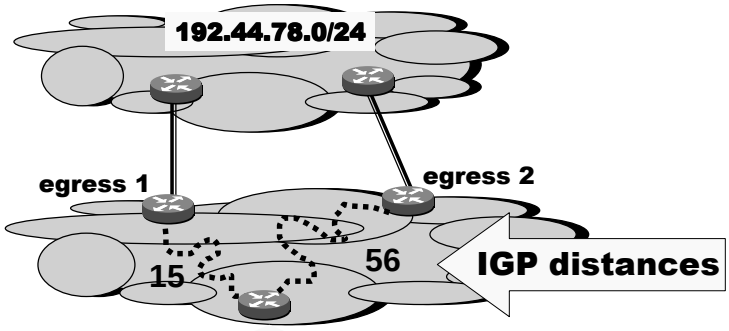
Interdomain Loop Prevention

BGP at AS YYY will never accept a route with ASPATH containing YYY.



80

Hot Potato Routing: Go for the Closest Egress Point

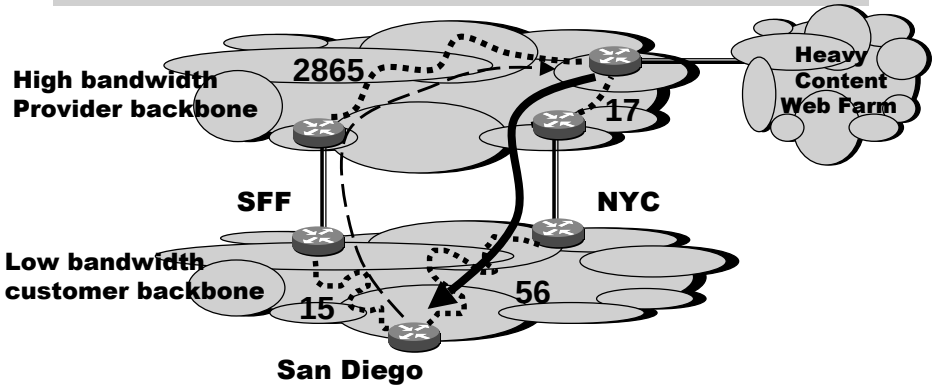


This Router has two BGP routes to 192.44.78.0/24.

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

81

Sometimes hot potato is not enough

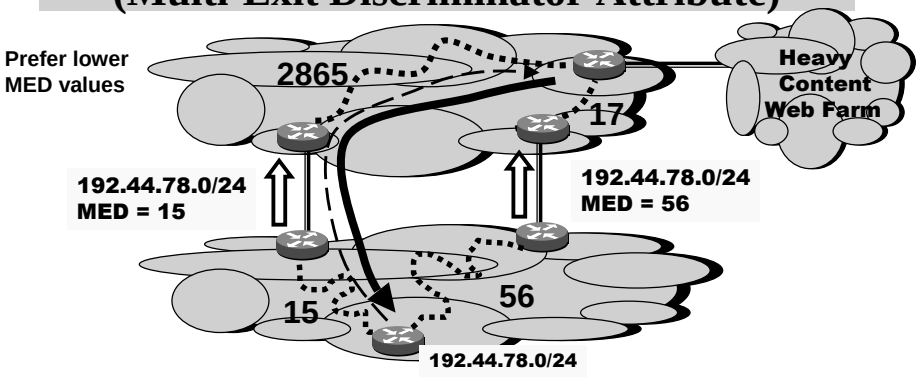


Many customers want their provider to carry the bits!

tiny ftp request
huge ftp reply

82

Cold Potato Routing with MEDs (Multi-Exit Discriminator Attribute)



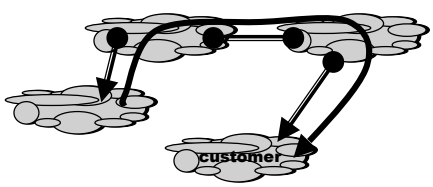
This means that MEDs must be considered BEFORE IGP distance!

Note1 : some providers will not listen to MEDs

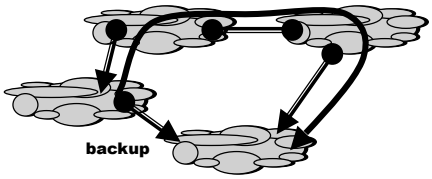
Note2 : MEDs need not be tied to IGP distance

83

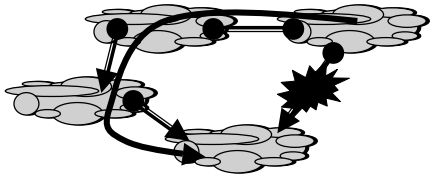
Policies Can Interact Strangely ("Route Pinning" Example)



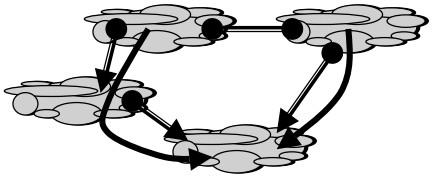
1



2 **Install backup link**



3 **Disaster strikes primary link and the backup takes over**



4 **Primary link is restored but some traffic remains *pinned* to backup**