CS 640 Introduction to Computer Networks

Lecture12

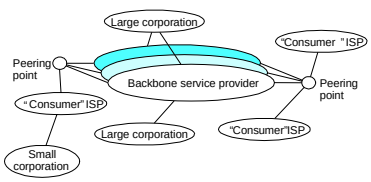
Based on slides by Tim Griffin

Today's lecture

- Inter-domain routing
  - Architecture and relationships between networks
  - BGP
    - Introduction
    - Implementing peering relationships
    - Backups and multihoming
    - Hot potato/cold potato

Internet Structure

Today



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.

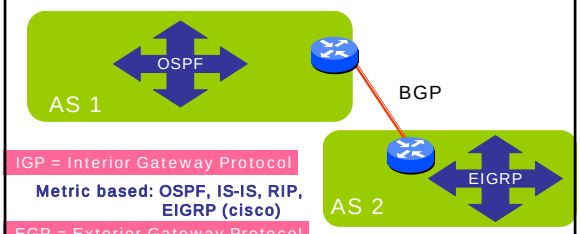
RFC 1930: Guidelines for creation, selection, and registration of an Autonomous System

How Are Forwarding Tables Populated to implement Routing?

Statically	Dynamically
Administrator manually configures forwarding table entries	Routers exchange network reachability information using ROUTING PROTOCOLS. Routers use this to compute best routes
+ More control	+ Can rapidly adapt to changes in network topology
+ Not restricted to destination-based forwarding	+ Can be made to scale well
- Doesn't scale	- Complex distributed algorithms
- Slow to adapt to network failures	- Consume CPU, Bandwidth, Memory
	- Debugging can be difficult
	- Current protocols are destination-based

In practice : a mix of these.
Static routing mostly at the "edge"

Architecture of Dynamic Routing



IGP = Interior Gateway Protocol

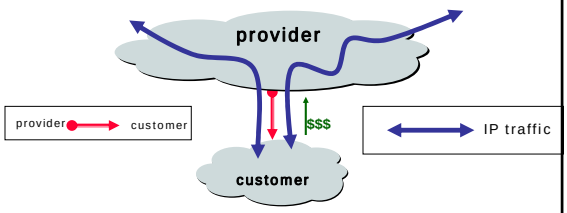
Metric based: OSPF, IS-IS, RIP, EIGRP (cisco)

EGP = Exterior Gateway Protocol

Policy based: BGP

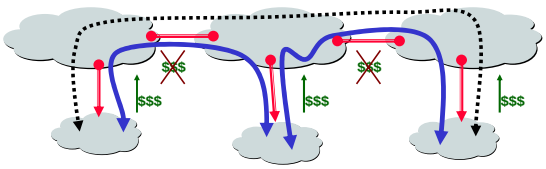
The Routing Domain of BGP is the entire Internet

Customers and Providers



Customer pays provider for access to the Internet

The "Peering" Relationship



peer — peer
provider → customer

traffic allowed

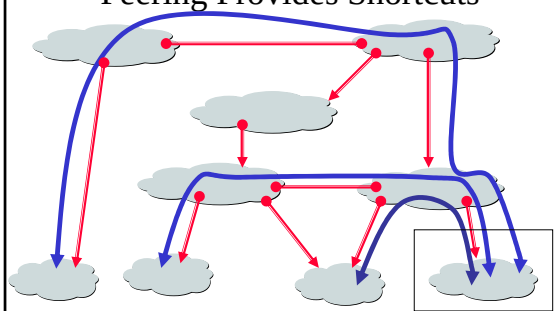
traffic NOT allowed

Peers provide transit between their respective customers

Peers do not provide transit between peers

Peers (often) do not exchange $$$

Peering Provides Shortcuts



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

peer — peer
provider → customer

Peering Wars

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.

Today's lecture

- Inter-domain routing
  - Architecture and relationships between networks
  - BGP
    - Introduction
    - Implementing peering relationships
    - Backups and multihoming
    - Hot potato/cold potato

BGP-4

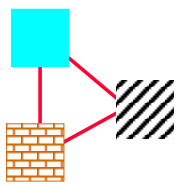
- **BGP** = **B**order **G**ateway **P**rotocol
- Aims to ensure *reachability* between ASes
  - "Doesn't know" about internals of ASes
  - Not based on "shortest distance"
  - Based on business relationships
- It is a *path vector* protocol (trivial to avoid loops)
  - Advertisements carry all ASes on the path to originator
- Relatively simple protocol but
  - Configuration is complex (captures business relationships)
  - The entire world can be impacted by your mistakes

A real-world AS Graph

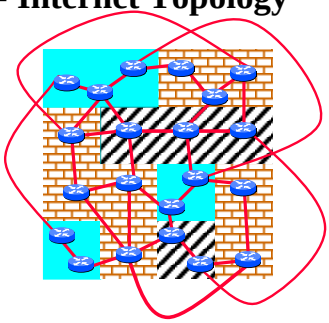
The subgraph showing all ASes that have more than 100 neighbors in full graph of 11,158 nodes. July 6, 2001.

AS Graph != Internet Topology

BGP was designed to throw away information!

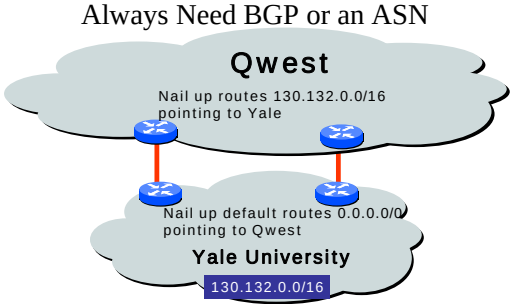


The AS graph may look like this.



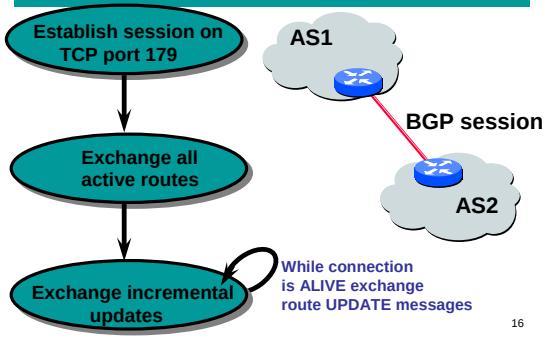
Reality may be closer to this...

Autonomous Routing Domains Don't Always Need BGP or an ASN



Static routing is the most common way of connecting an autonomous routing domain to the Internet.
This helps explain why BGP is a mystery to many ...

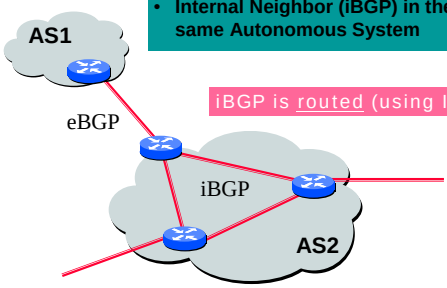
BGP Operations (Simplified)



16

Two Types of BGP Neighbor Relations

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

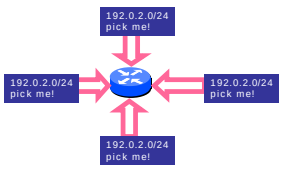


iBGP is routed (using IGP!)

17

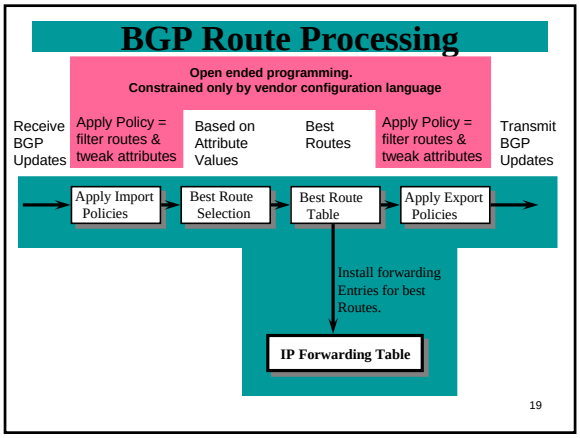
BGP announcement

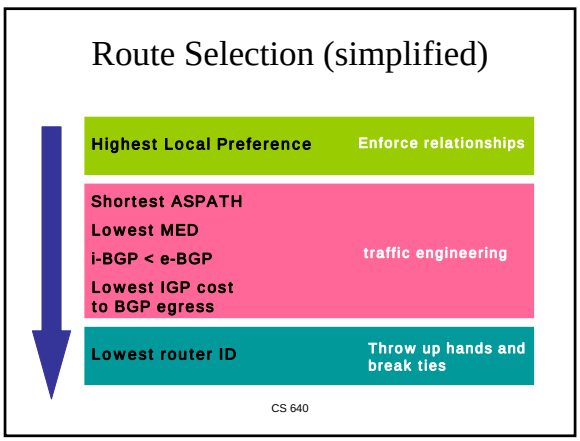
- Announcement = prefix + attributes
- BGP has many attributes, not all present in each announcement
  - AS path
  - Local preference
  - Multi-exit discriminator MED
  - Community (unspecified meaning)

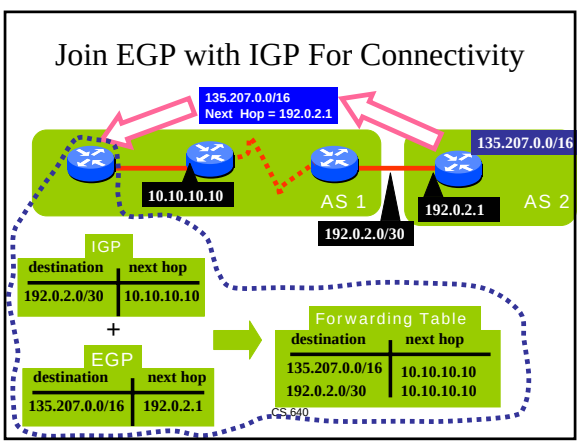


Given multiple routes to the same prefix, a BGP speaker must pick at most one best route

(Note: it could reject them all!)


BGP Route Processing
Open ended programming.
Constrained only by vendor configuration language
Receive BGP Updates
Apply Policy = filter routes & tweak attributes
Based on Attribute Values
Best Routes
Apply Policy = filter routes & tweak attributes
Transmit BGP Updates
Apply Import Policies
Best Route Selection
Best Route Table
Apply Export Policies
Install forwarding Entries for best Routes.
IP Forwarding Table
19


Route Selection (simplified)
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
CS 640


Join EGP with IGP For Connectivity
135.207.0.0/16
Next Hop = 192.0.2.1
135.207.0.0/16
10.10.10.10
AS 1
192.0.2.1
AS 2
192.0.2.0/30
IGP
destination
next hop
192.0.2.0/30
10.10.10.10
+
EGP
destination
next hop
135.207.0.0/16
192.0.2.1
Forwarding Table
destination
next hop
135.207.0.0/16
10.10.10.10
192.0.2.0/30
10.10.10.10
CS 640

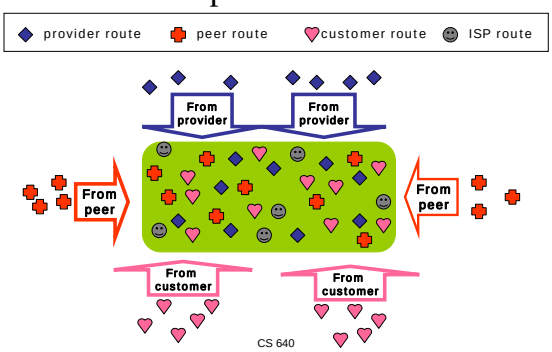
Today's lecture

- Inter-domain routing
  - Architecture and relationships between networks
  - BGP
    - Introduction
    - Implementing peering relationships
    - Backups and multihoming
    - Hot potato/cold potato

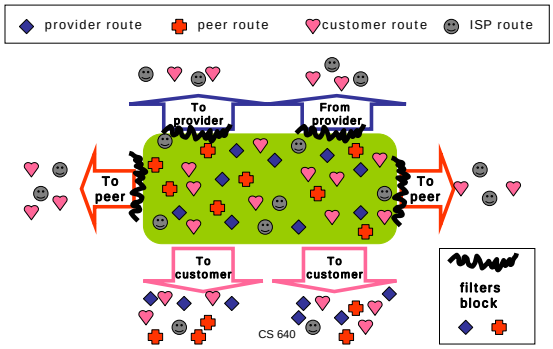
Implementing Customer/Provider and Peer/Peer relationships

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

Import Routes



Export Routes



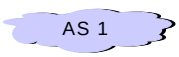
Communities Example

- 1:100
  - Customer routes
- 1:200
  - Peer routes
- 1:300
  - Provider Routes

Import

- To Customers
  - 1:100, 1:200, 1:300
- To Peers
  - 1:100
- To Providers
  - 1:100

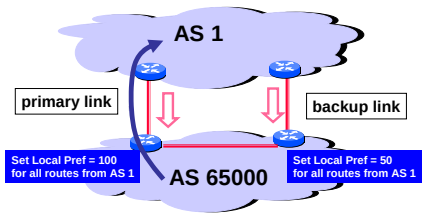
Export



Today's lecture

- Inter-domain routing
  - Architecture and relationships between networks
  - BGP
    - Introduction
    - Implementing peering relationships
    - Backups and multihoming
    - Hot potato/cold potato

Implementing Backup Links with Local Preference (Outbound Traffic)

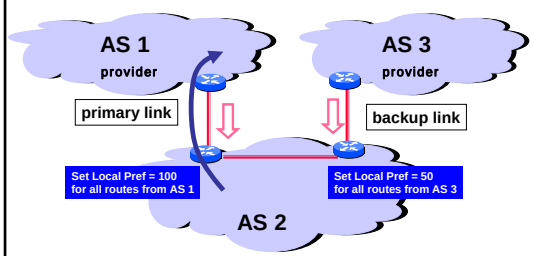


Forces outbound traffic to take primary link, unless link is down.

We'll talk about inbound traffic soon …

28

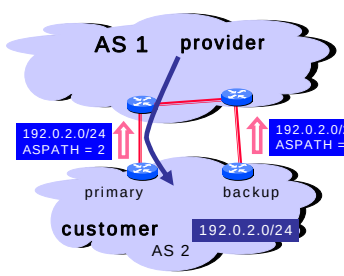
Multihomed Backups (Outbound Traffic)



Forces outbound traffic to take primary link, unless link is down.

29

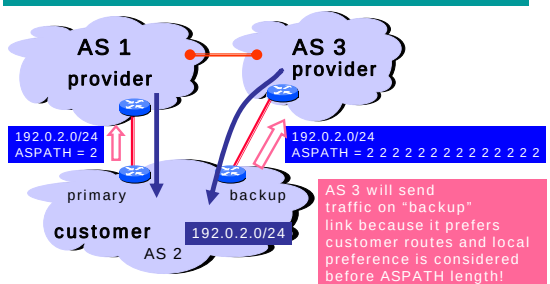
Shedding Inbound Traffic with ASPATH Padding



Padding will (usually) force inbound traffic from AS 1 to take primary link

30

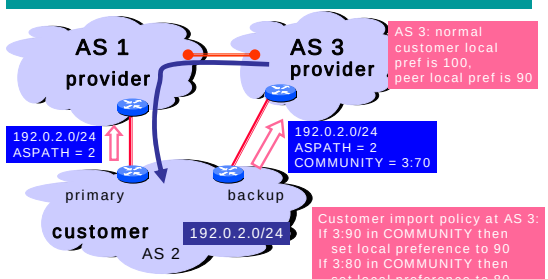
… But Padding Does Not Always Work



AS 3 will send traffic on "backup" link because it prefers customer routes and local preference is considered before ASPATH length!

Padding in this way is often used as a form of load balancing

COMMUNITY Attribute to the Rescue!



AS 3: normal customer local pref is 100, peer local pref is 90

Customer import policy at AS 3:
If 3:90 in COMMUNITY then
set local preference to 90
If 3:80 in COMMUNITY then
set local preference to 80
If 3:70 in COMMUNITY then
set local preference to 70

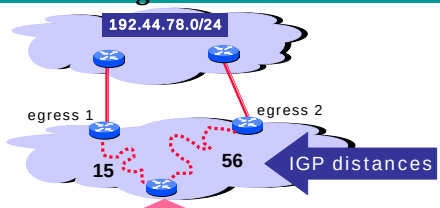
32

Today's lecture

- Inter-domain routing
  - Architecture and relationships between networks
  - BGP
    - Introduction
    - Implementing peering relationships
    - Backups and multihoming
    - Hot potato/cold potato

CS 640

Hot Potato Routing: Go for the Closest Egress Point

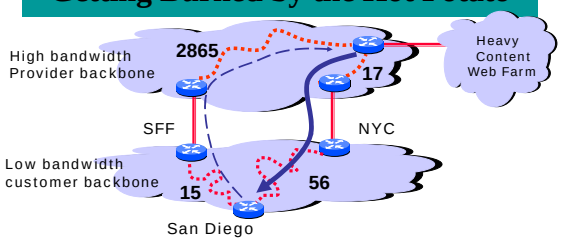


This Router has two BGP routes to 192.44.78.0/24.

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

34

Getting Burned by the Hot Potato

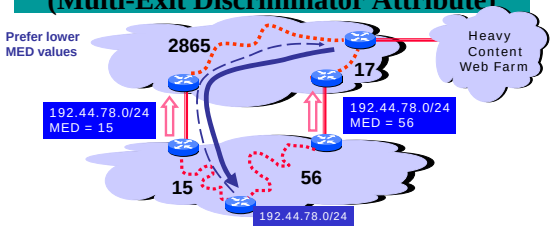


Many customers want their provider to carry the bits!

tiny http request
huge http reply

35

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

36