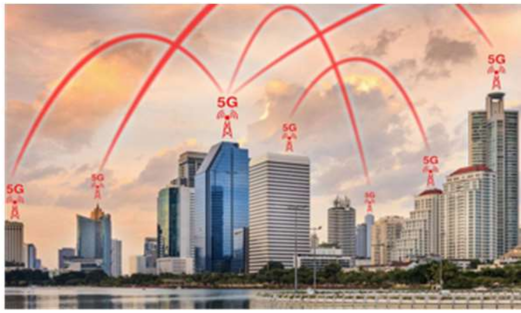


Computer Networks

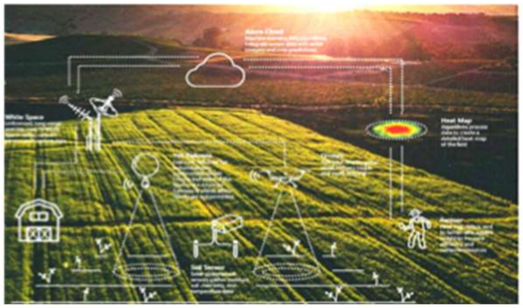
CMSC 417 : Spring 2024

COMPUTER SCIENCE
UNIVERSITY OF MARYLAND

Topic: BGP – Part2 (Textbook chapter 4)

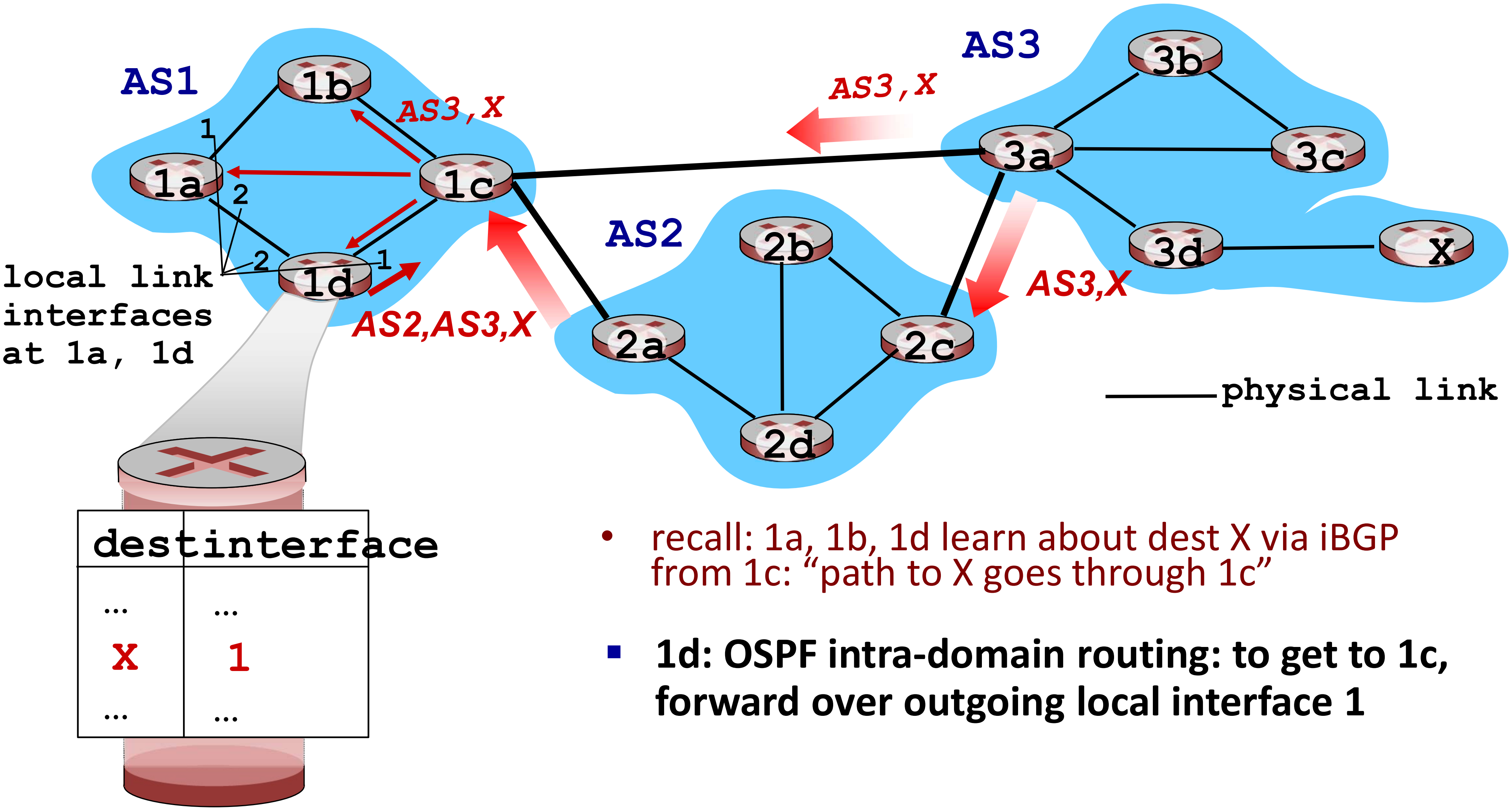
Nirupam Roy
Tu-Th 2:00-3:15pm
CSI 2117

May 1st, 2024



BGP, OSPF, forwarding table entries

Q: how does router set forwarding table entry to distant prefix?

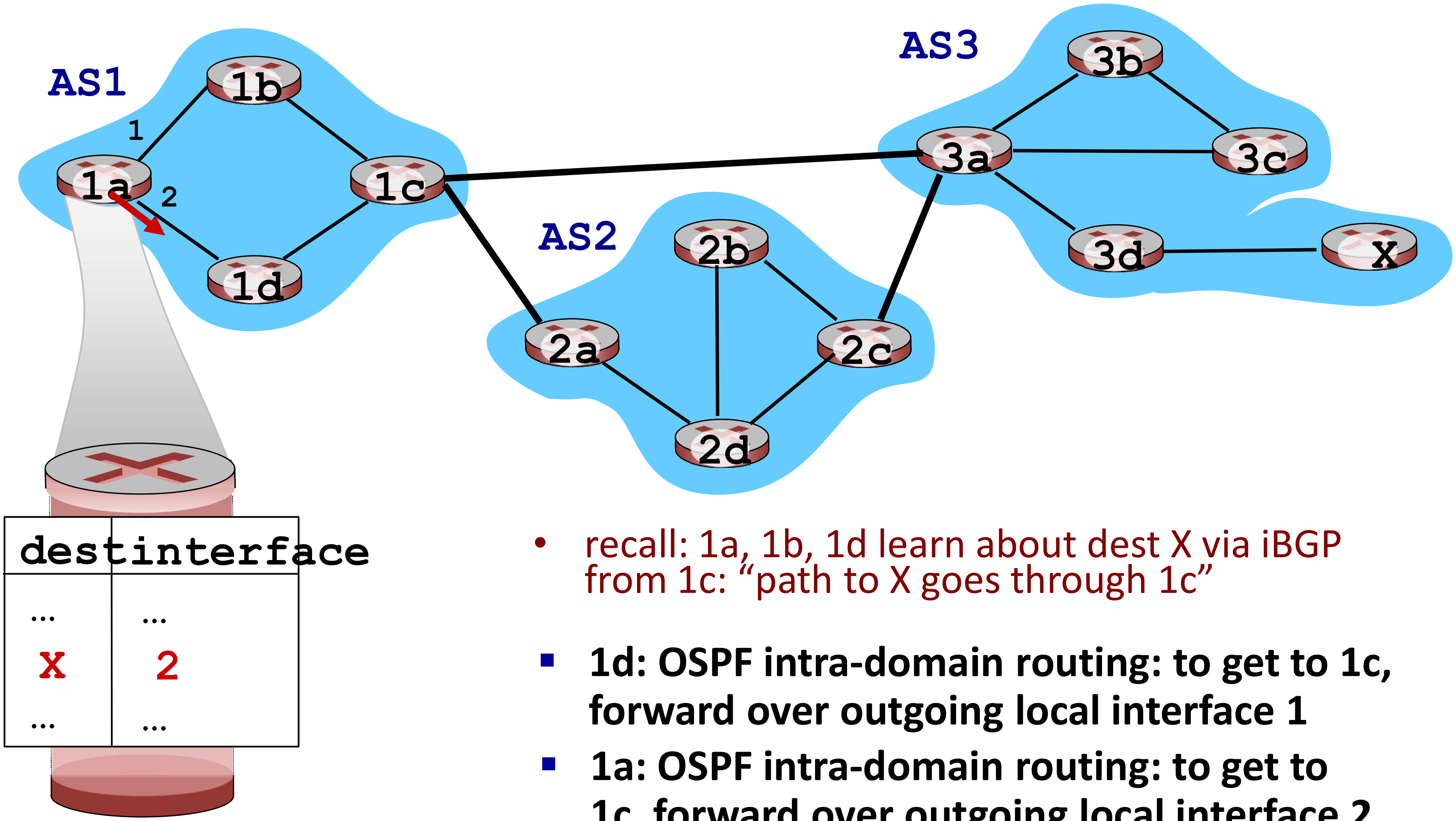


dest	interface
…	…
X	1
…	…

- recall: 1a, 1b, 1d learn about dest X via iBGP from 1c: "path to X goes through 1c"
- **1d: OSPF intra-domain routing: to get to 1c, forward over outgoing local interface 1**

BGP, OSPF, forwarding table entries

Q: how does router set forwarding table entry to distant prefix?



dest	interface
…	…
x	2
…	…

- recall: 1a, 1b, 1d learn about dest X via iBGP from 1c: “path to X goes through 1c”

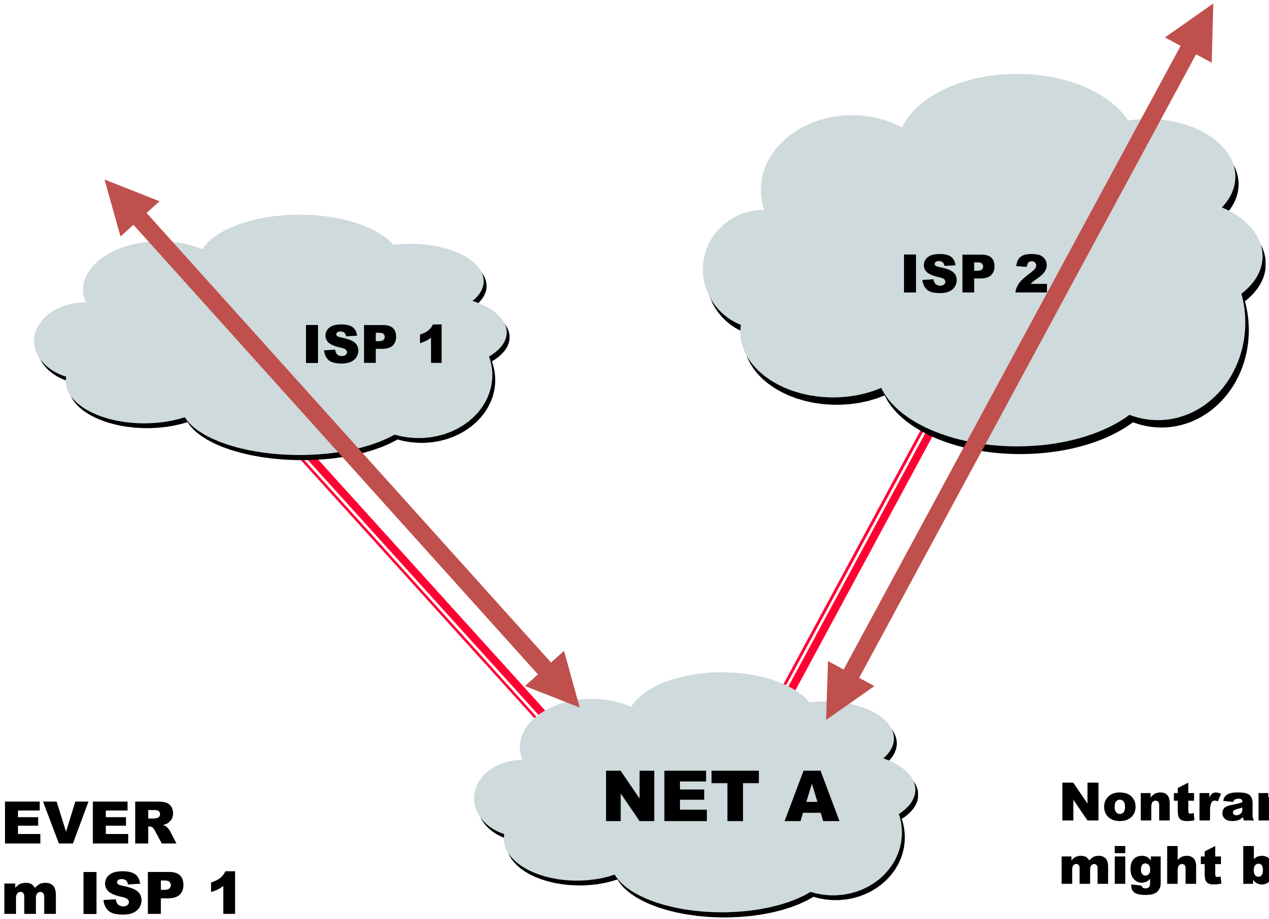
- **1d: OSPF intra-domain routing: to get to 1c, forward over outgoing local interface 1**
- **1a: OSPF intra-domain routing: to get to 1c, forward over outgoing local interface 2**

BGP route selection

- router may learn about more than one route to destination AS, selects route based on:
  1. local preference value attribute: policy decision
  2. shortest AS-PATH
  3. closest NEXT-HOP router: hot potato routing
  4. additional criteria

A dive into the BGP policies

Nontransit vs. Transit ASes



Internet Service providers (often) have transit networks

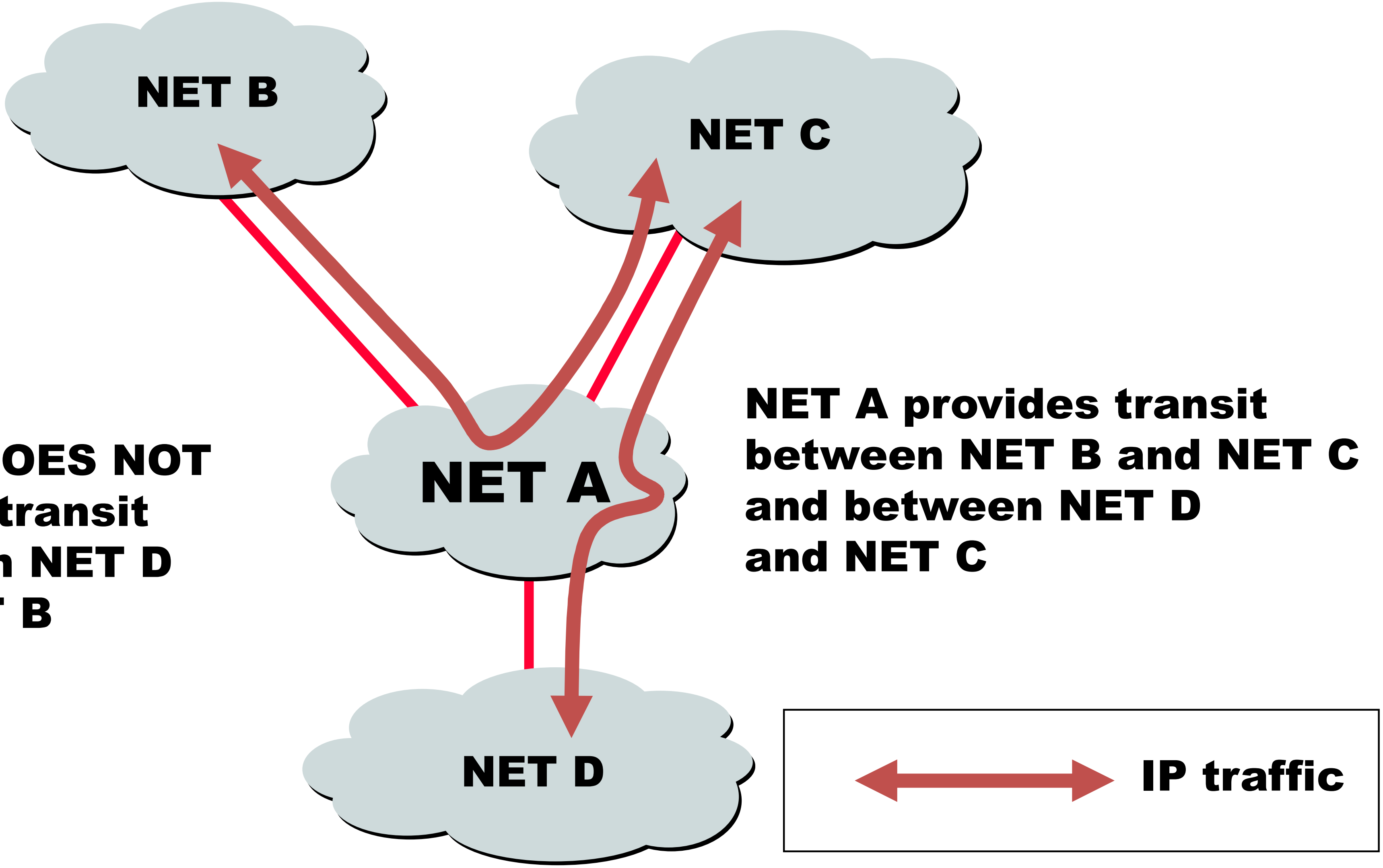
Traffic NEVER flows from ISP 1 through NET A to ISP 2 (At least not intentionally!)

Nontransit AS might be a corporate or campus network. Could be a "content provider"

⟷ IP traffic

Selective Transit

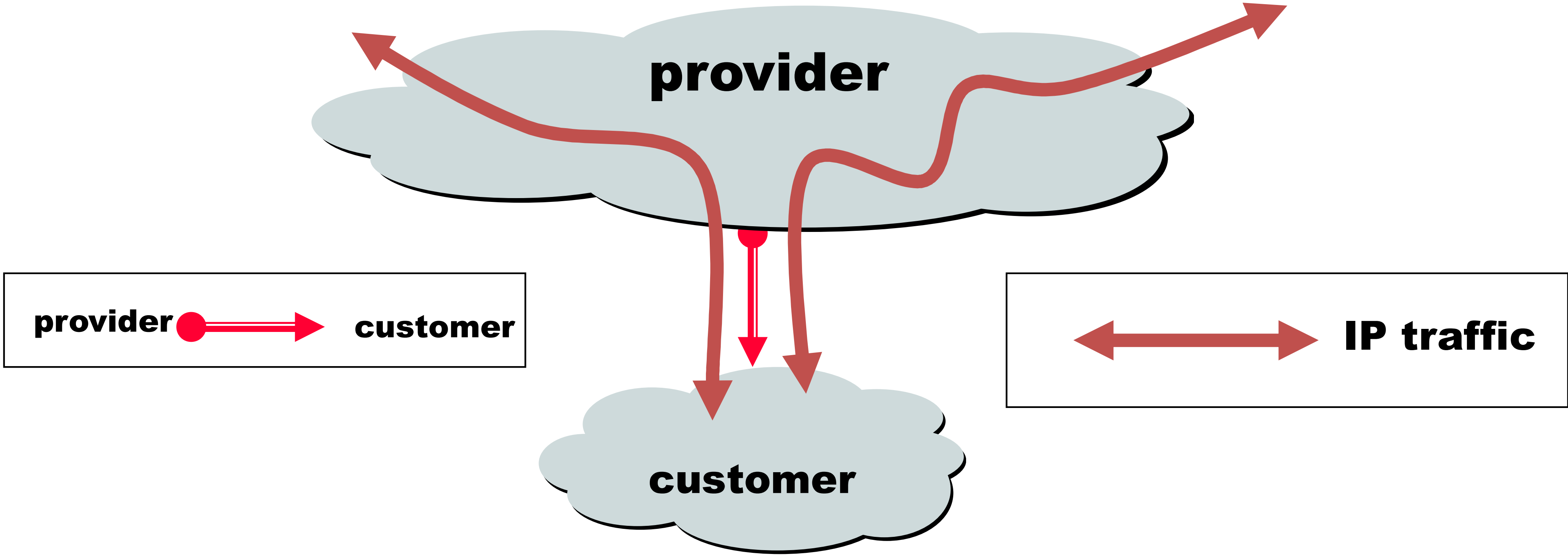
NET A DOES NOT provide transit Between NET D and NET B



NET A provides transit between NET B and NET C and between NET D and NET C

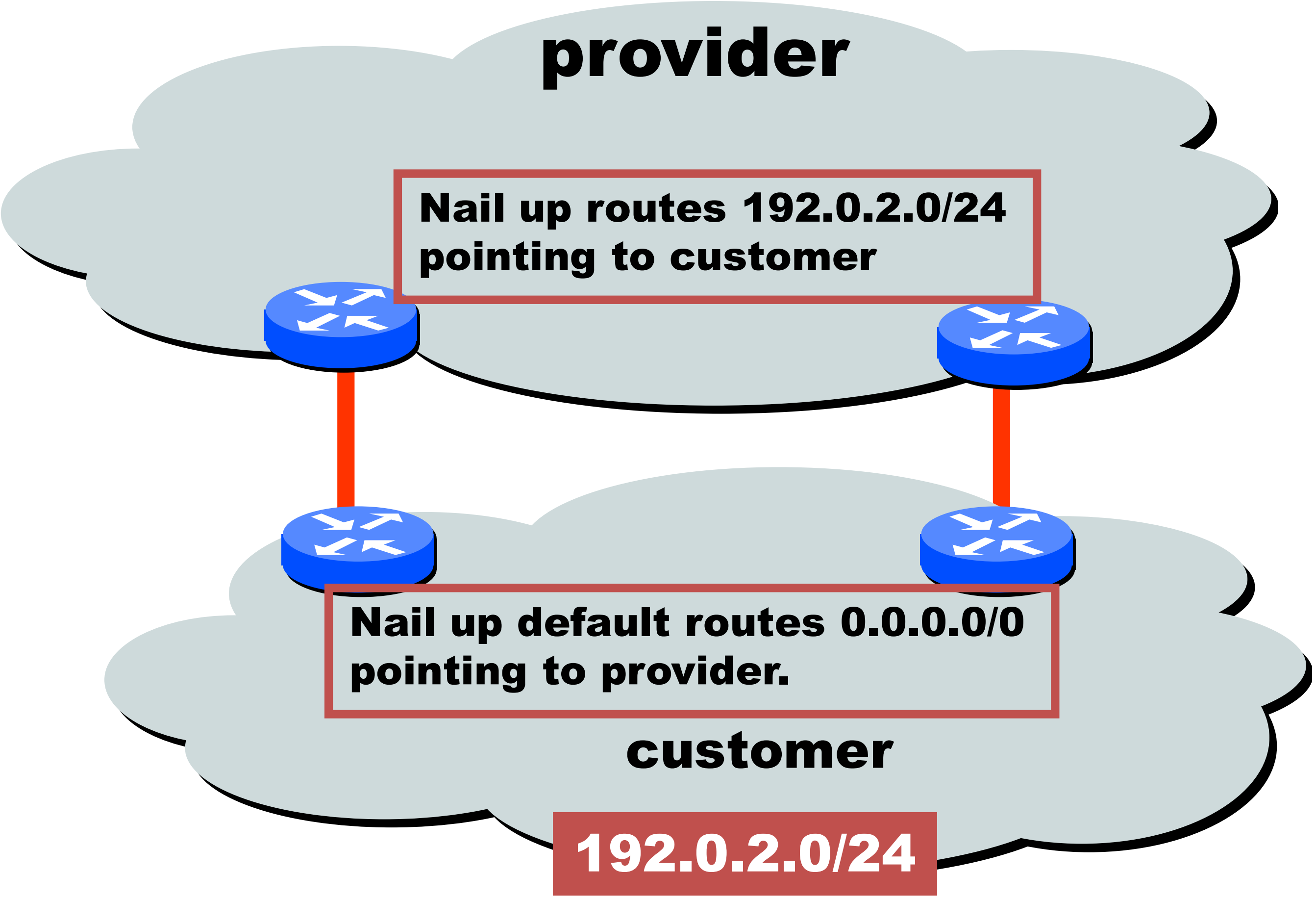
Most transit networks transit in a selective manner…

Customers and Providers



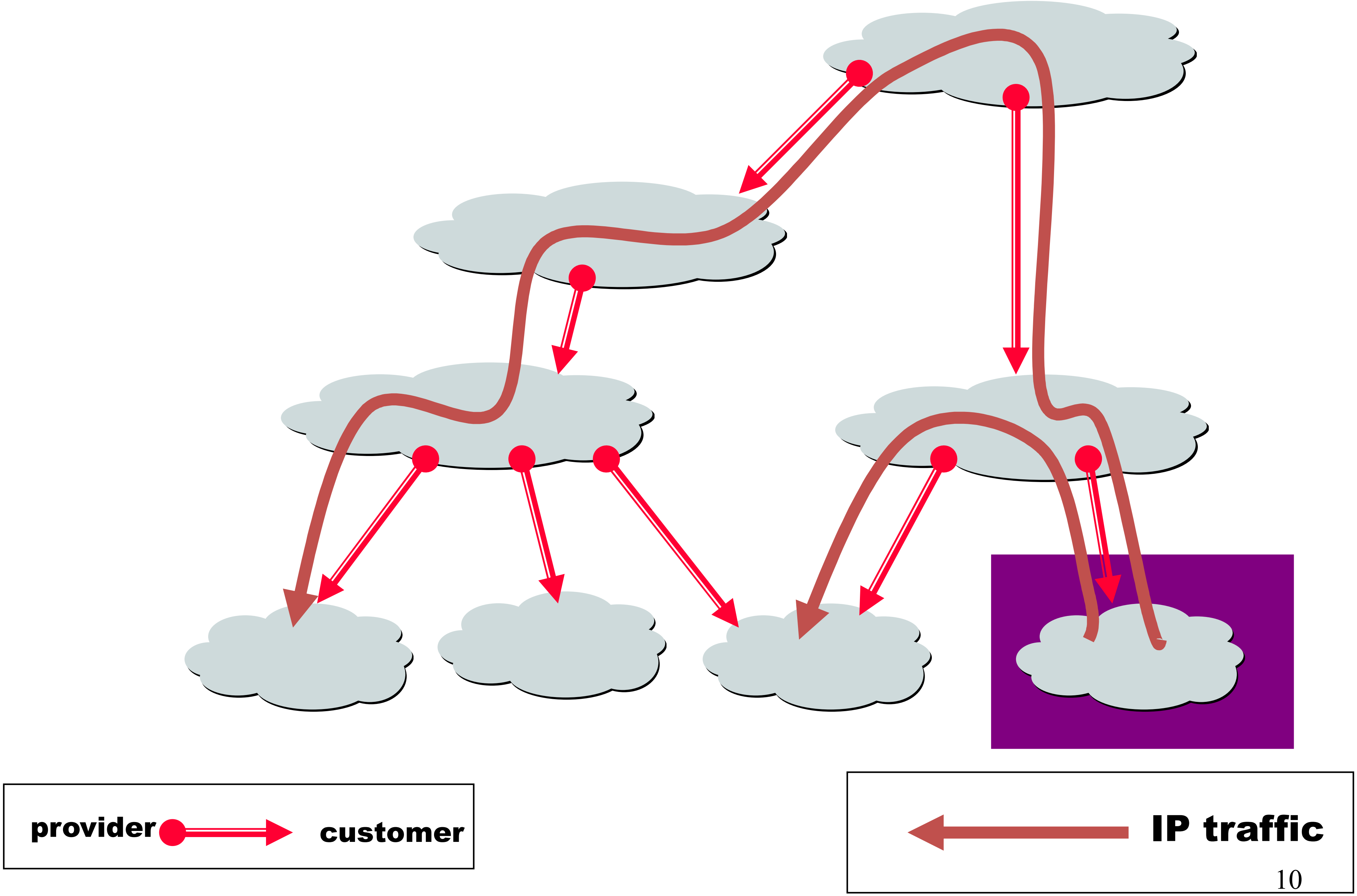
Customer pays provider for access to the Internet

Customers Don’t Always Need BGP

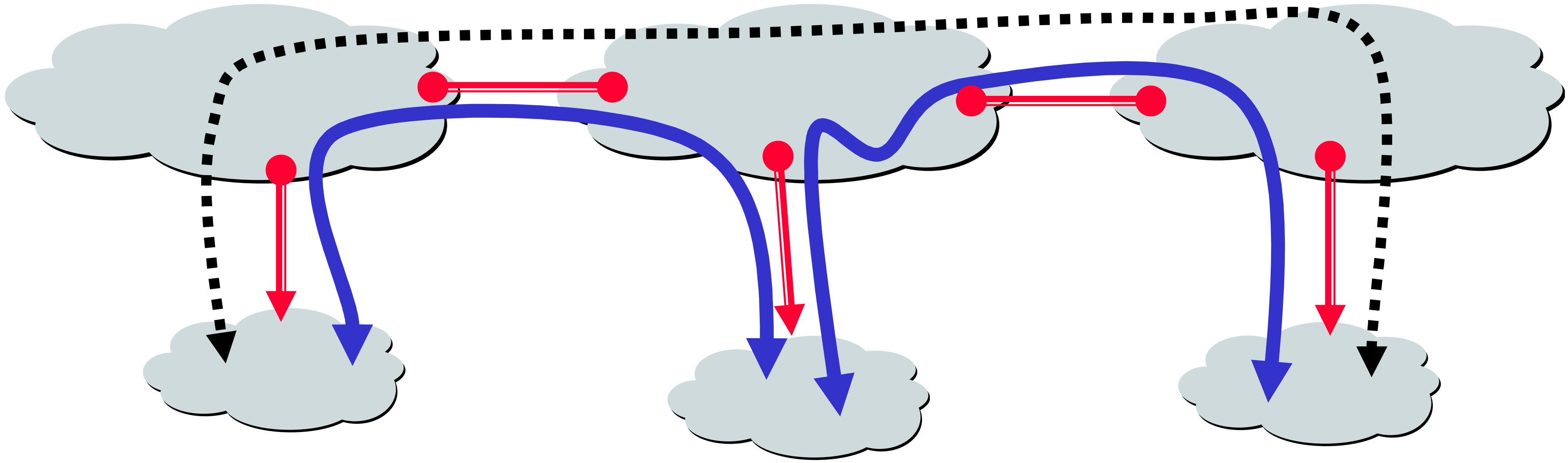


Static routing is the most common way of connecting an autonomous routing domain to the Internet.
This helps explain why BGP is a mystery to many …

Customer-Provider Hierarchy



The Peering Relationship



peer ●══● peer

provider ●══▶ customer

◀━━▶ traffic allowed

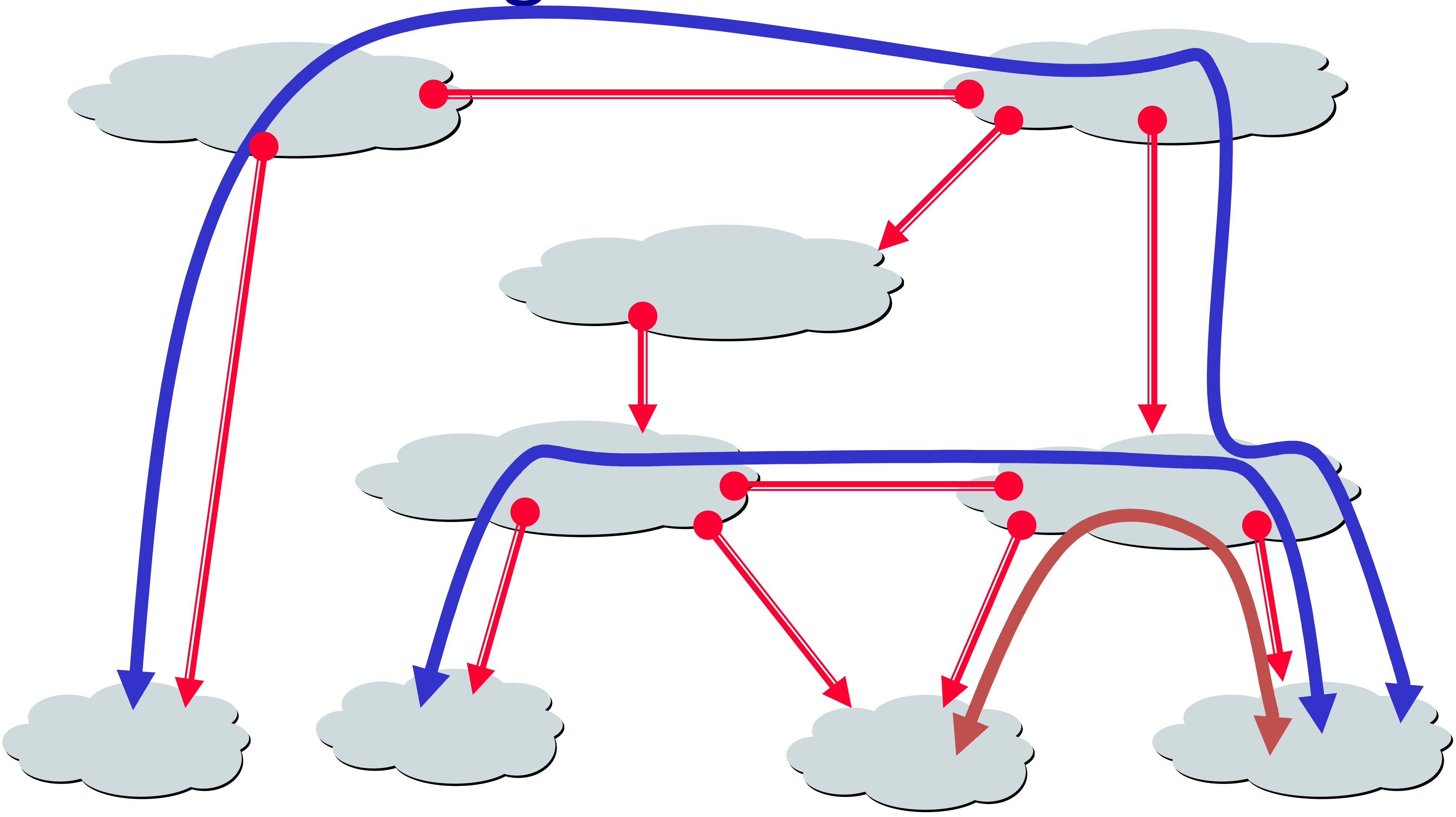
◀■■■■■▶ traffic NOT allowed

Peers provide transit between their respective customers

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

Peers (often) do not exchange $$$

Peering Provides Shortcuts



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

peer ●—● peer

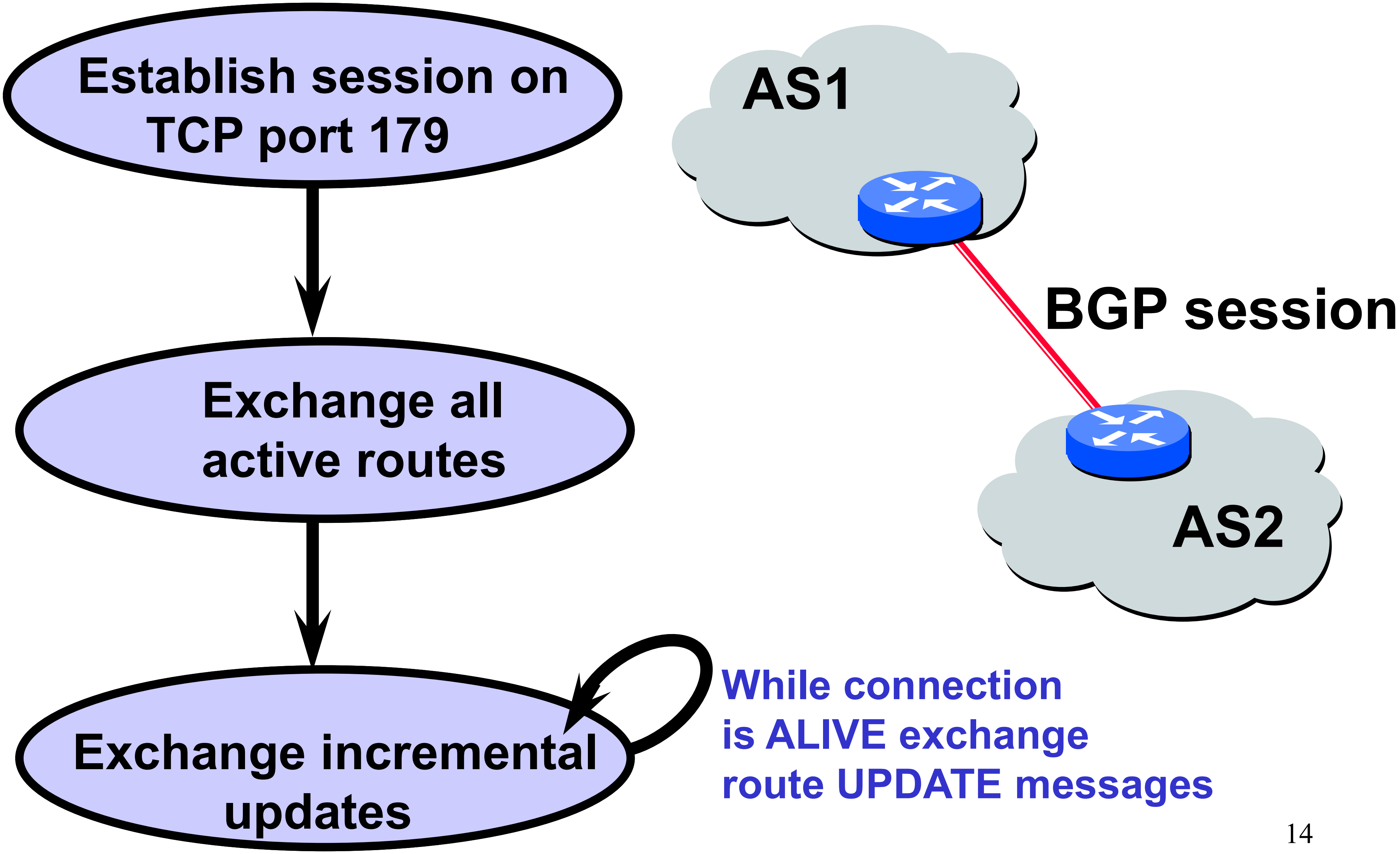
provider ●→ customer

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)



Establish session on TCP port 179

Exchange all active routes

Exchange incremental updates

While connection is ALIVE exchange route UPDATE messages

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

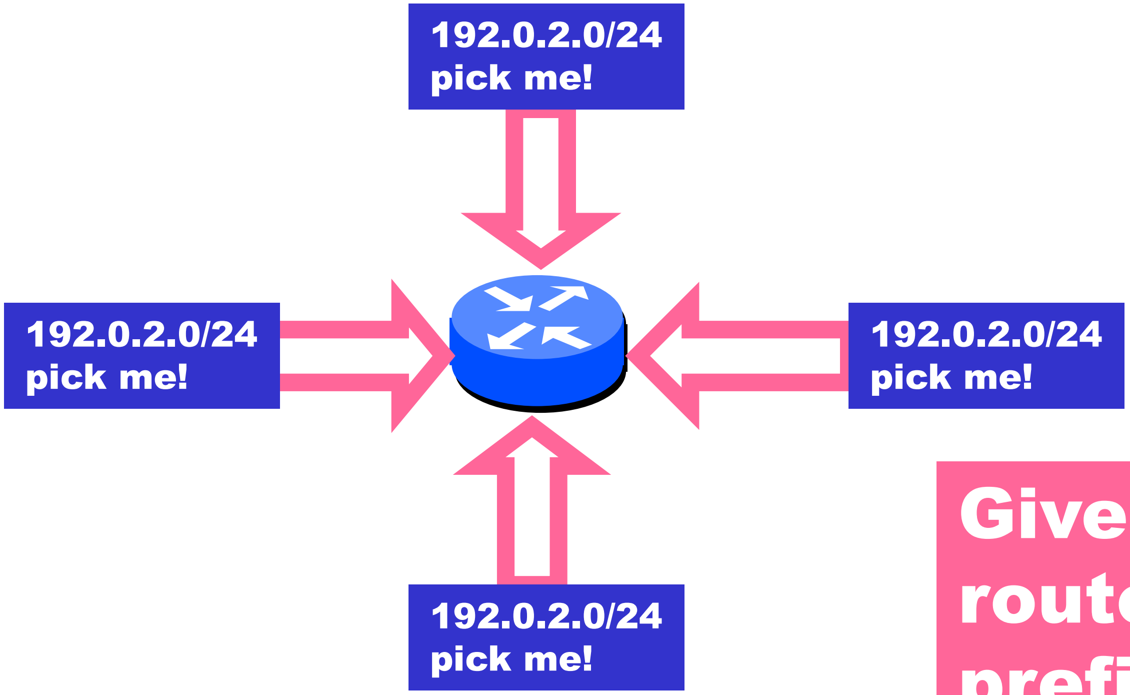
Value	Code	Reference
1	ORIGIN	[RFC1771]
2	AS_PATH	[RFC1771]
3	NEXT_HOP	[RFC1771]
4	MULTI_EXIT_DISC	[RFC1771]
5	LOCAL_PREF	[RFC1771]
6	ATOMIC_AGGREGATE	[RFC1771]
7	AGGREGATOR	[RFC1771]
8	COMMUNITY	[RFC1997]
9	ORIGINATOR_ID	[RFC2796]
10	CLUSTER_LIST	[RFC2796]
11	DPA	[Chen]
12	ADVERTISER	[RFC1863]
13	RCID_PATH / 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



192.0.2.0/24
pick me!

192.0.2.0/24
pick me!

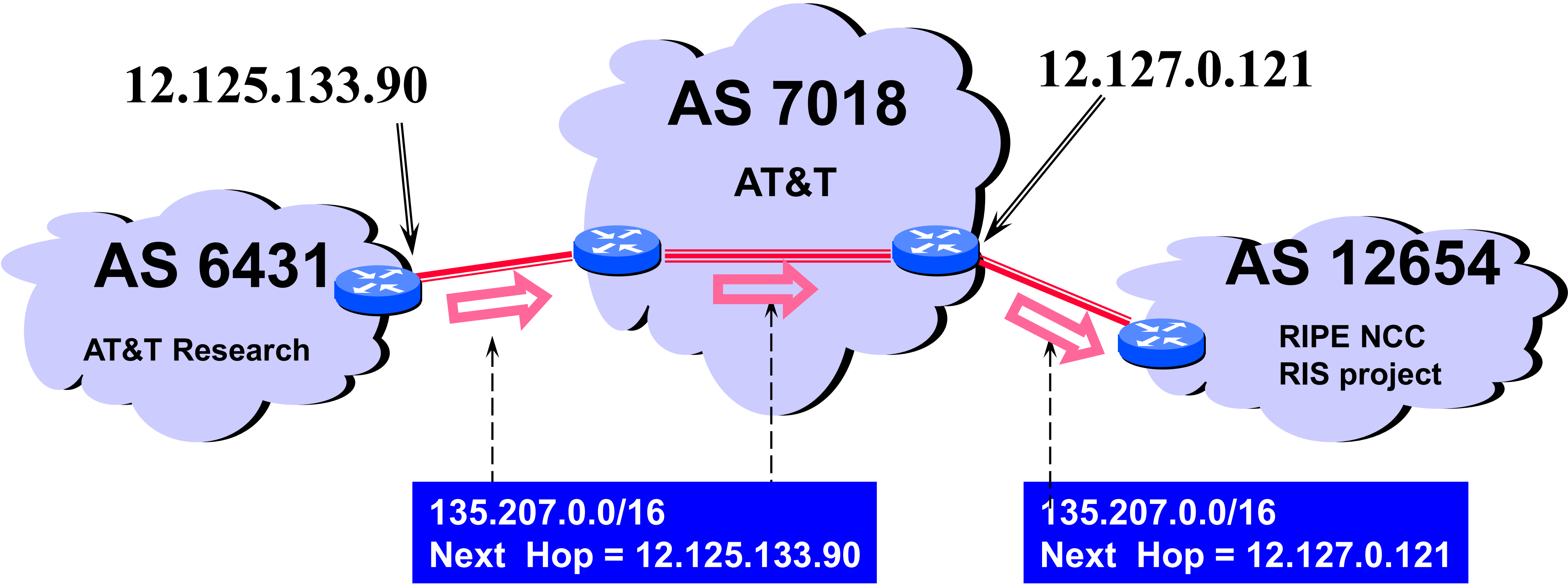
192.0.2.0/24
pick me!

192.0.2.0/24
pick me!

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!)

BGP Next Hop Attribute

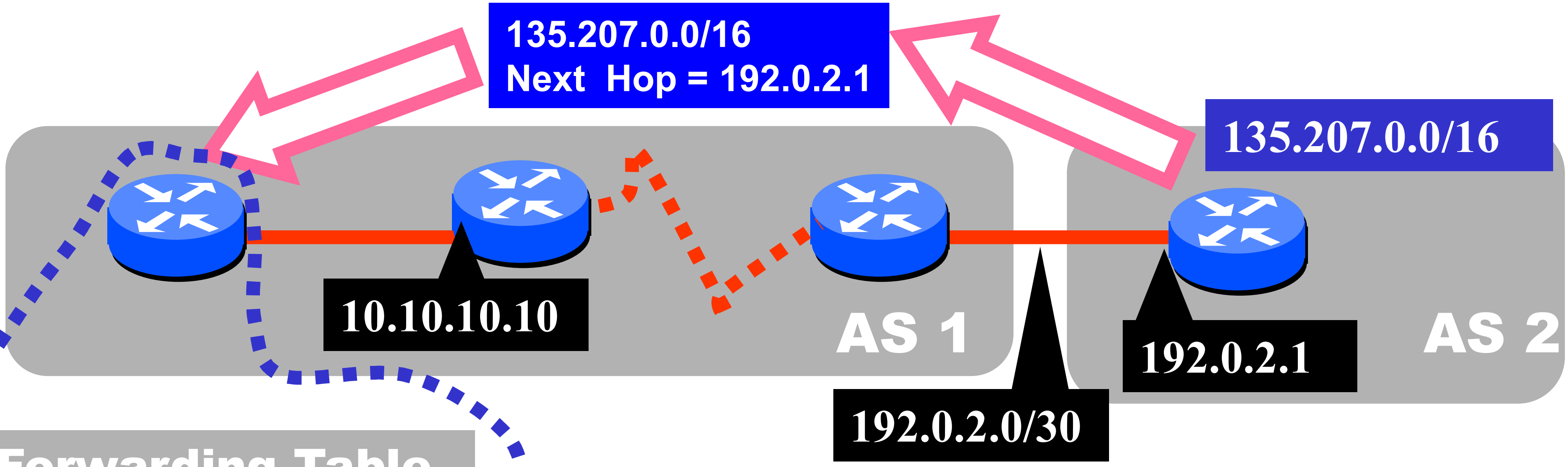


135.207.0.0/16
Next Hop = 12.125.133.90

135.207.0.0/16
Next Hop = 12.127.0.121

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.

Join EGP with IGP For Connectivity



Forwarding Table

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

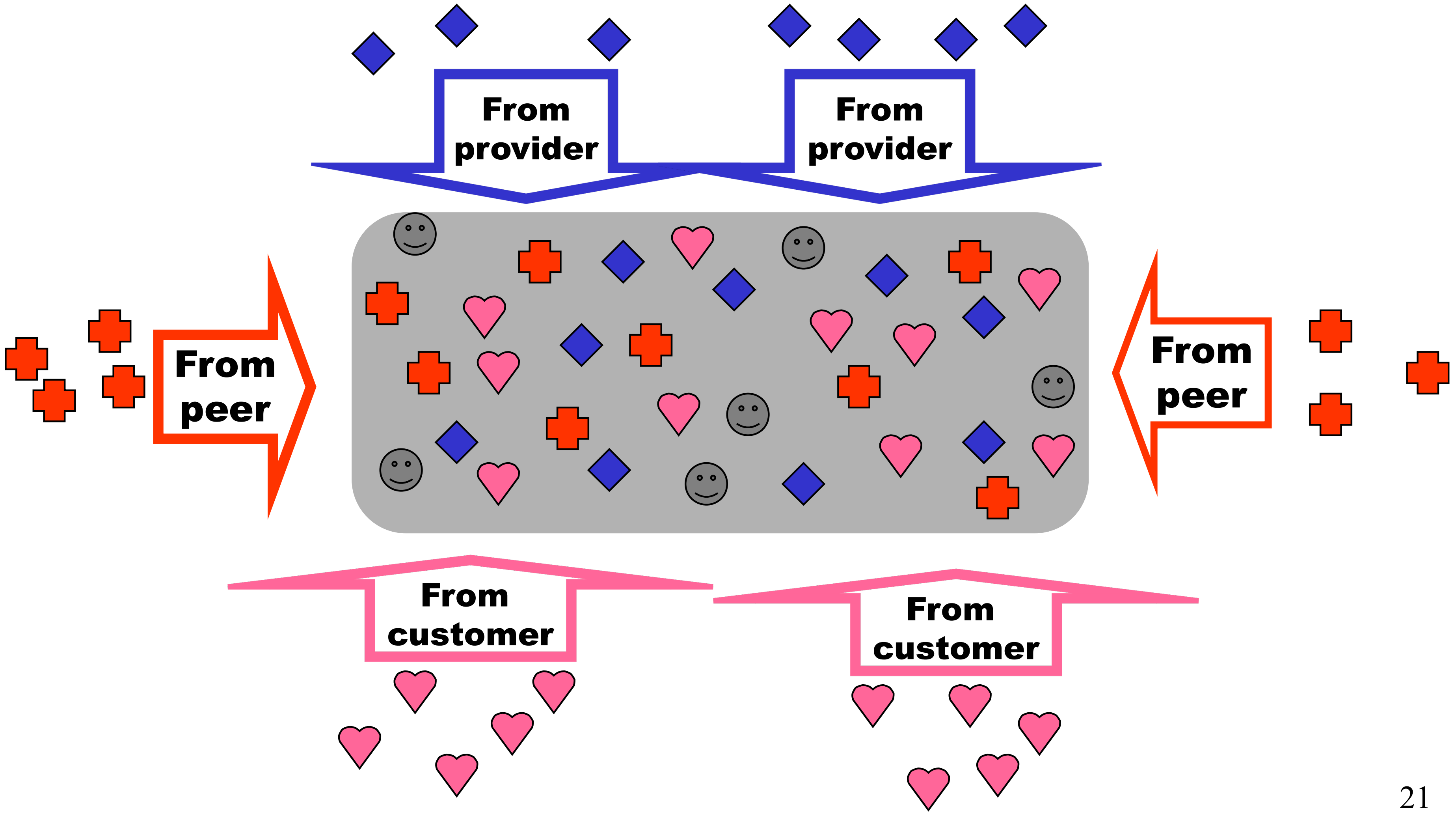
Implementing Customer/Provider and Peer/Peer relationships

Two parts:

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

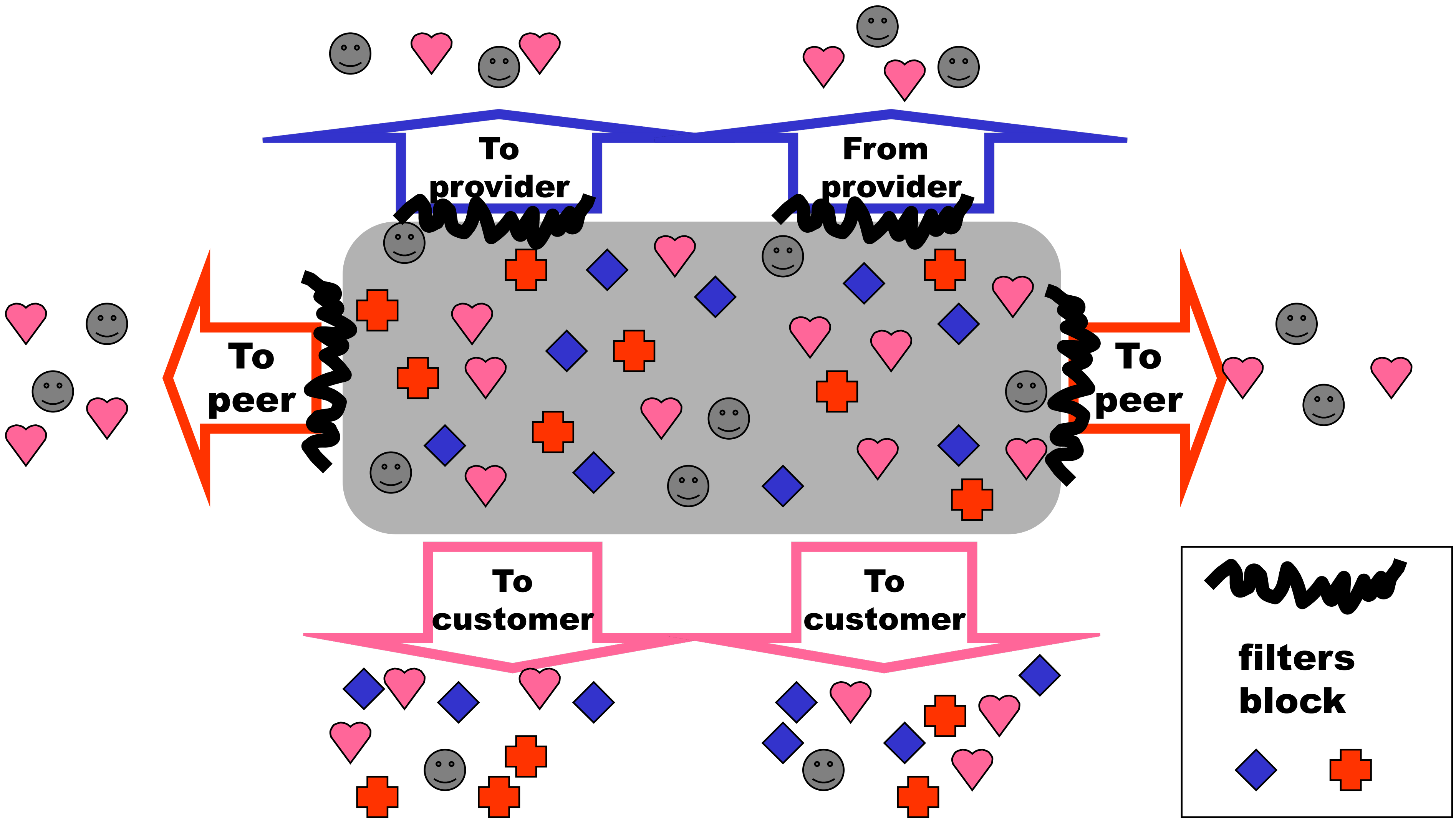
Import Routes

◆ **provider route** ✚ **peer route** ♥ **customer route** ☺ **ISP route**



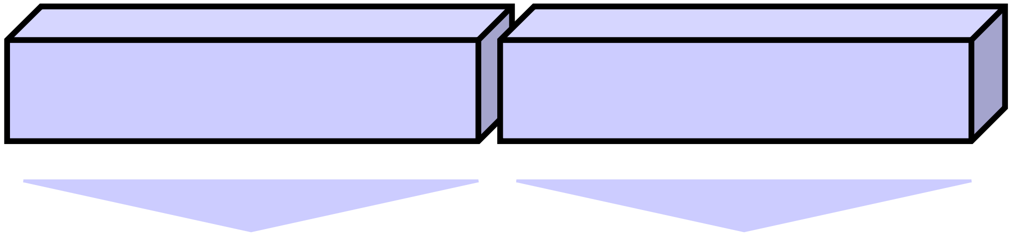
Export Routes

◆ **provider route** ✚ **peer route** ♥ **customer route** ☺ **ISP route**



How Can Routes be Colored? BGP Communities!

A community value is 32 bits



By convention, first 16 bits is ASN indicating who is giving it an interpretation

community number

Used for signalling within and between ASes

Very powerful BECAUSE it has no (predefined) meaning

Community Attribute = a list of community values. (So one route can belong to multiple communities)

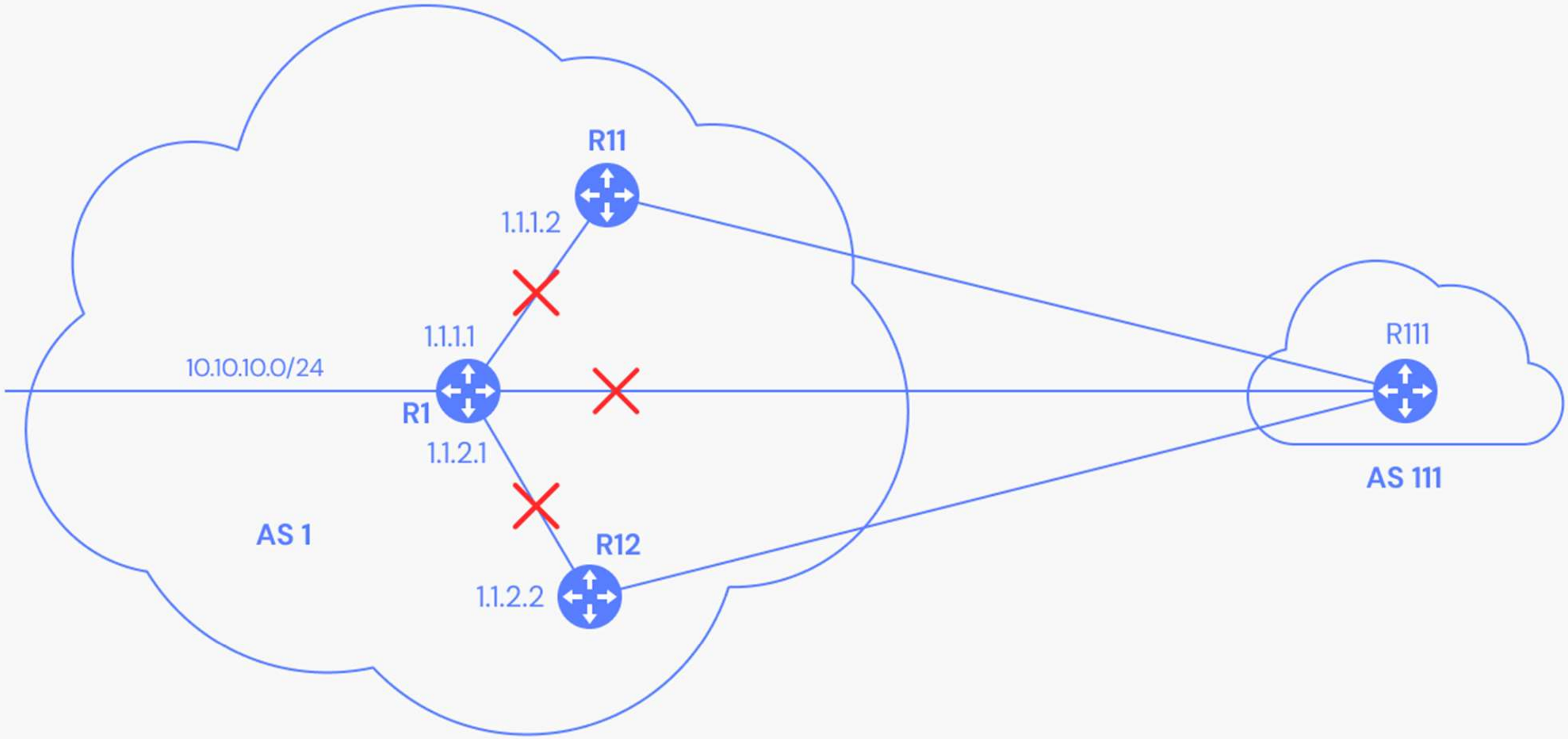
Two reserved communities

no_export = 0xFFFFFF01: don't export out of AS

no_advertise 0xFFFFFF02: don't pass to BGP neighbors

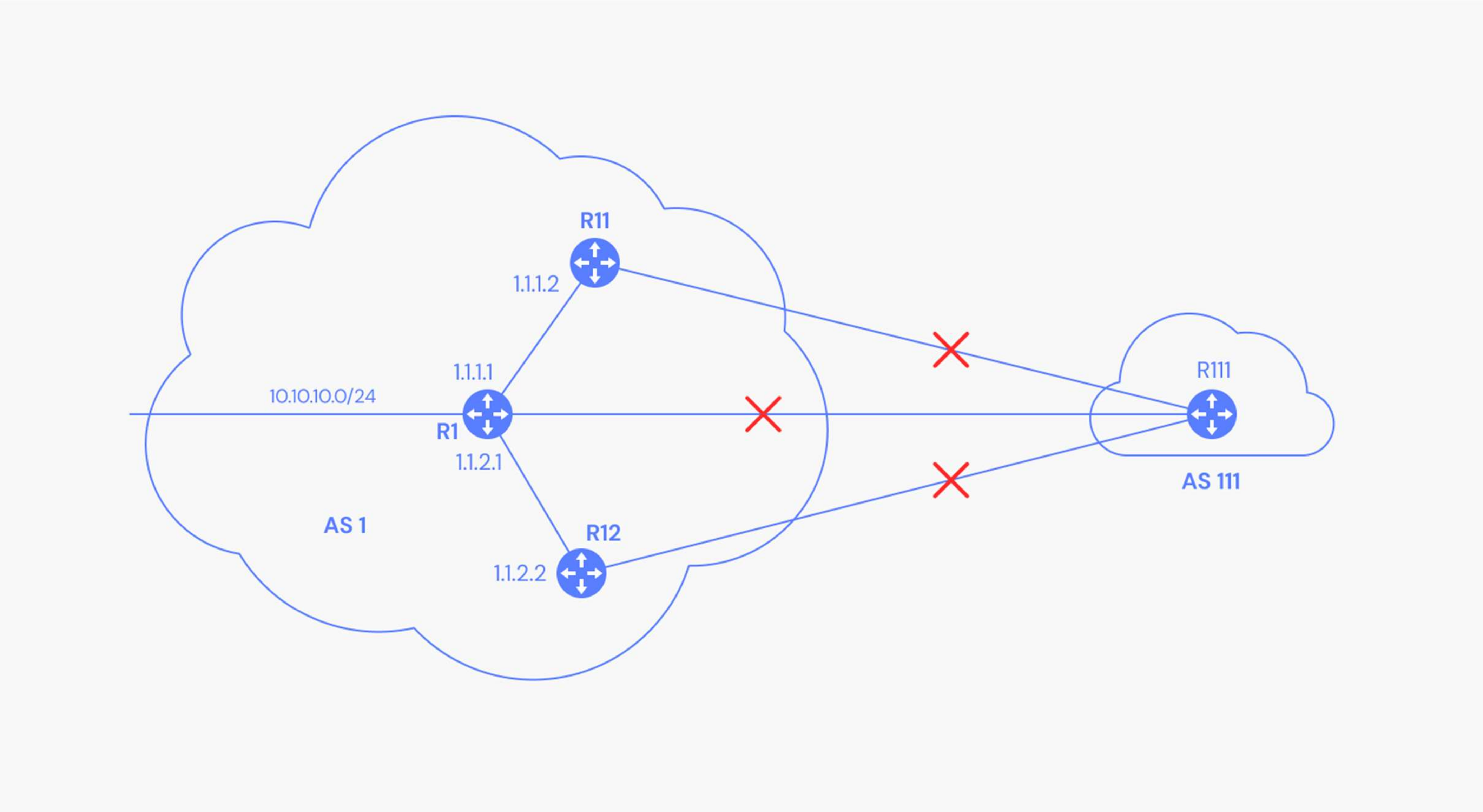
RFC 1997 (August 1996)

Community attribute: No_Advertise



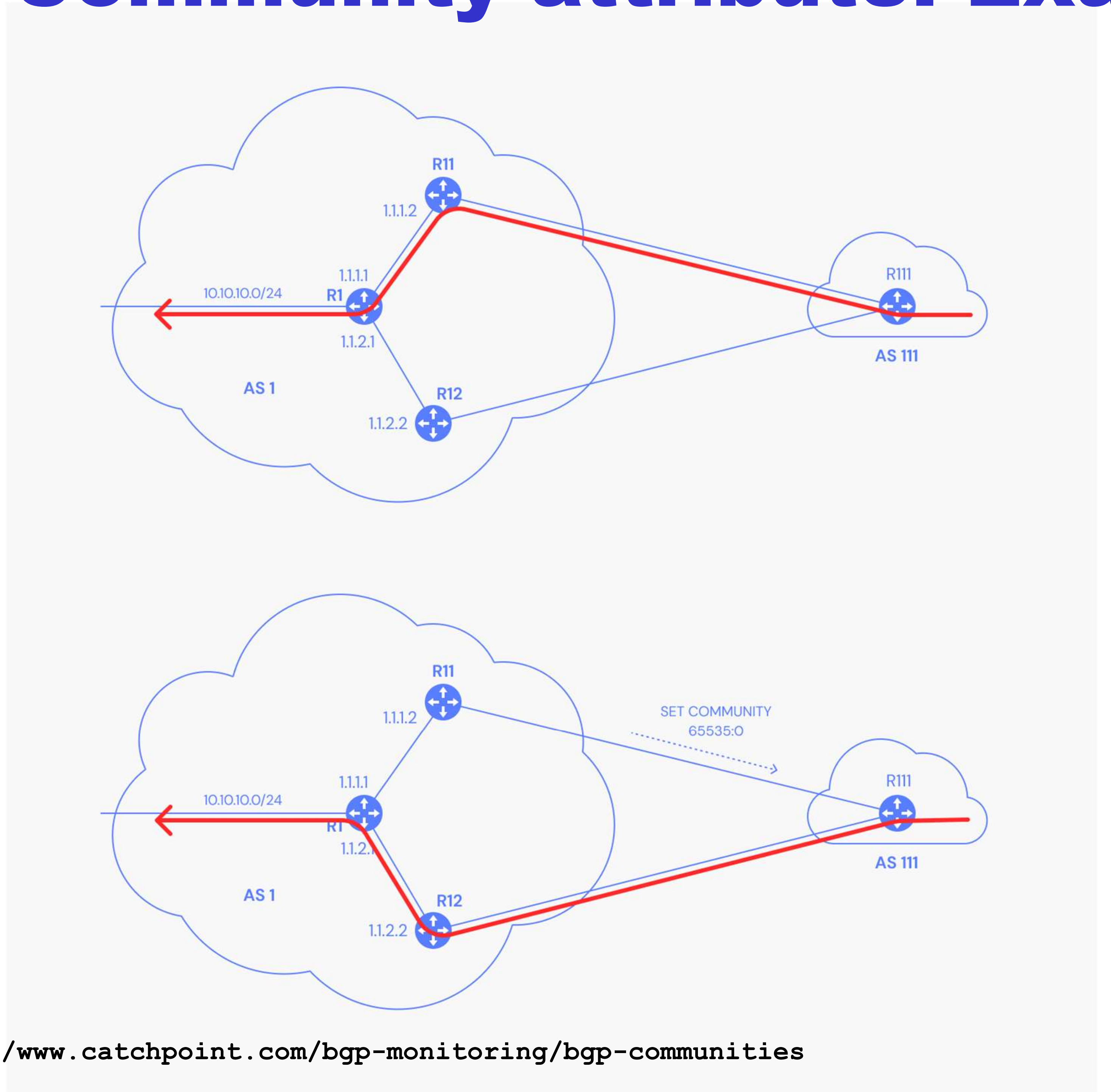
Pic: https://www.catchpoint.com/bgp-monitoring/bgp-communities

Community attribute: No_Export



Pic: https://www.catchpoint.com/bgp-monitoring/bgp-communities

BGP Community attribute: Example

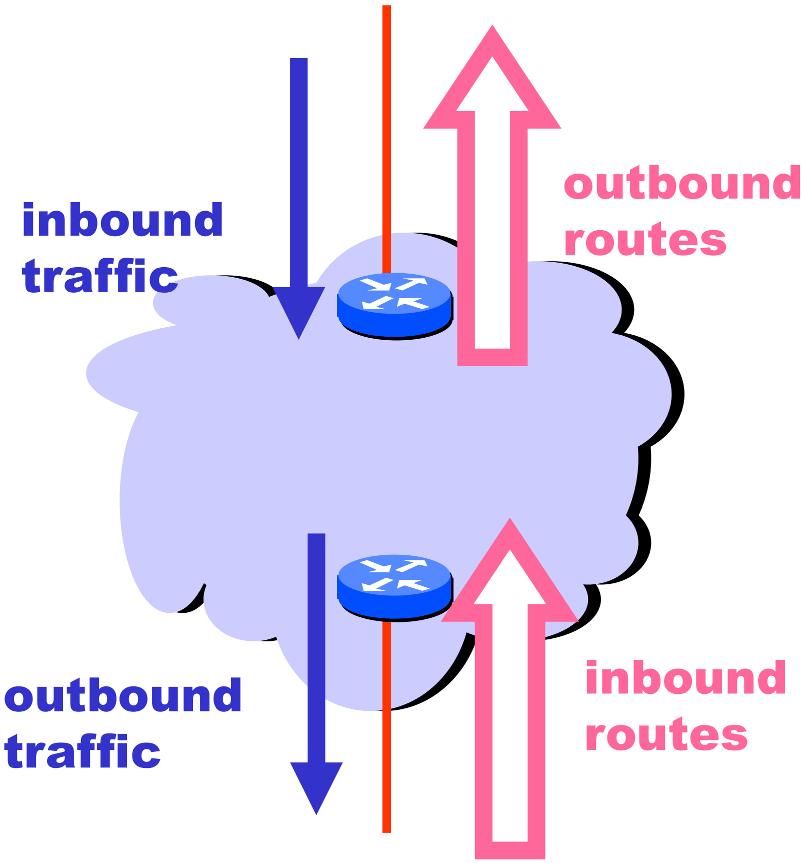


Pic: https://www.catchpoint.com/bgp-monitoring/bgp-communities

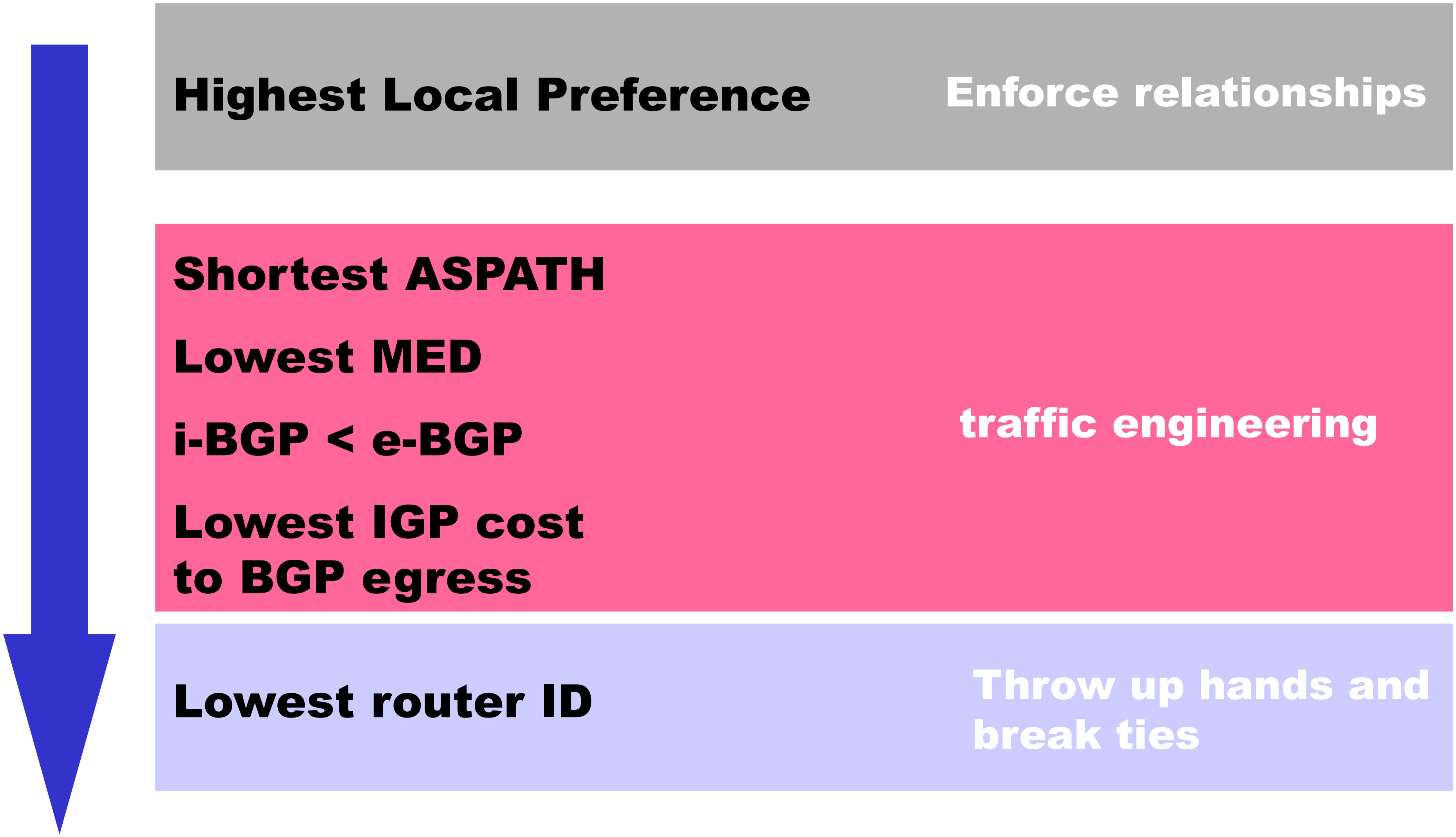
Tweak Tweak Tweak

- For inbound traffic
  - Filter outbound routes
  - Tweak attributes on outbound routes in the hope of influencing your neighbor’s best route selection
- For outbound traffic
  - Filter inbound routes
  - Tweak attributes on inbound routes to influence best route selection

In general, an AS has more control over outbound traffic



Route Selection Summary



Highest Local Preference

Enforce relationships

Shortest ASPATH

Lowest MED

i-BGP < e-BGP

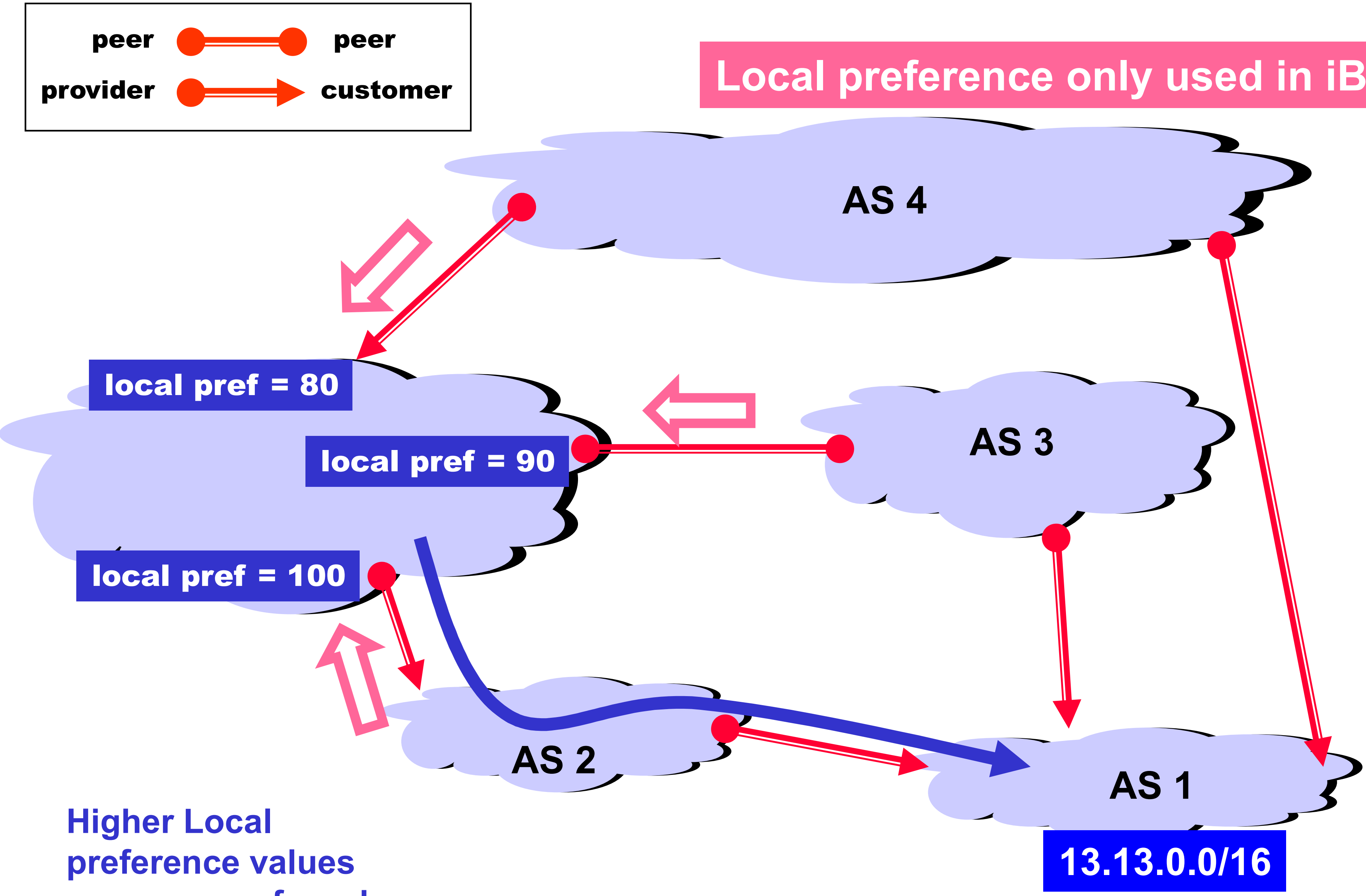
Lowest IGP cost to BGP egress

traffic engineering

Lowest router ID

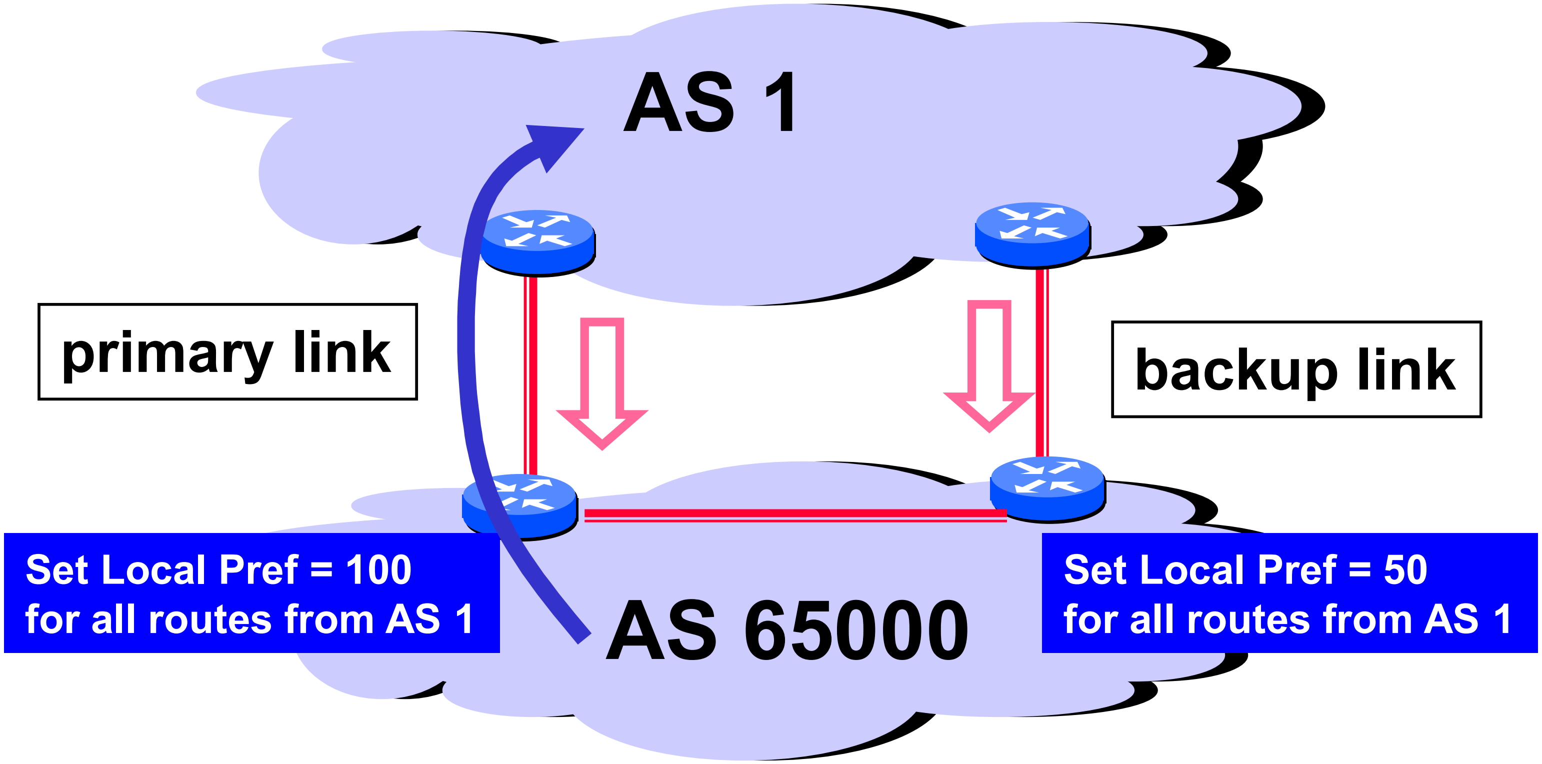
Throw up hands and break ties

Local Preference Attribute



Higher Local preference values are more preferred

Implementing Backup Links with Local Preference (Outbound Traffic)

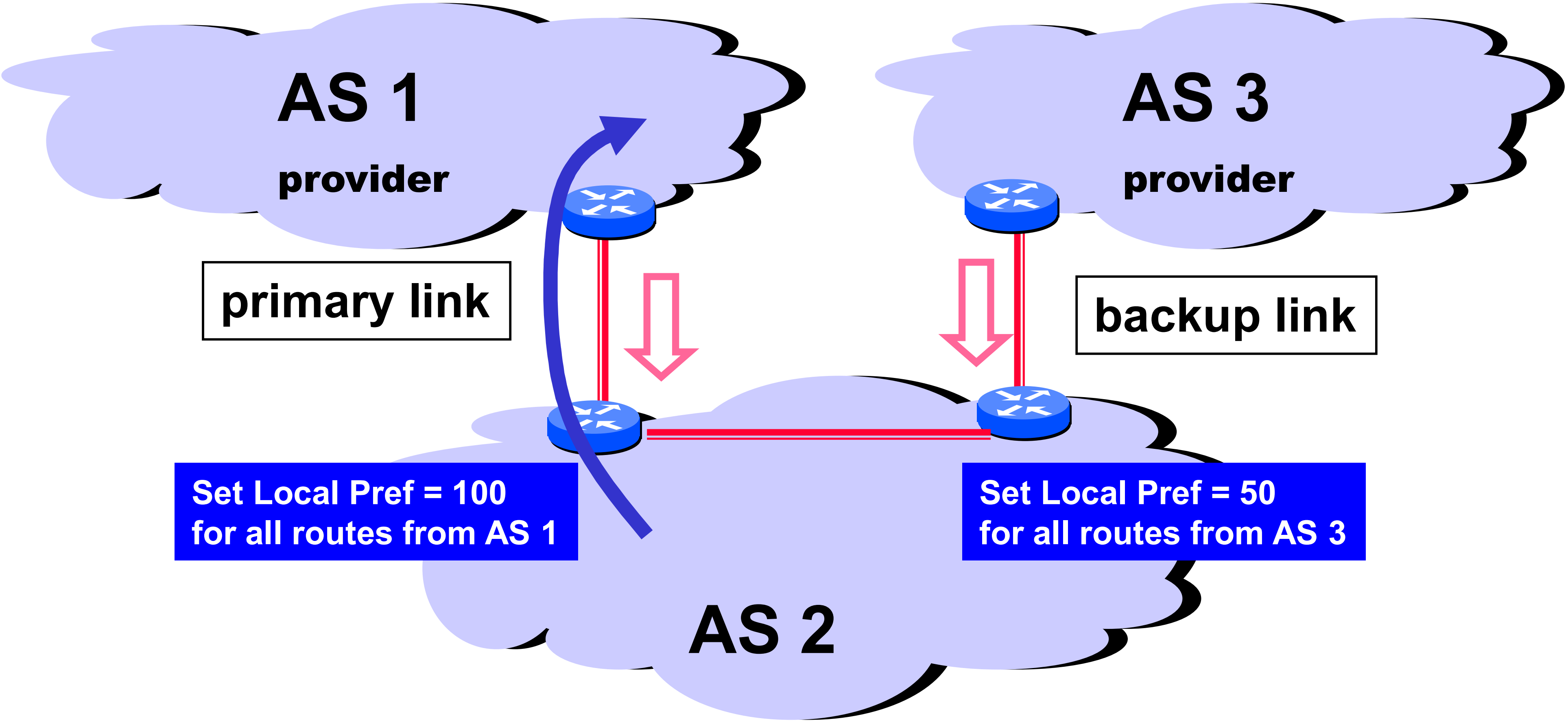


Set Local Pref = 100
for all routes from AS 1

Set Local Pref = 50
for all routes from AS 1

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

Multihomed Backups (Outbound Traffic)

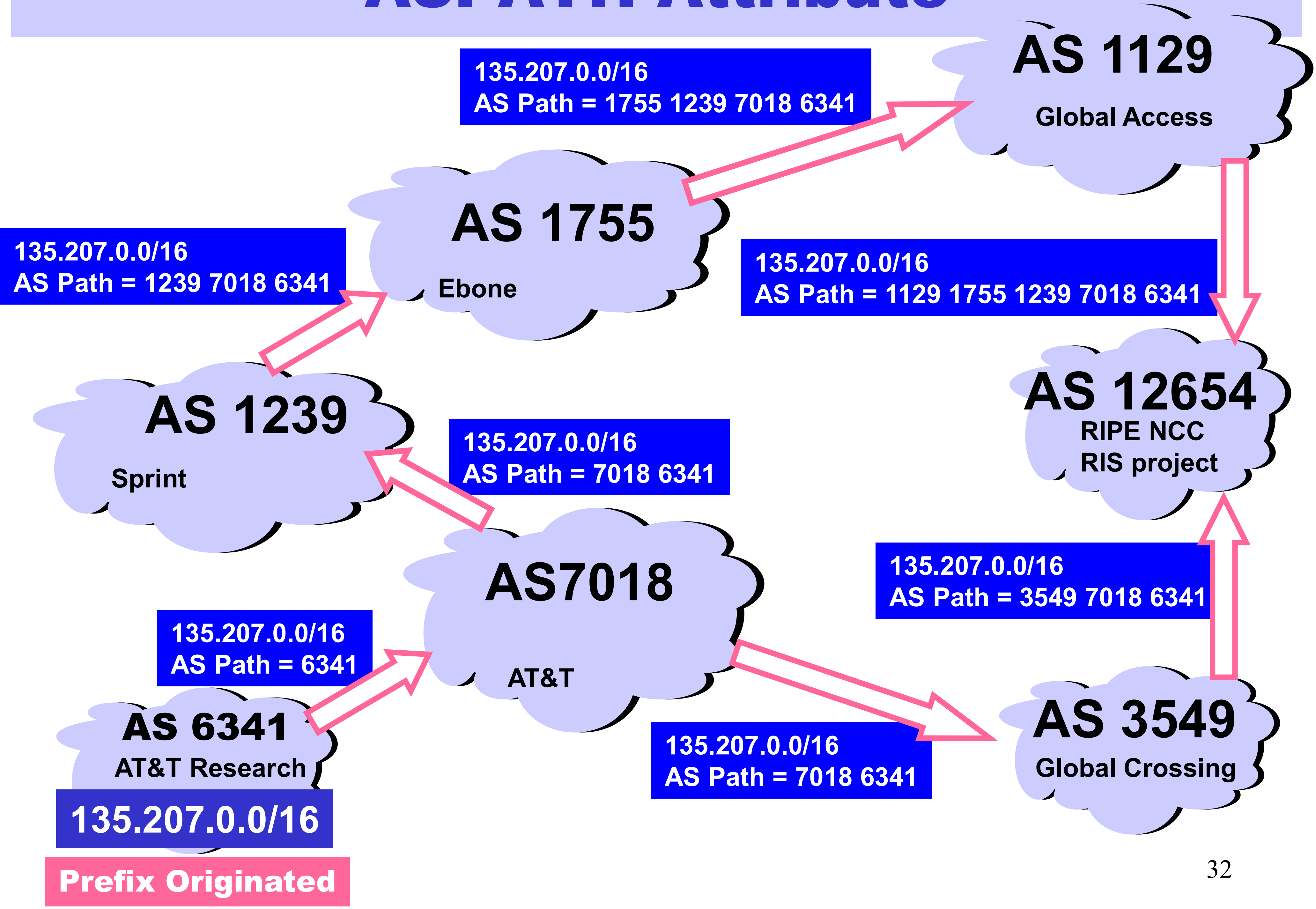


Set Local Pref = 100
for all routes from AS 1

Set Local Pref = 50
for all routes from AS 3

Forces <u>outbound </u>traffic to take primary link, unless link is down.

ASPATH Attribute



AS 6341
AT&T Research

135.207.0.0/16

Prefix Originated

135.207.0.0/16
AS Path = 6341

AS7018
AT&T

135.207.0.0/16
AS Path = 7018 6341

AS 1239
Sprint

135.207.0.0/16
AS Path = 1239 7018 6341

AS 1755
Ebone

135.207.0.0/16
AS Path = 1755 1239 7018 6341

AS 1129
Global Access

135.207.0.0/16
AS Path = 1129 1755 1239 7018 6341

135.207.0.0/16
AS Path = 7018 6341

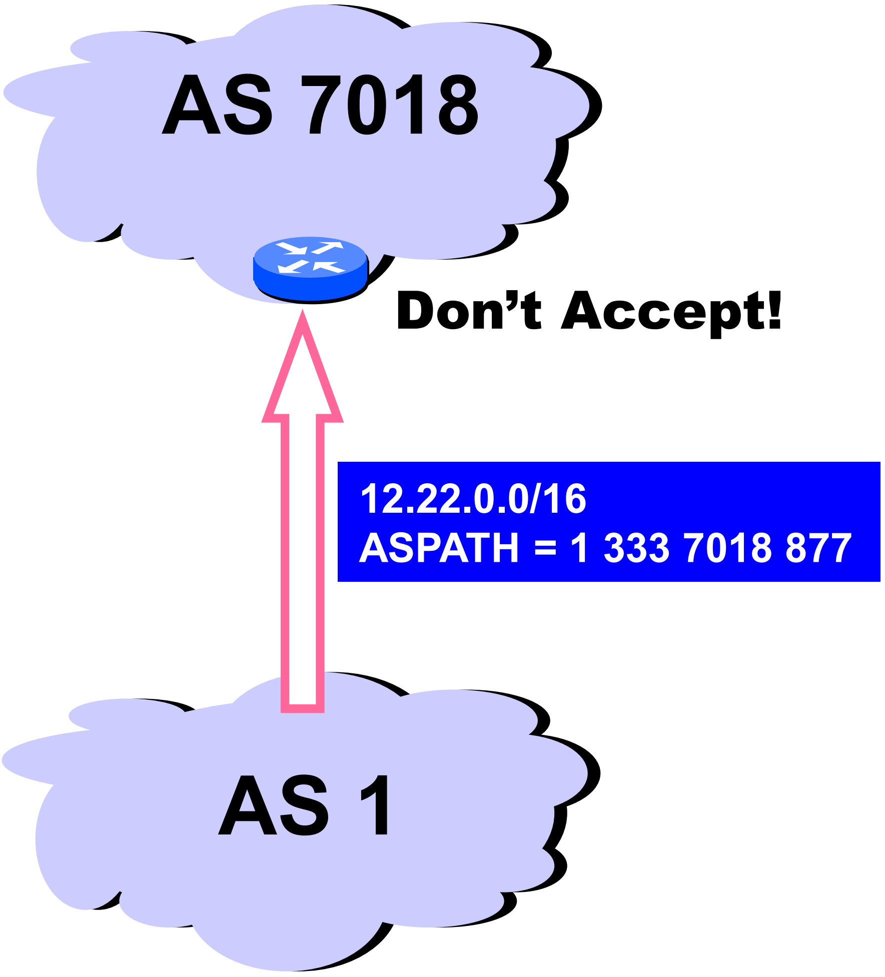
AS 3549
Global Crossing

135.207.0.0/16
AS Path = 3549 7018 6341

AS 12654
RIPE NCC
RIS project

Interdomain Loop Prevention

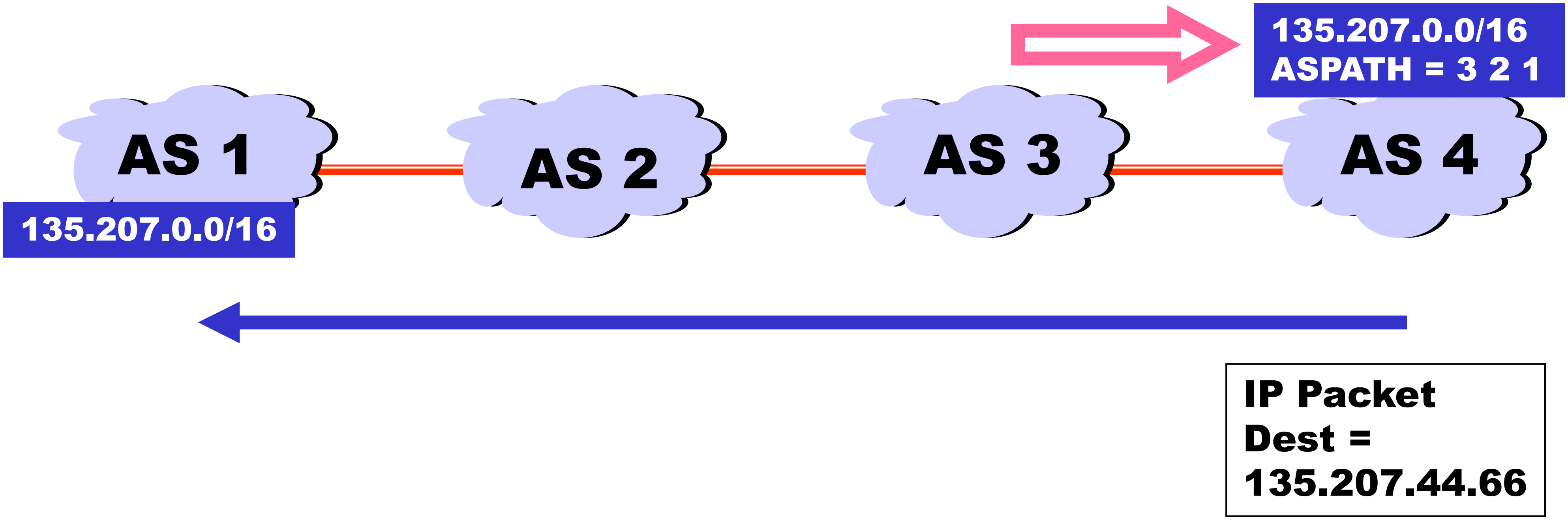
BGP at AS YYY will never accept a route with ASPATH containing YYY.



Don’t Accept!

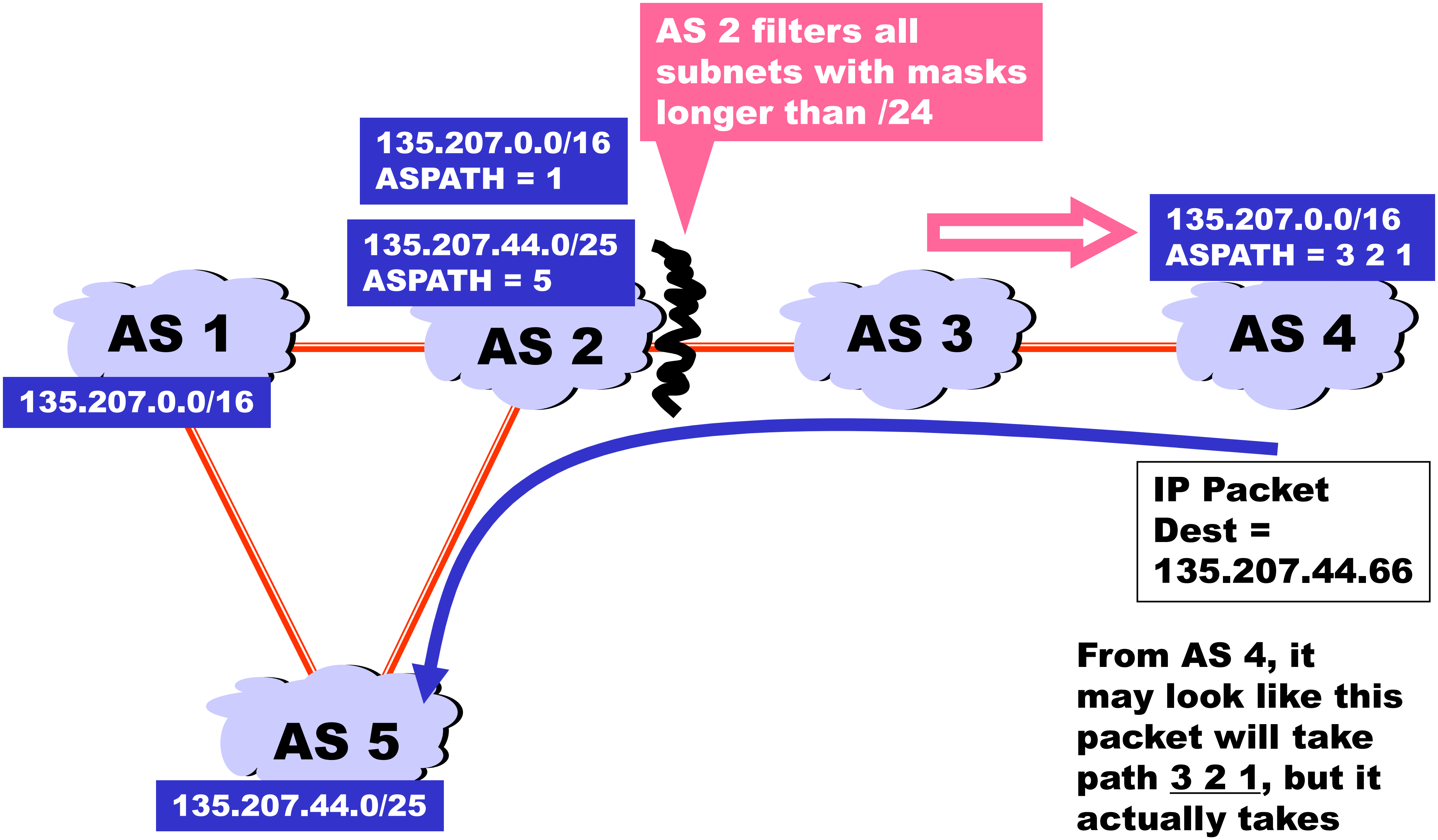
12.22.0.0/16
ASPATH = 1 333 7018 877

Traffic Often Follows ASPATH



IP Packet
Dest =
135.207.44.66

... But It Might Not

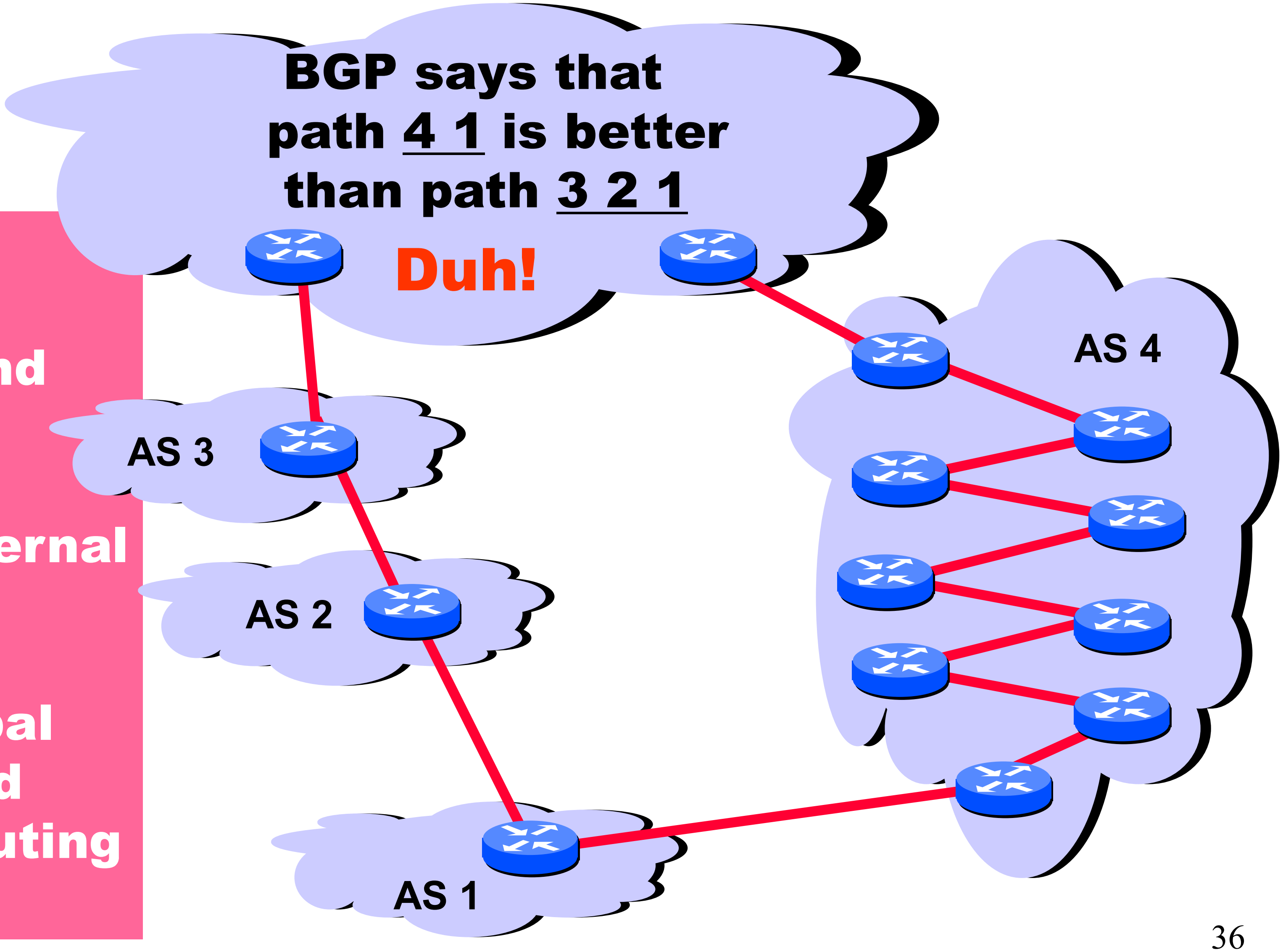


From AS 4, it may look like this packet will take path 3 2 1, but it actually takes path 3 2 5

Shorter Doesn’t Always Mean Shorter

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

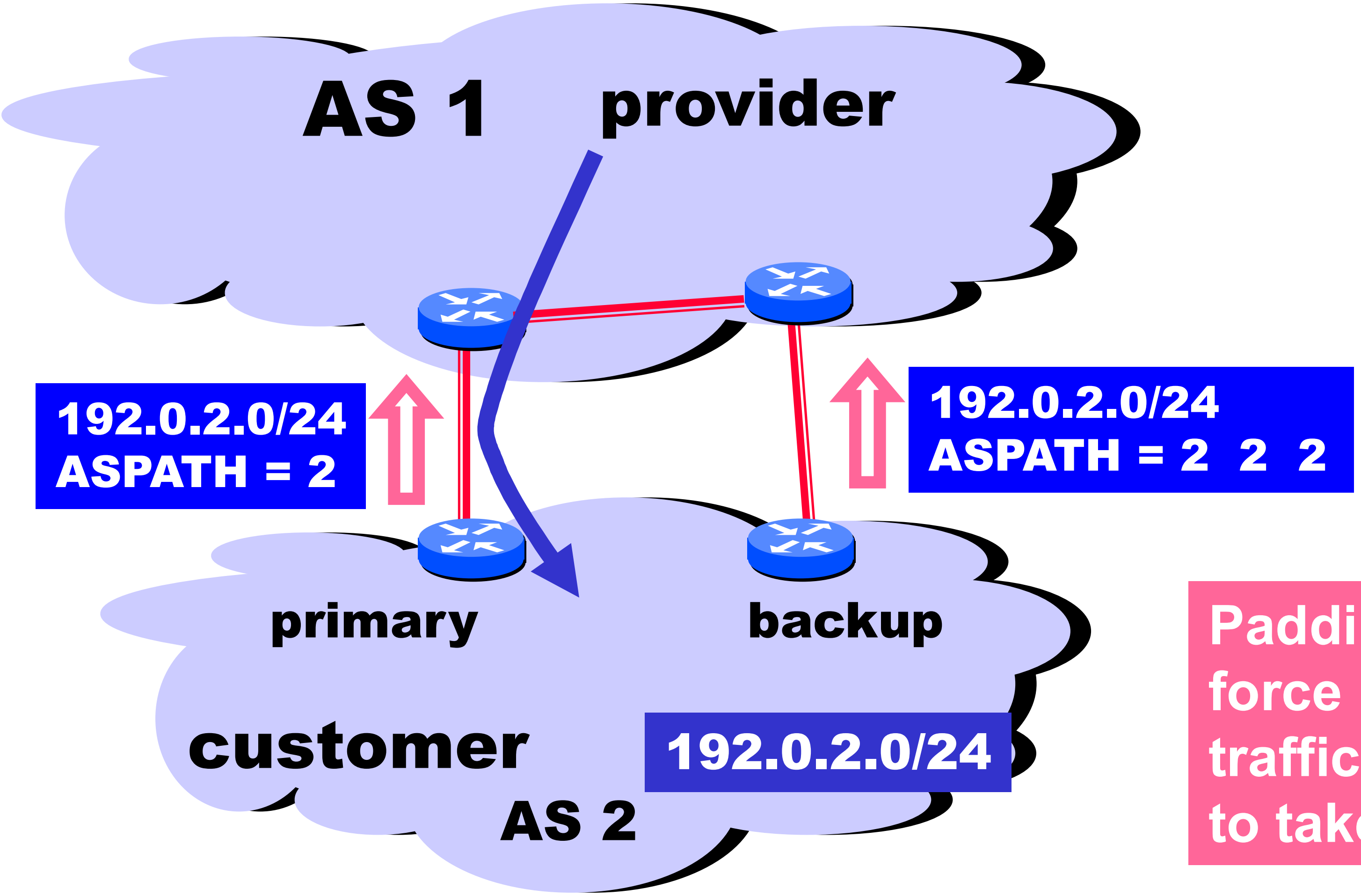
Exporting internal state would dramatically increase global instability and amount of routing state



BGP says that path 4 1 is better than path 3 2 1

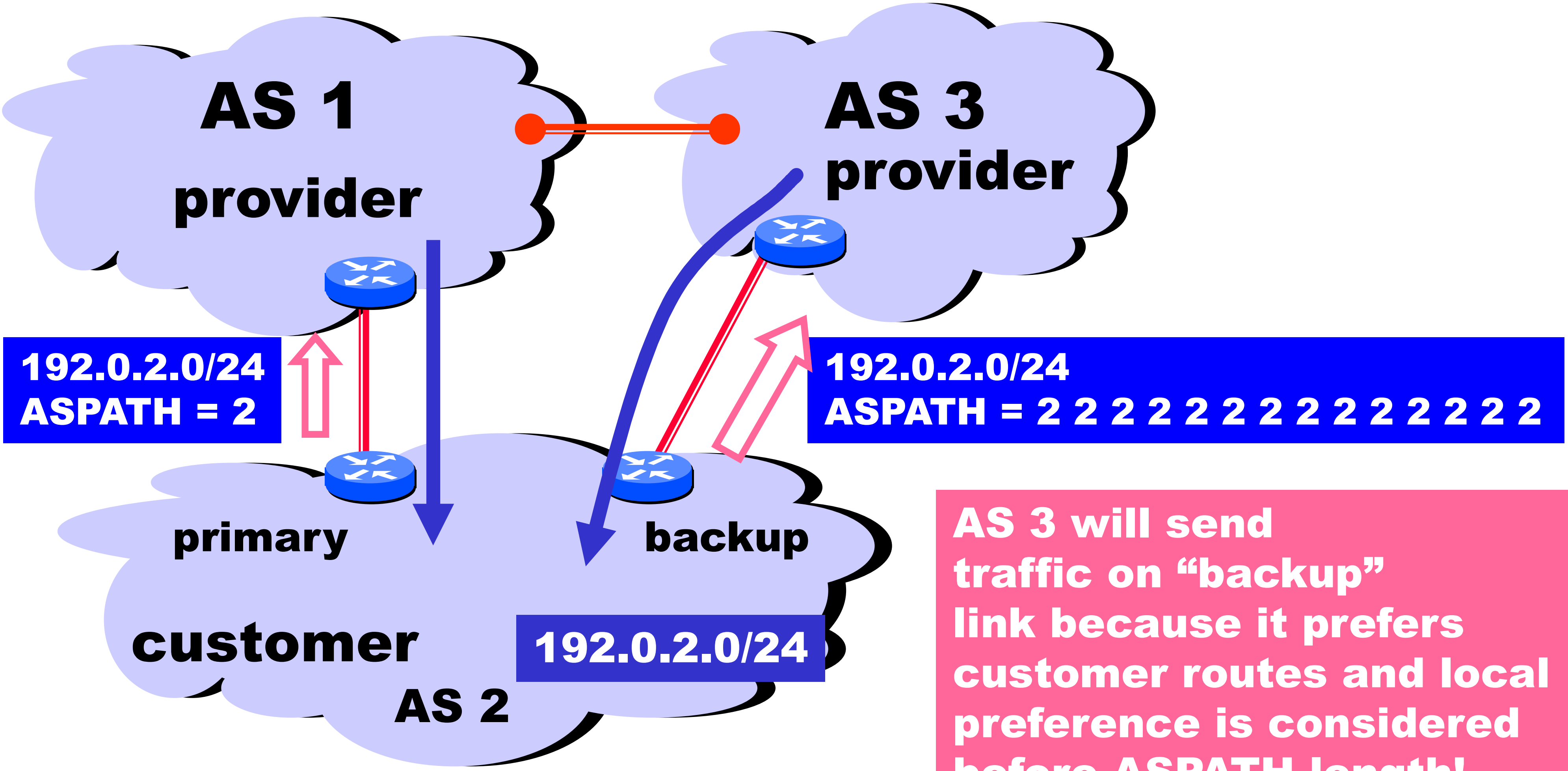
Duh!

Shedding Inbound Traffic with ASPATH Padding Hack



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

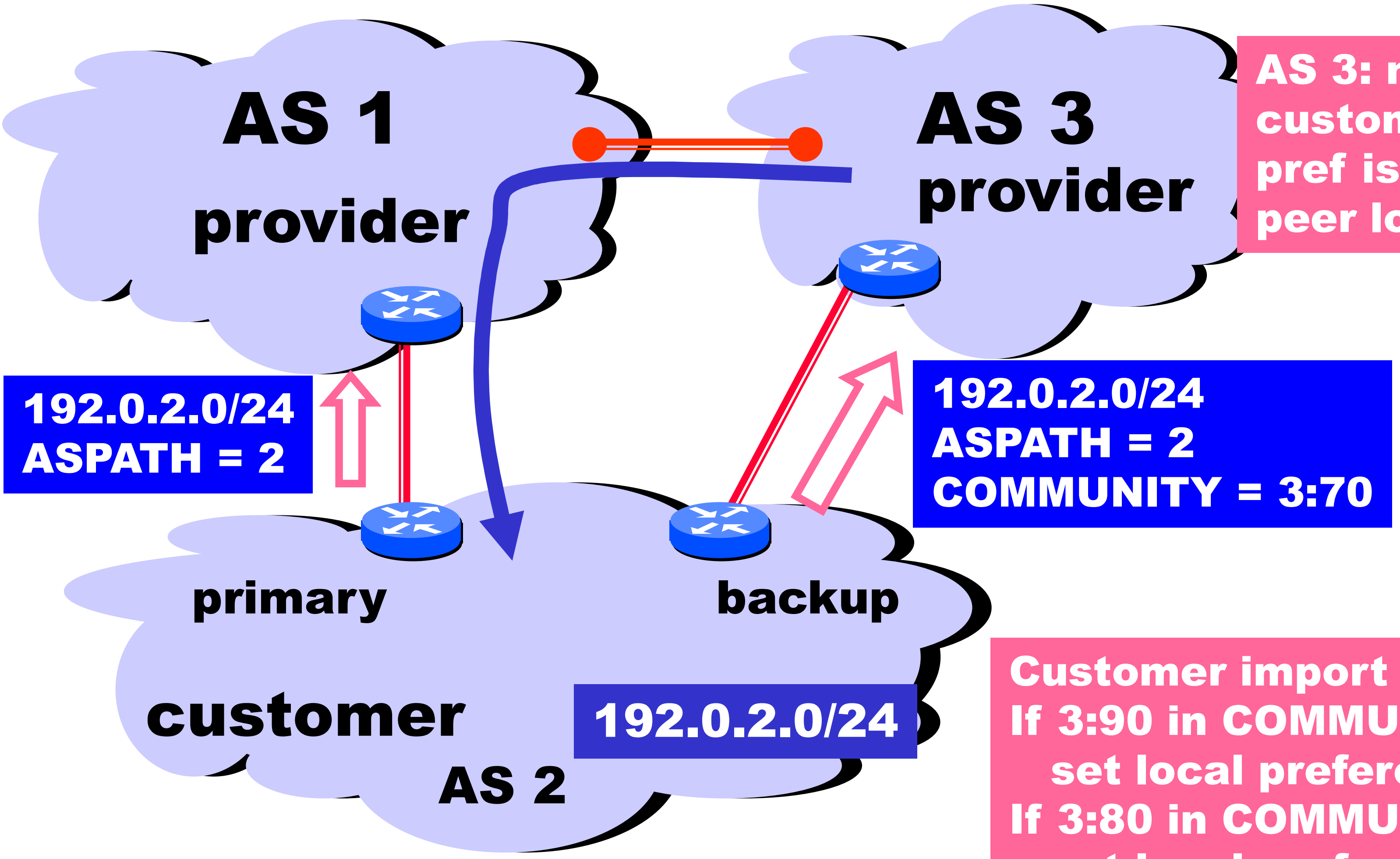
Padding May Not Shut Off All Traffic



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

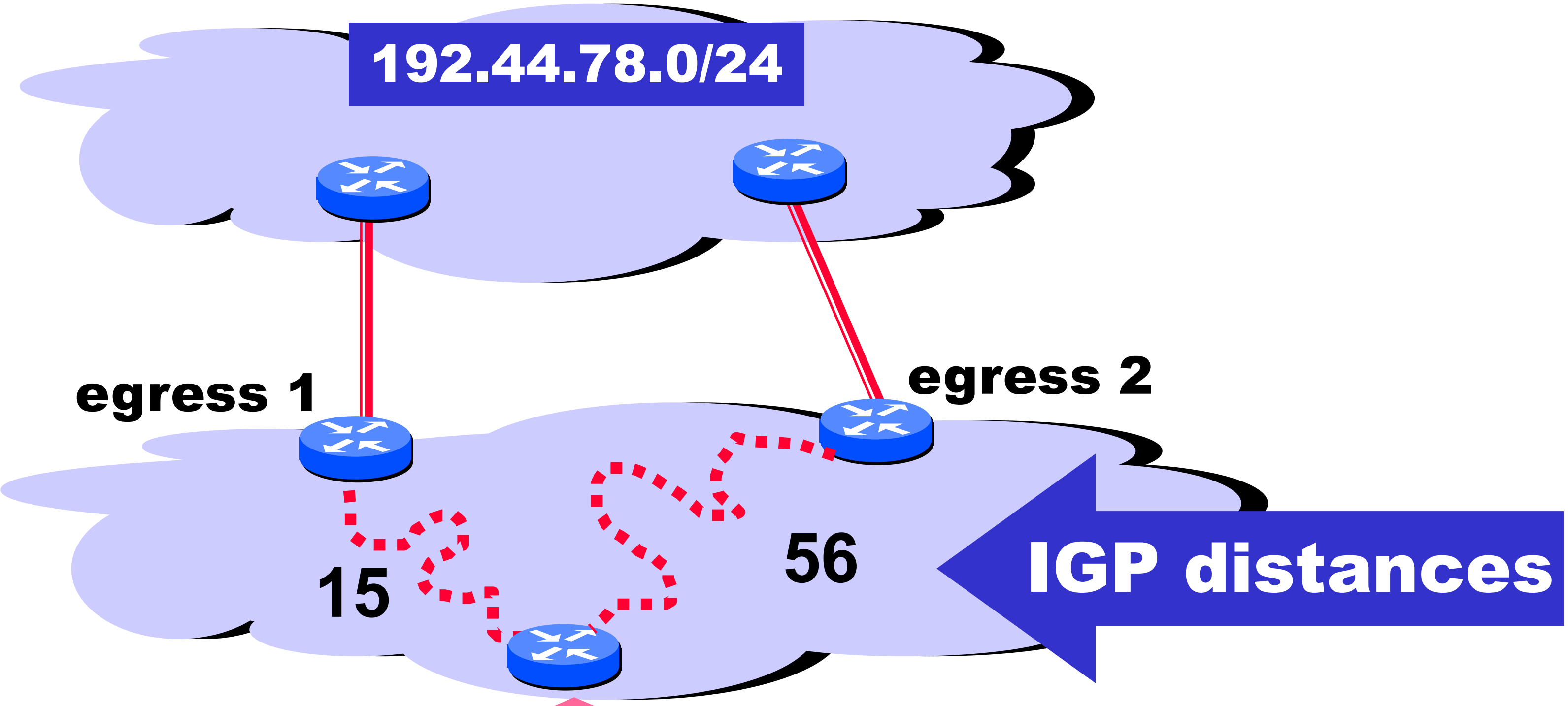
COMMUNITY Attribute to the Rescue!



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

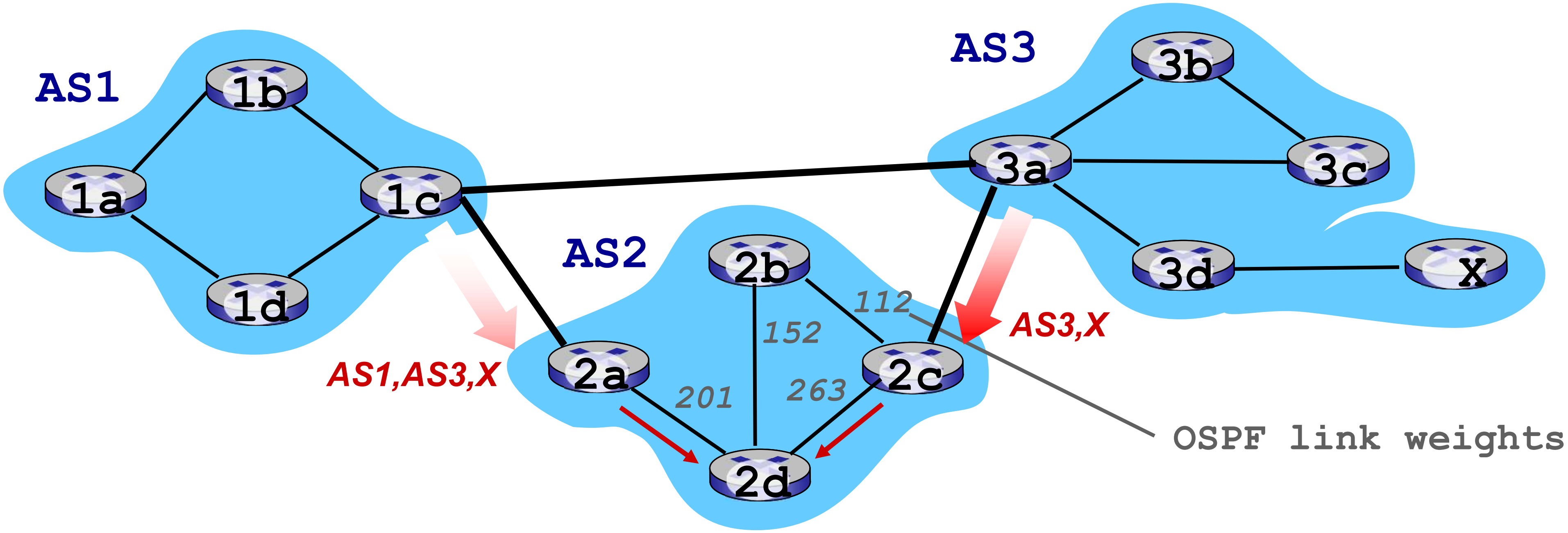
Hot Potato Routing: Go for the Closest Egress Point



This Router has two BGP routes to 192.44.78.0/24.

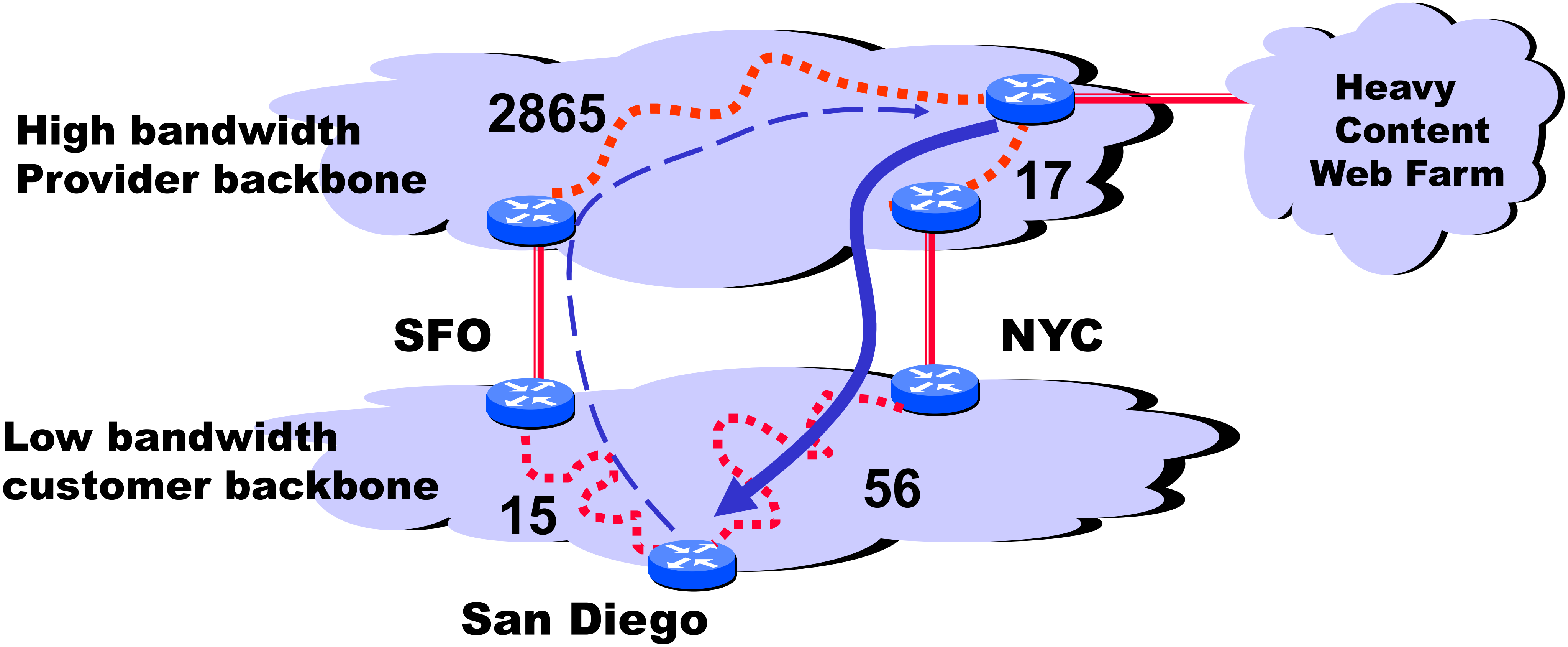
Hot potato: get traffic off of your network as
Soon as possible. Go for egress 1!

Hot Potato Routing



- 2d learns (via iBGP) it can route to X via 2a or 2c
- ***hot potato routing:*** choose local gateway that has least intra-domain cost (e.g., 2d chooses 2a, even though more AS hops to *X*): don't worry about inter-domain cost!

Getting Burned by the Hot Potato



High bandwidth Provider backbone

2865

17

Heavy Content Web Farm

SFO

NYC

Low bandwidth customer backbone

15

56

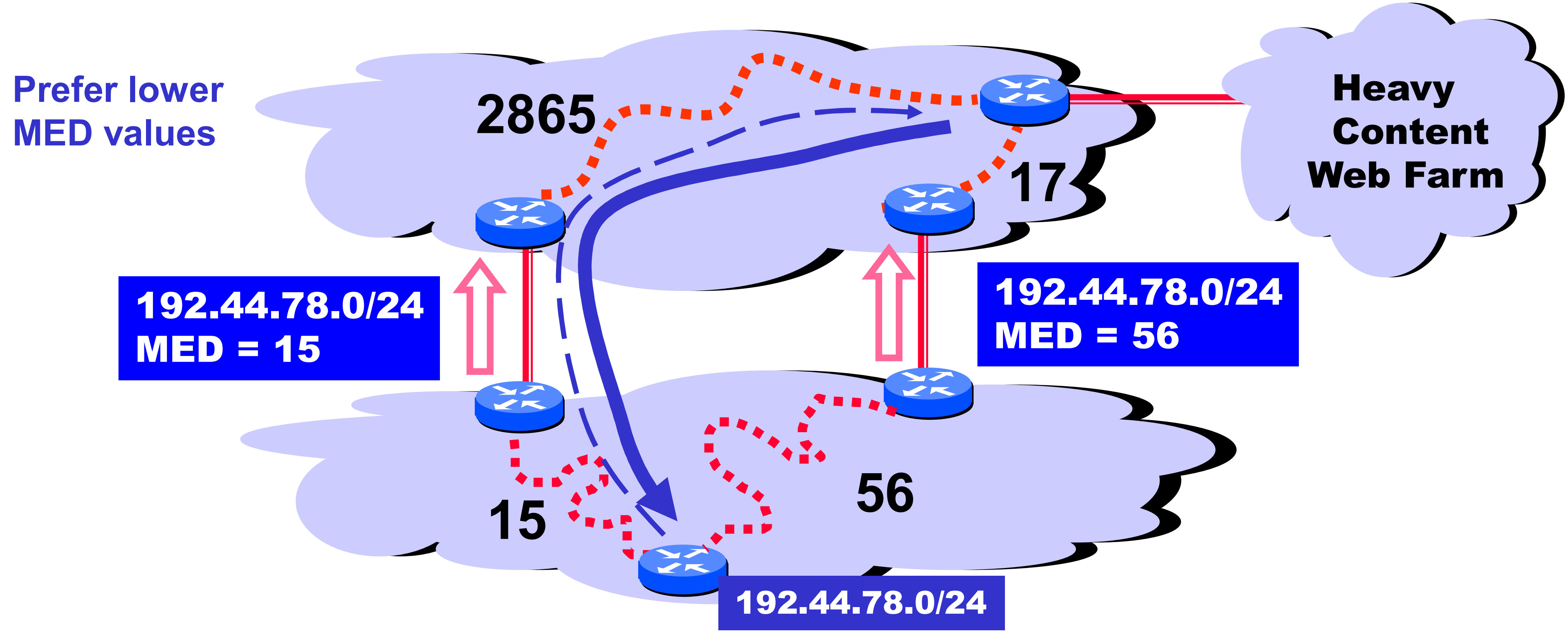
San Diego

Many customers want their provider to carry the bits!

- - - → tiny http request

━━➡ huge http reply

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

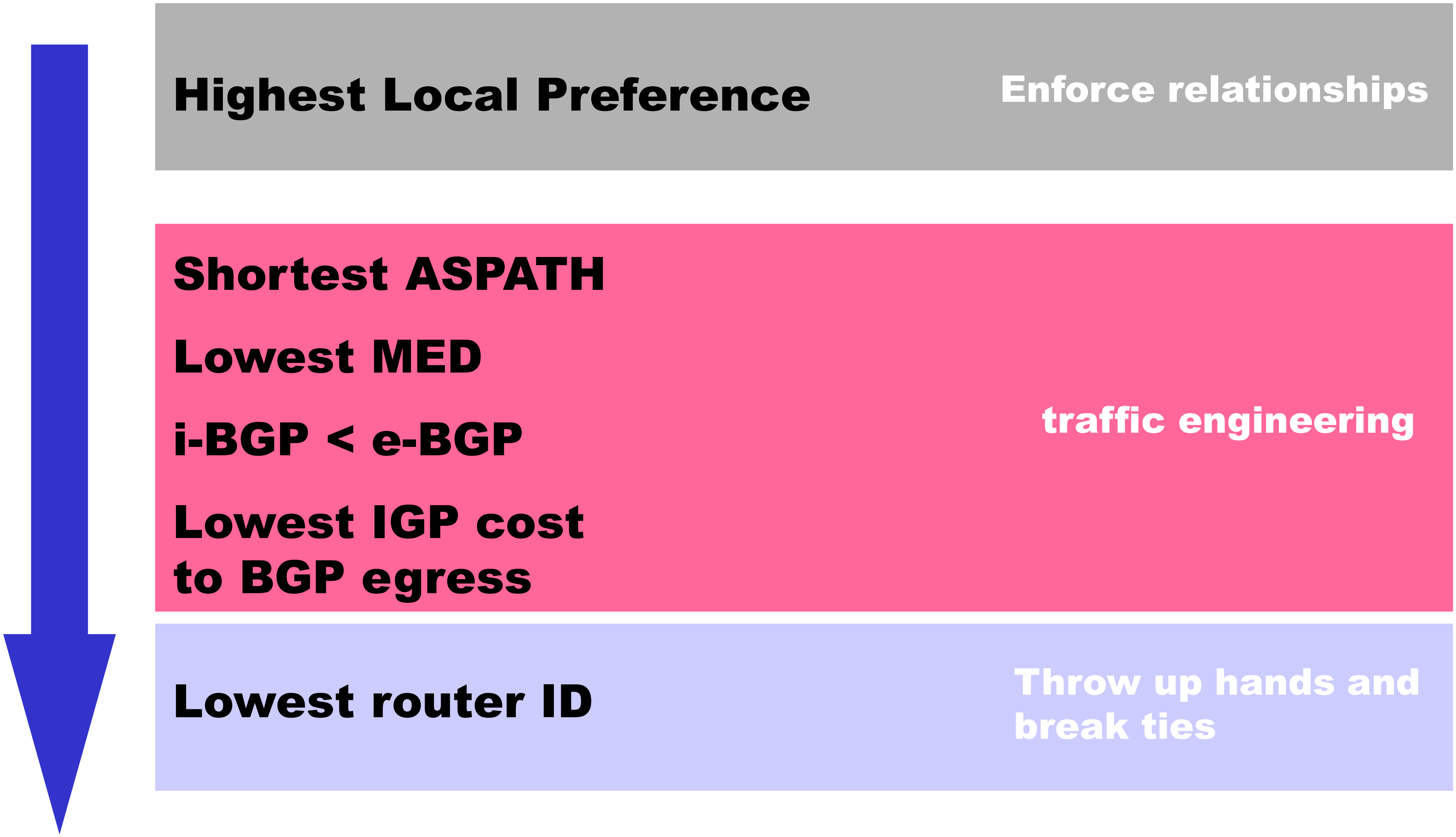


This means that MEDs must be considered BEFORE IGP distance!

Note1 : some providers will not listen to MEDs

Note2 : MEDs need not be tied to IGP distance

Route Selection Summary



Highest Local Preference

Enforce relationships

Shortest ASPATH

Lowest MED

i-BGP < e-BGP

Lowest IGP cost to BGP egress

traffic engineering

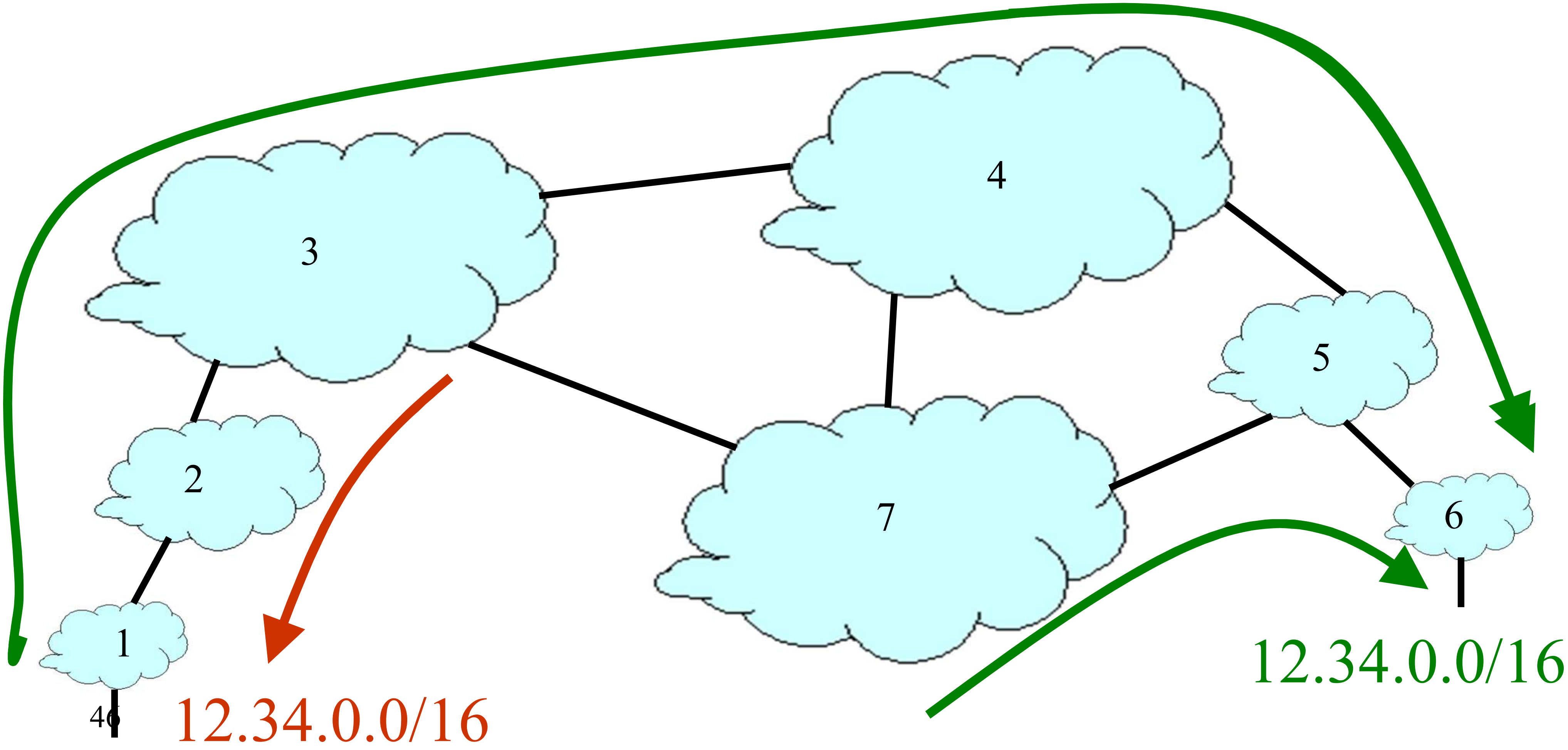
Lowest router ID

Throw up hands and break ties

BGP Attacks/Misconfiguration

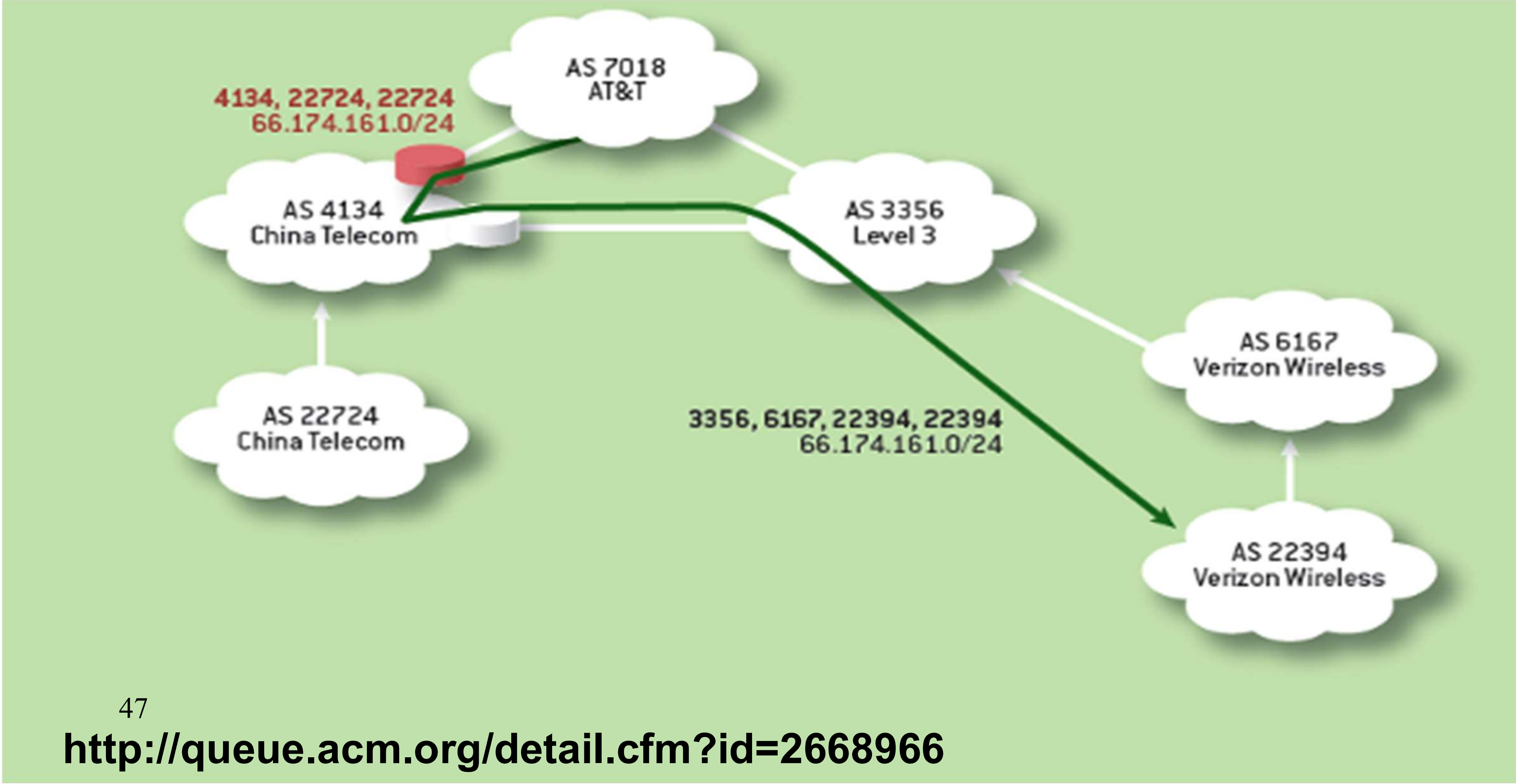
Prefix Hijacking

- Originating someone else's prefix
  - What fraction of the Internet believes it?



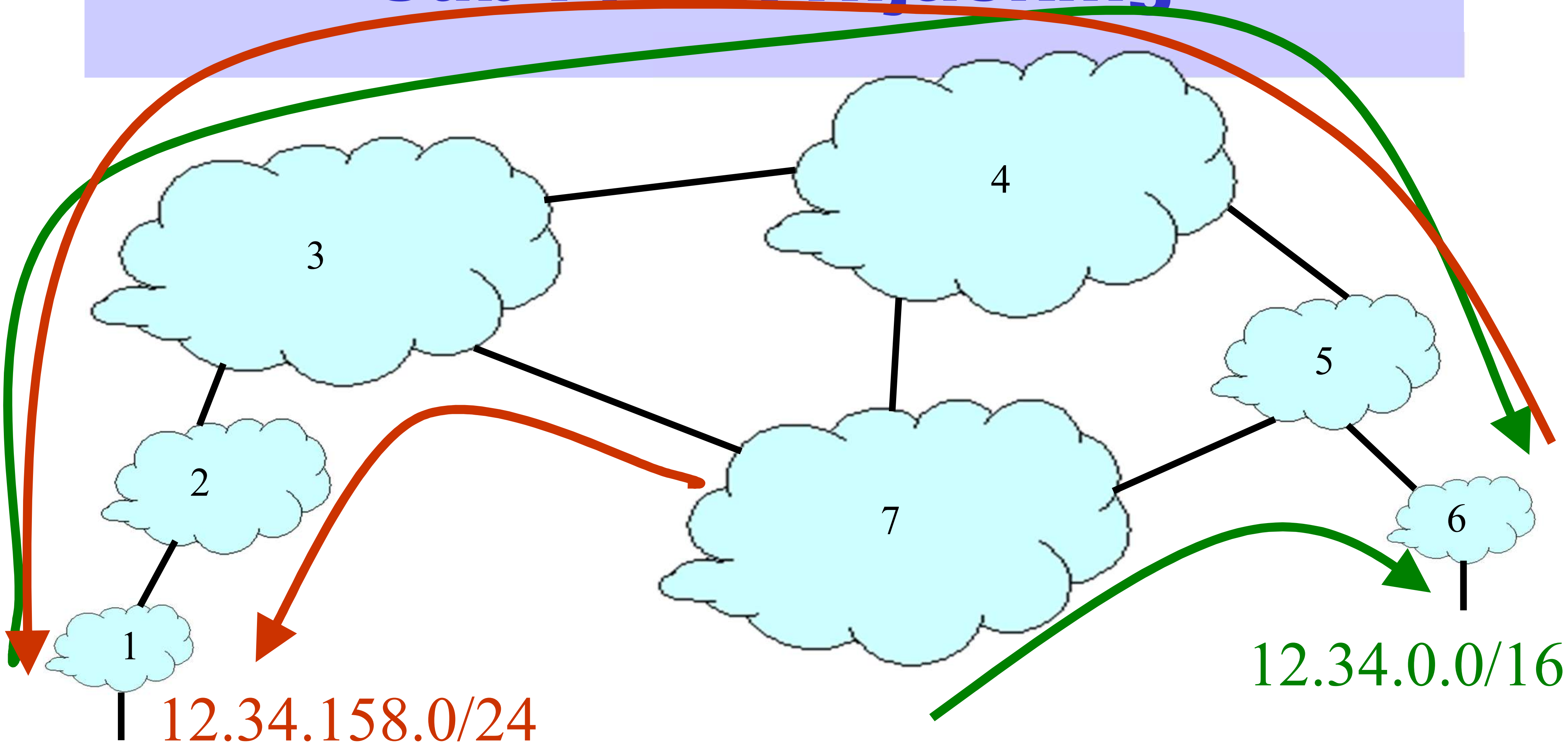
Prefix highjack

Prefix misconfiguration incident with Verizon Wireless



http://queue.acm.org/detail.cfm?id=2668966

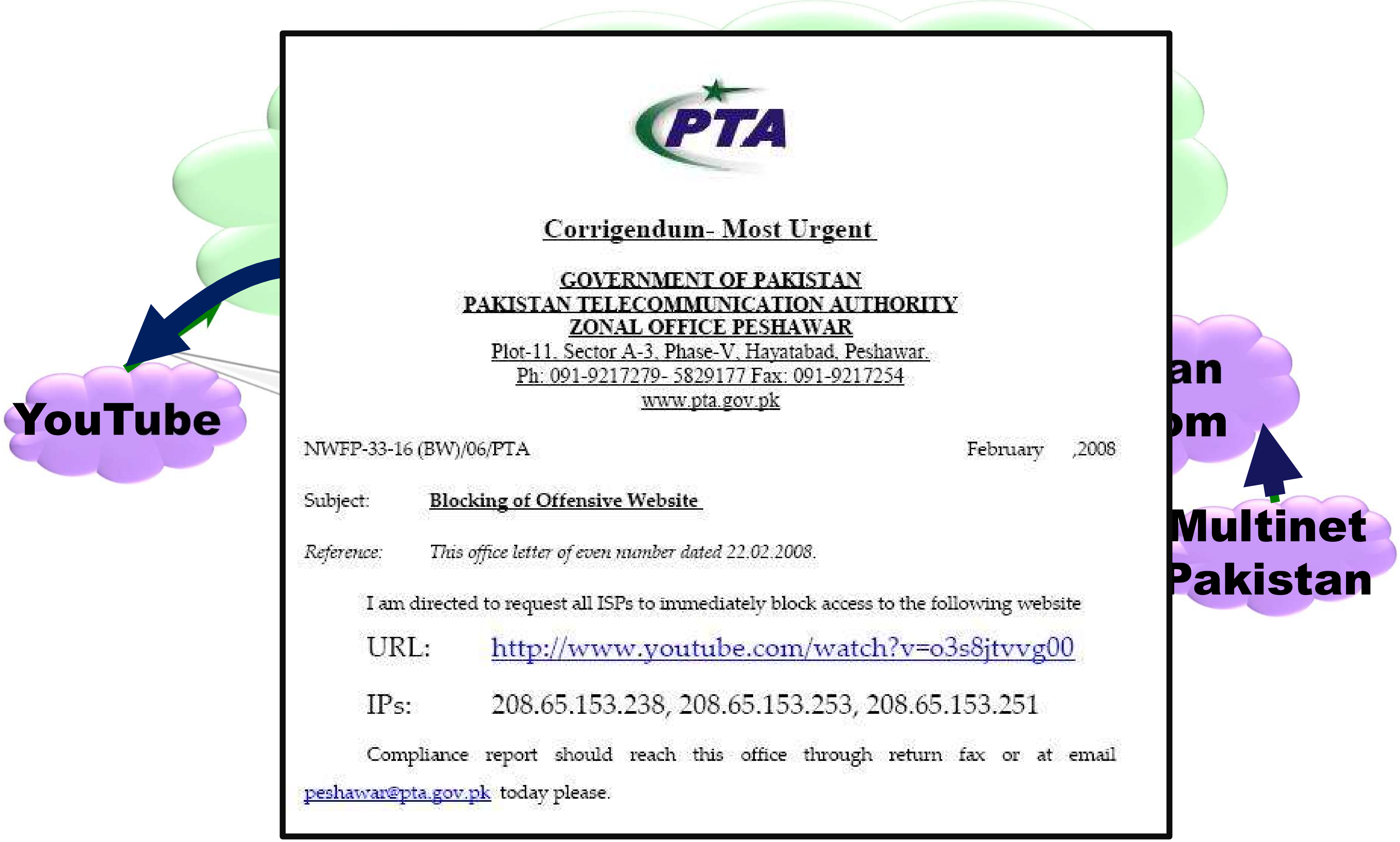
Sub-Prefix Hijacking



- **Originating a more-specific prefix**
  - **Every AS picks the bogus route for that prefix**
  - **Traffic follows the longest matching prefix**

Sub-prefix hijack

February 2008 : YouTube traffic was highjacked for a couple of hours


YouTube
an
om
Multinet
Pakistan

PTA

Corrigendum- Most Urgent

GOVERNMENT OF PAKISTAN
PAKISTAN TELECOMMUNICATION AUTHORITY
ZONAL OFFICE PESHAWAR
Plot-11, Sector A-3, Phase-V, Hayatabad, Peshawar.
Ph: 091-9217279- 5829177 Fax: 091-9217254
www.pta.gov.pk

NWFP-33-16 (BW)/06/PTA February ,2008

Subject: **Blocking of Offensive Website**

Reference: This office letter of even number dated 22.02.2008.

I am directed to request all ISPs to immediately block access to the following website

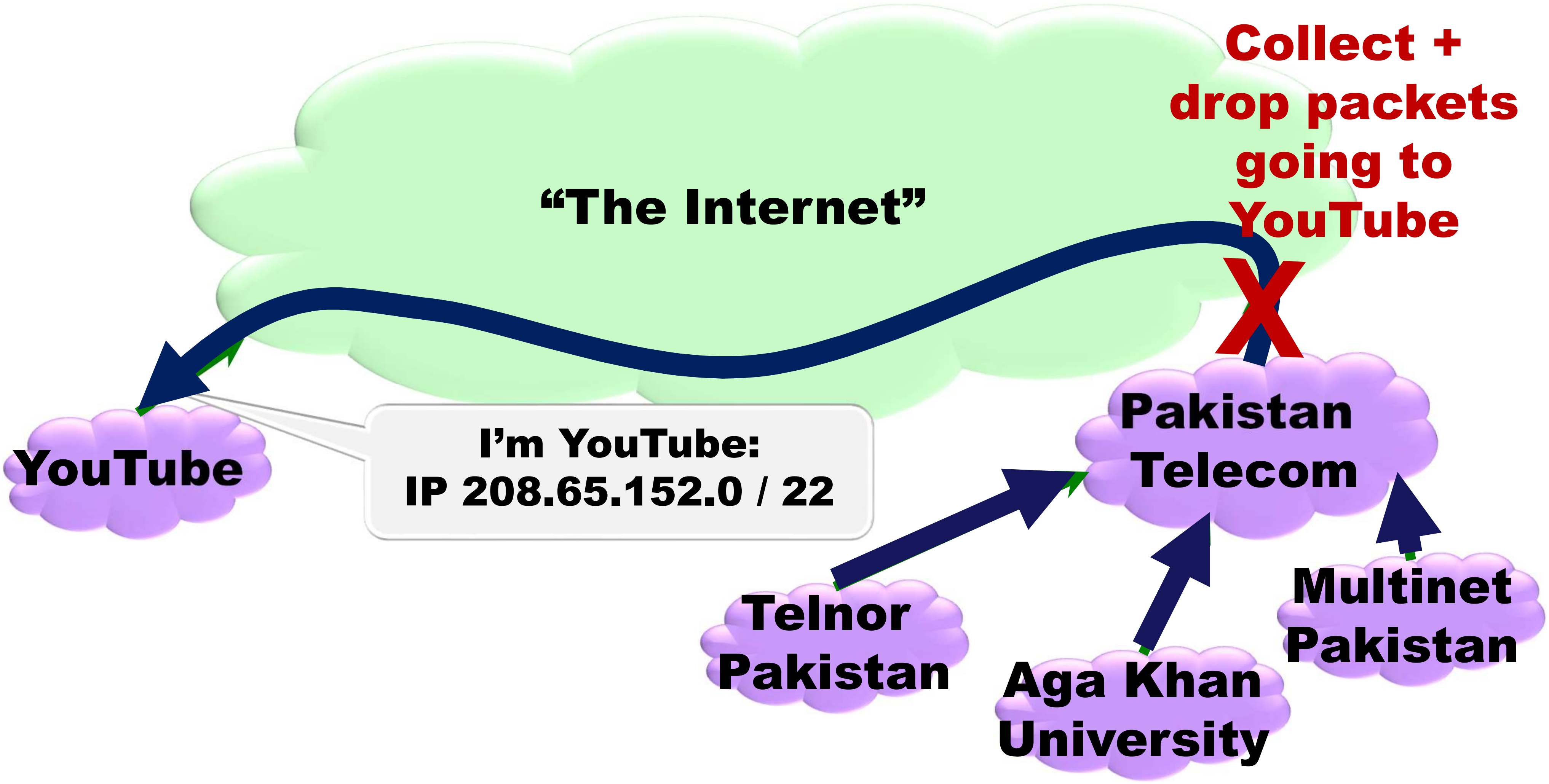
URL: http://www.youtube.com/watch?v=o3s8jtvvg00

IPs: 208.65.153.238, 208.65.153.253, 208.65.153.251

Compliance report should reach this office through return fax or at email peshawar@pta.gov.pk today please.

Pakistan Telecom: Sub-prefix hijack

Here’s what should have happened….

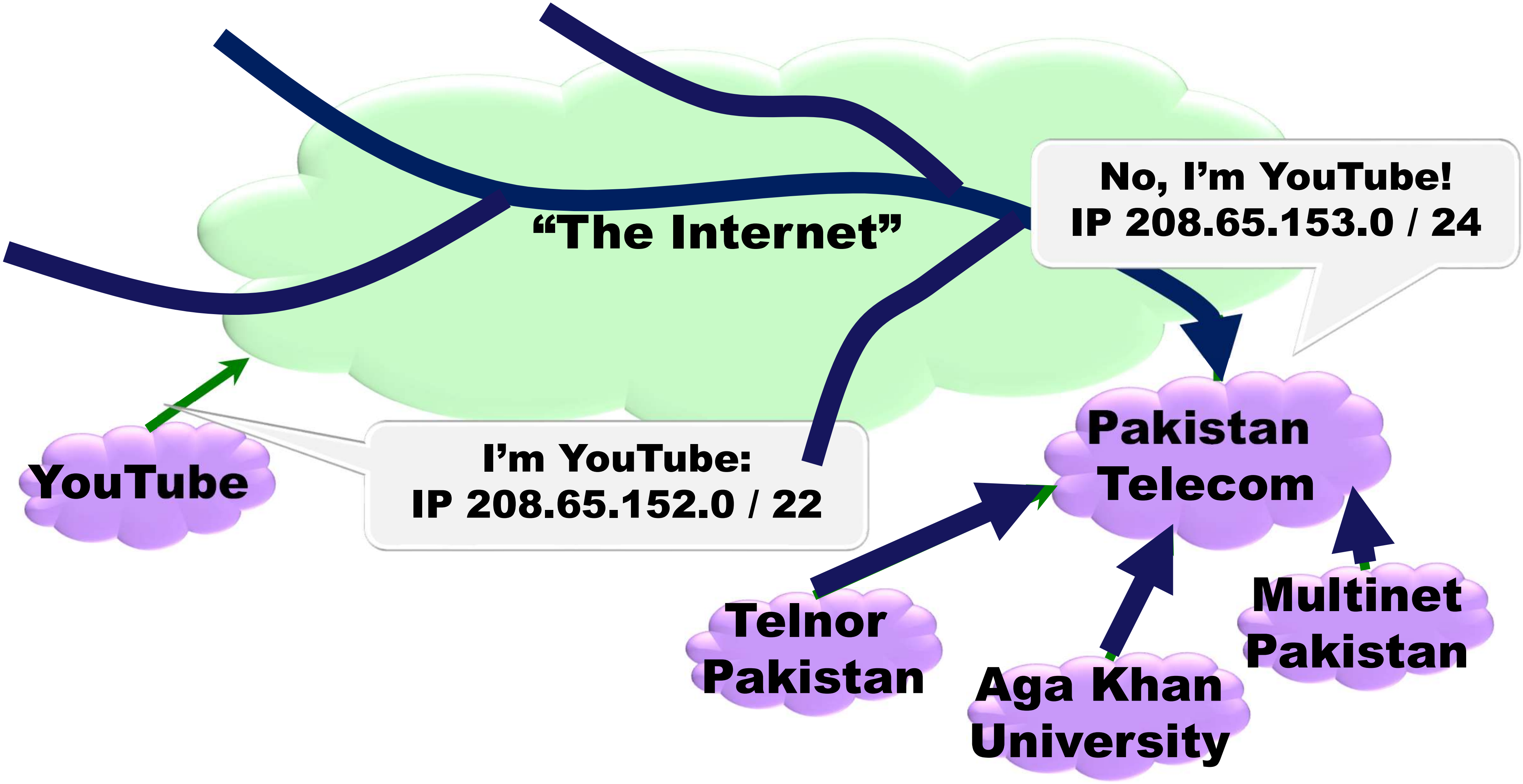


Block your own customers.

Target IPs:[208.65.153.238, 208.65.153.253, 208.65.153.251]

Pakistan Telecom: Sub-prefix hijack

But here's what Pakistan ended up doing…

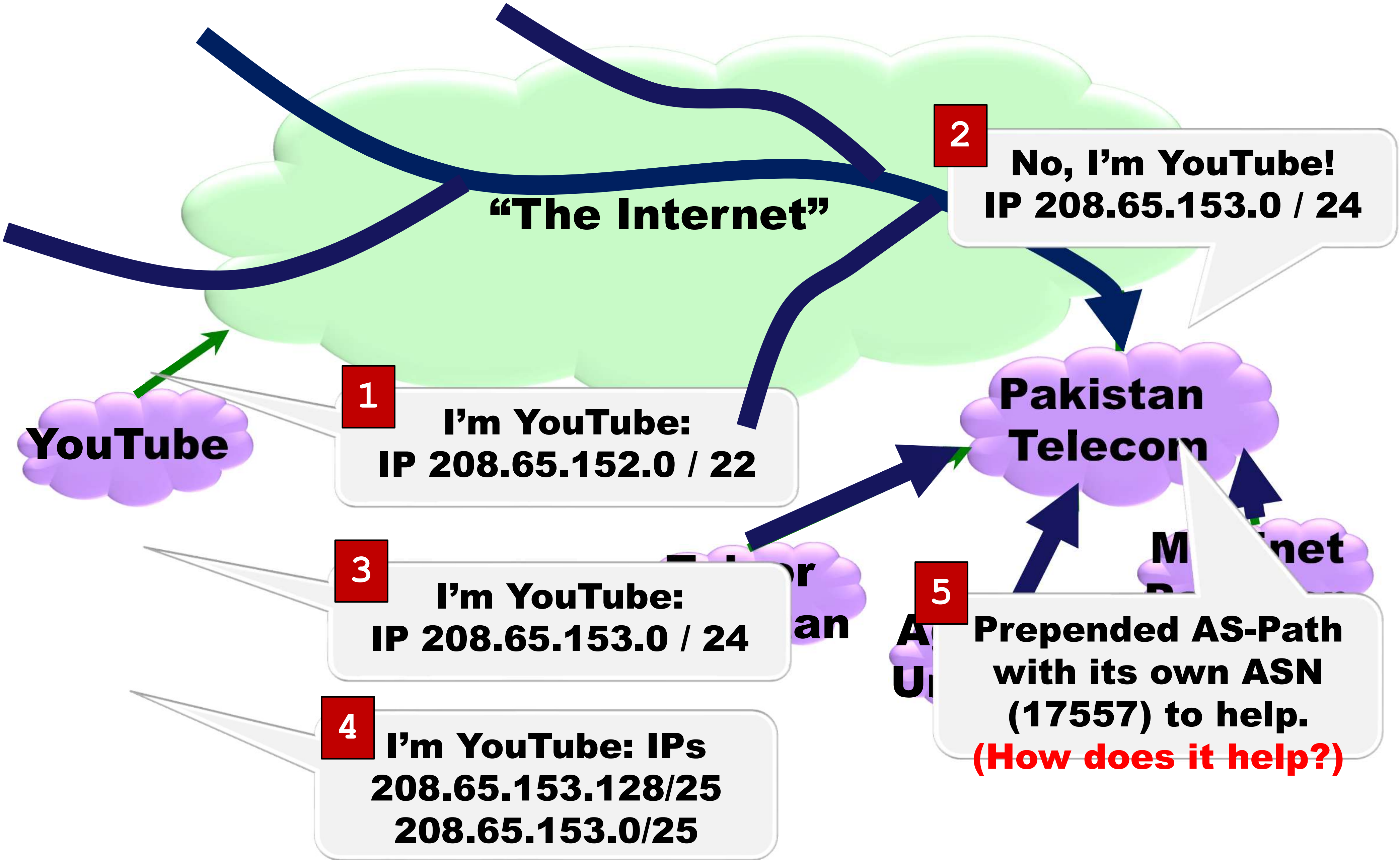


No, I'm YouTube!
IP 208.65.153.0 / 24

I'm YouTube:
IP 208.65.152.0 / 22

Pakistan Telecom: Sub-prefix hijack

But here's what Pakistan ended up doing…



1 I'm YouTube:
IP 208.65.152.0 / 22

2 No, I'm YouTube!
IP 208.65.153.0 / 24

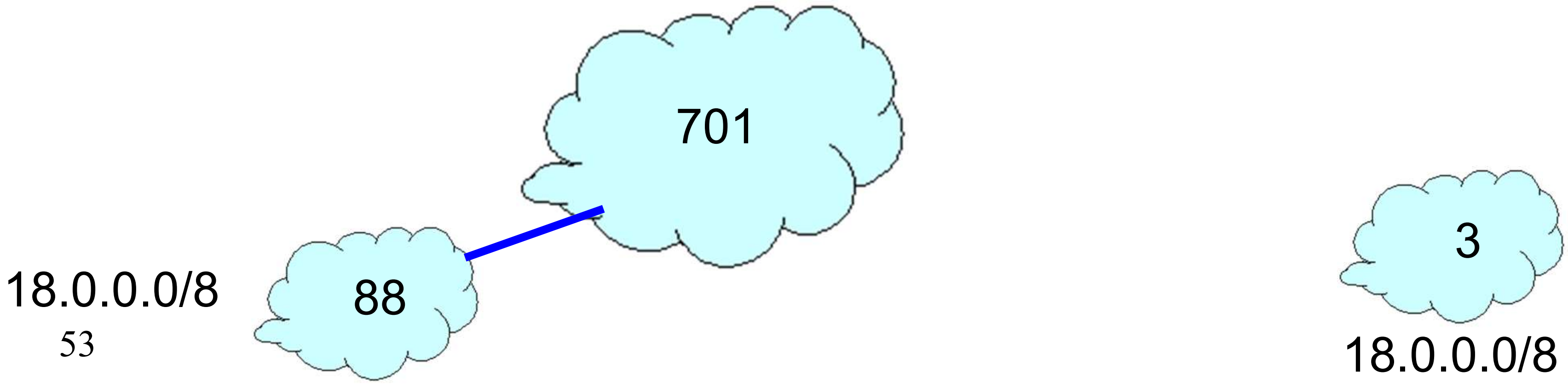
3 I'm YouTube:
IP 208.65.153.0 / 24

4 I'm YouTube: IPs
208.65.153.128/25
208.65.153.0/25

5 Prepended AS-Path with its own ASN (17557) to help.
(How does it help?)

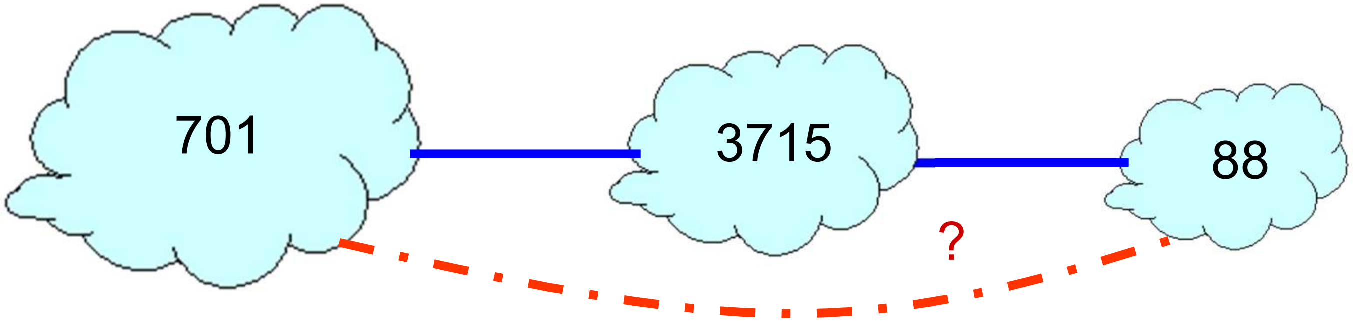
Bogus AS Paths to Hide Hijacking

- **Adds AS hop(s) at the end of the path**
  - **E.g., turns “701 88” into “701 88 3”**
- **Motivations**
  - **Evade detection for a bogus route**
  - **E.g., by adding the legitimate AS to the end**
- **Hard to tell that the AS path is bogus…**
  - **Even if other ASes filter based on prefix ownership**

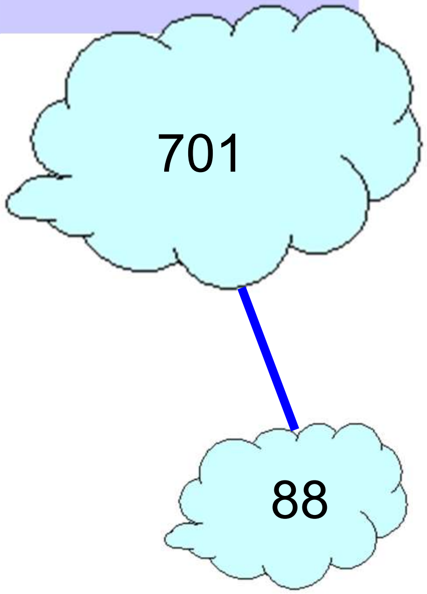


Path-Shortening Attacks

- Remove ASes from the AS path
  - E.g., turn “701 3715 88” into “701 88”
- Motivations
  - Make the AS path look shorter than it is
  - Attract sources that normally try to avoid AS 3715
  - Help AS 88 look like it is closer to the Internet’s core
- Who can tell that this AS path is a lie?
  - Maybe AS 88 *does* connect to AS 701 directly



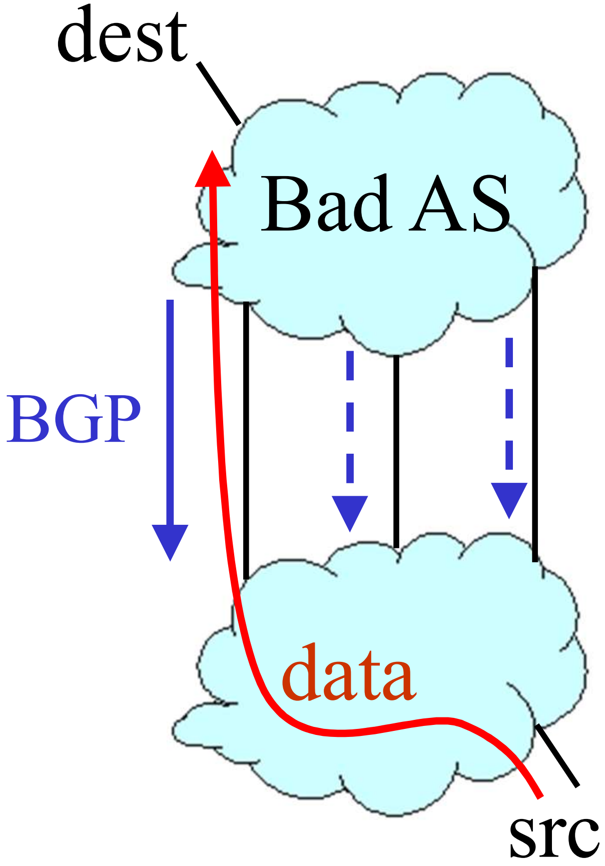
Attacks that Add a Bogus AS Hop



- **Add ASes to the path**
  - **E.g., turn "701 88" into "701 3715 88"**
- **Motivations**
  - **Trigger loop detection in AS 3715**
    - **Denial-of-service attack on AS 3715**
    - **Or, blocking unwanted traffic coming from AS 3715!**
  - **Make your AS look like is has richer connectivity**
- **Who can tell the AS path is a lie?**
  - **AS 3715 could, if it could see the route**
  - **AS 88 could, but would it really care as long as it received data traffic meant for it?**

Violating "Consistent Export" to Peers

- Peers require consistent export
  - Prefix advertised at all peering points
  - Prefix advertised with same AS path length
- Reasons for violating the policy
  - Trick neighbor into "cold potato"
  - Configuration mistake
- Main defense
  - Analyzing BGP updates
  - … or data traffic
  - … for signs of inconsistency



Other Attacks

- Attacks on BGP sessions
  - Confidentiality of BGP messages
  - Denial-of-service on BGP session
  - Inserting, deleting, modifying, or replaying messages
- Resource exhaustion attacks
  - Too many IP prefixes (e.g., BGP "512K Day")
  - Too many BGP update messages
- Data-plane attacks
  - Announce one BGP routes, but use another

Solution Techniques

- **Protective filtering**
  - **Know your neighbors**
- **Anomaly detection**
  - **Suspect the unexpected**
- **Checking against registries**
  - **Establish ground truth for prefix origination**
- **Signing and verifying**
  - **Prevent bogus AS PATHs**
- **Data-plane verification**
  - **Ensure the path is actually followed**