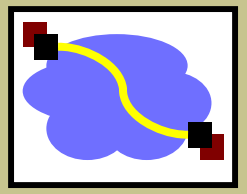
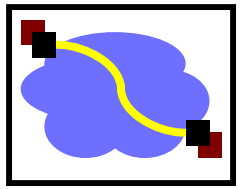


Inter-Domain Routing

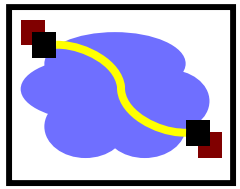
BGP (Border Gateway Protocol)

Summary



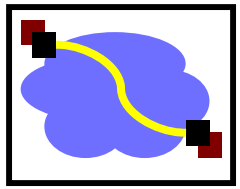
- The Story So Far…
  - Routing protocols generate the forwarding table
  - Two styles: distance vector, link state
  - Scalability issues:
    - Distance vector protocols suffer from count-to-infinity
    - Link state protocols must flood information throughout network
- Questions
  - How to make routing protocols support large networks
  - How to make routing protocols support business policies

Outline



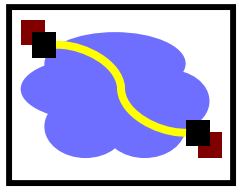
- Routing hierarchy
- Internet structure
- External BGP (E-BGP)
- Internal BGP (I-BGP)

Routing Hierarchies



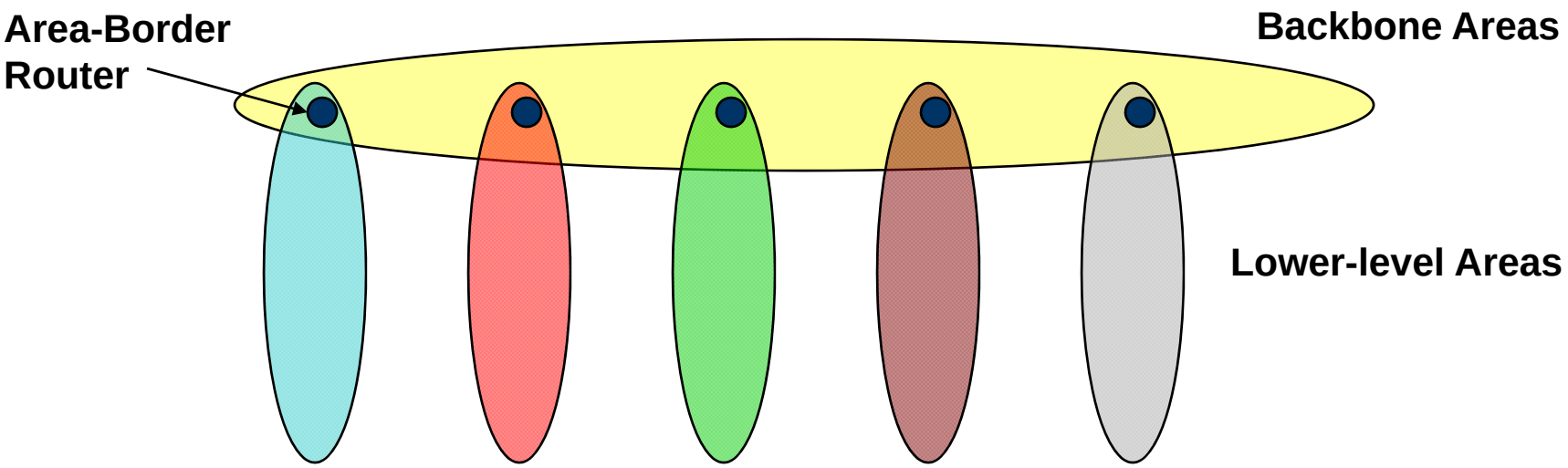
- Flat routing does not scale
  - Storage → Each node cannot be expected to store routes to every destination (or destination network)
  - Convergence times increase
  - Communication → Total message count increases
- Key observation
  - Need less information with increasing distance to destination
  - Need lower diameter networks
- Solution: area hierarchy

Areas



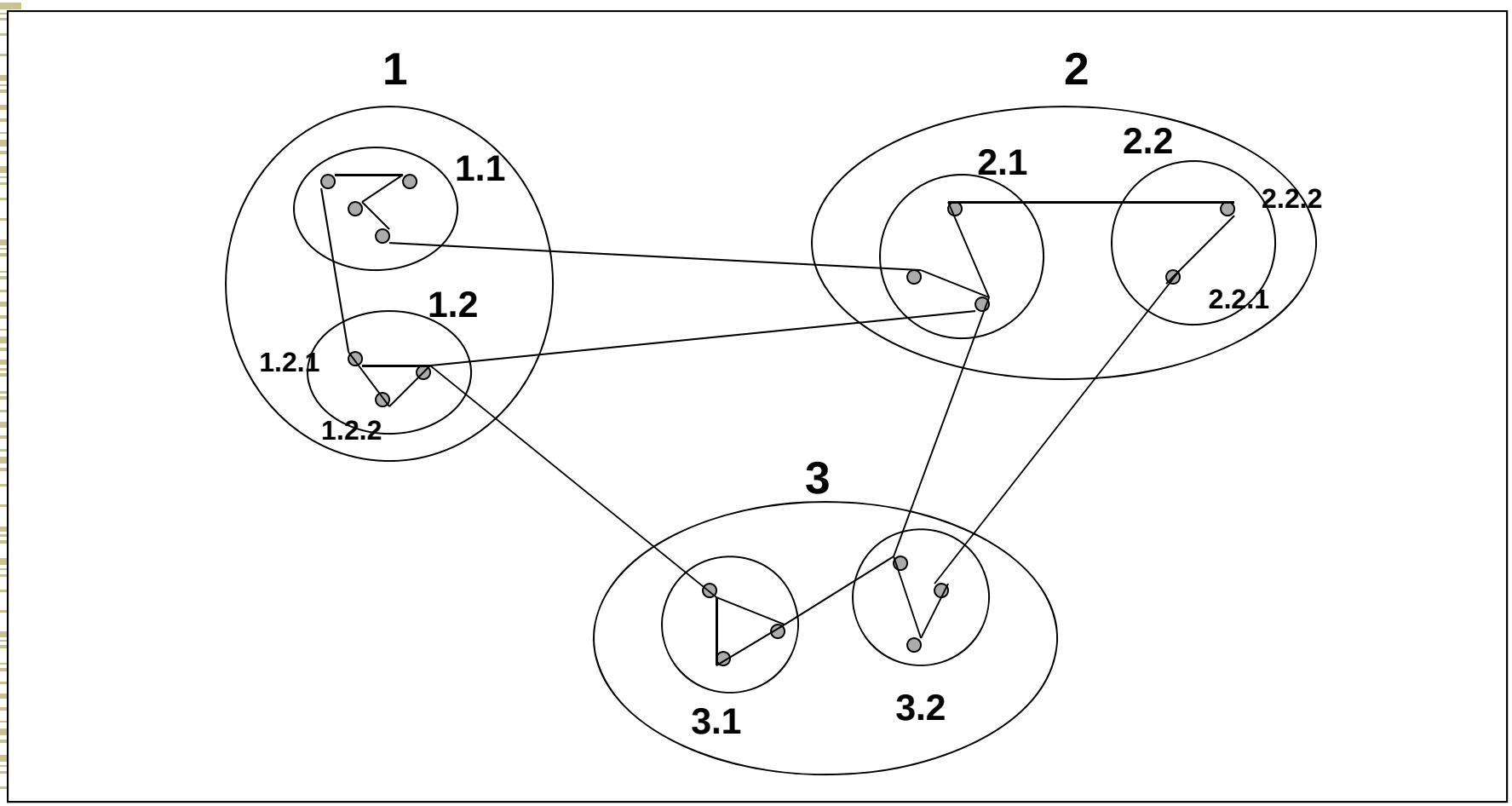
- Divide network into areas
  - Areas can have nested sub-areas
- Hierarchically address nodes in a network
  - Sequentially number top-level areas
  - Sub-areas of area are labeled relative to that area
  - Nodes are numbered relative to the smallest containing area

Routing Hierarchy



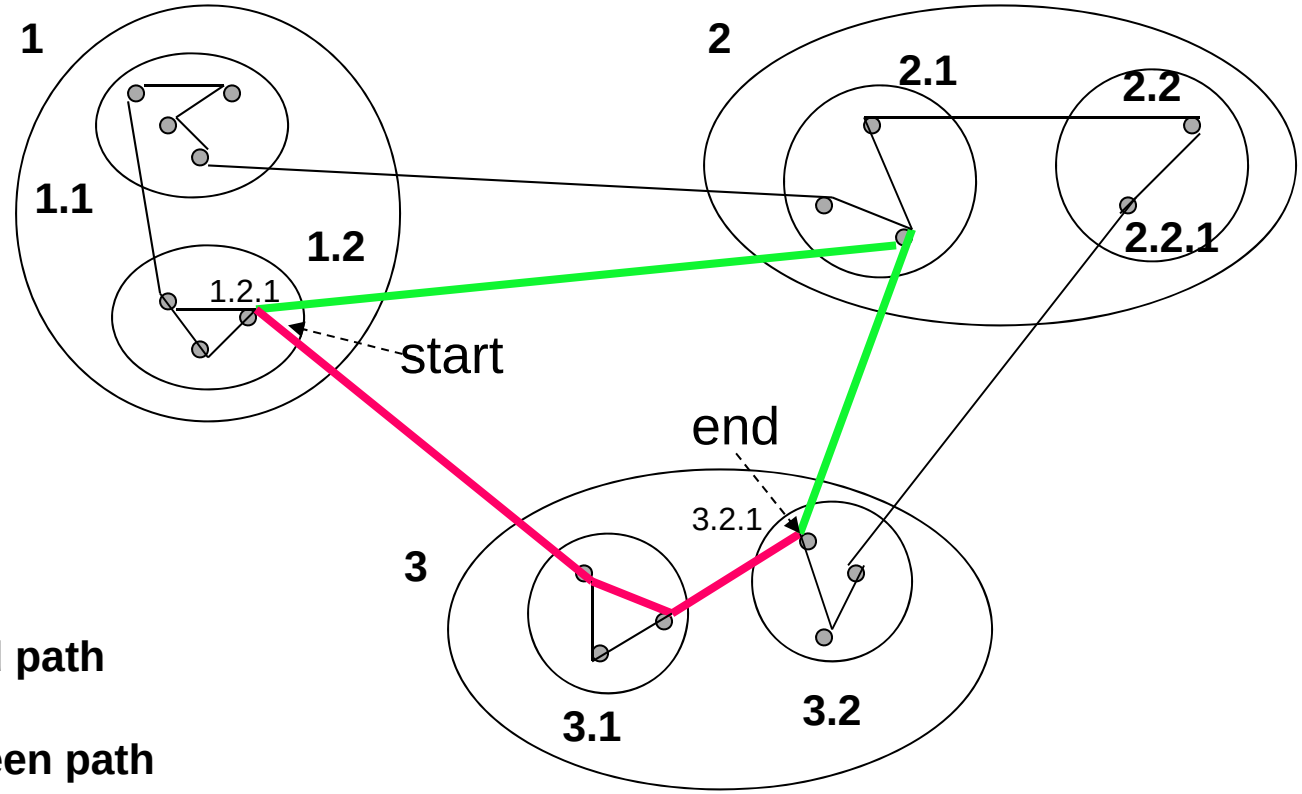
- Partition Network into “Areas”
  - Within area
    - Each node has routes to every other node
  - Outside area
    - Each node has routes for other top-level areas only
    - Inter-area packets are routed to nearest appropriate border router
- Constraint: no path between two sub-areas of an area can exit that area

Area Hierarchy Addressing

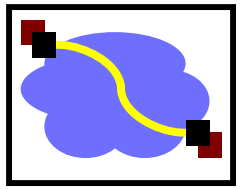


Path Sub-optimality

- Can result in sub-optimal paths

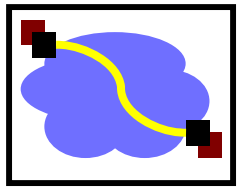


Outline

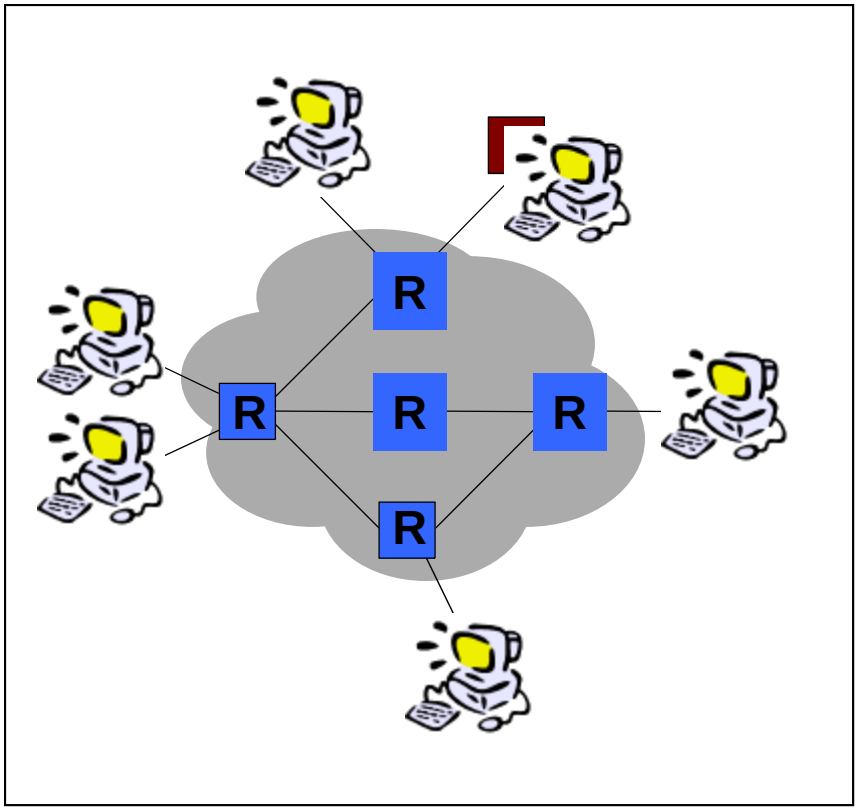


- Routing hierarchy
- Internet structure
- External BGP (E-BGP)
- Internal BGP (I-BGP)

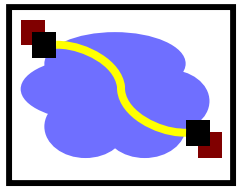
A Logical View of the Internet?



- After looking at RIP/OSPF descriptions
  - End-hosts connected to routers
  - Routers exchange messages to determine connectivity
- NOT TRUE!

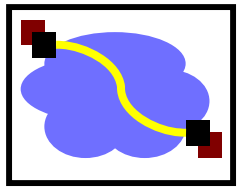


Internet's Area Hierarchy



- What is an Autonomous System (AS)?
  - A set of routers under a single technical administration, using an *interior gateway protocol (IGP)* and common metrics to route packets within the AS and using an *exterior gateway protocol (EGP)* to route packets to other AS's
  - Sometimes AS's use multiple IGPs and metrics, but appear as single AS's to other AS's
- Each AS assigned unique ID
- AS's peer at network exchanges
- http://www.cidr-report.org/as2.0/

Hierarchical routing

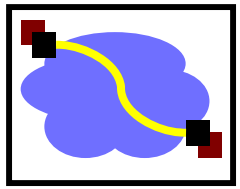


- aggregate routers into regions, "autonomous systems" (AS)
- routers in same AS run same routing protocol
  - "intra-AS" routing protocol
  - routers in different AS can run different intra-AS routing protocol

gateway router:

- at "edge" of its own AS
- has link to router in another AS

Autonomous Systems

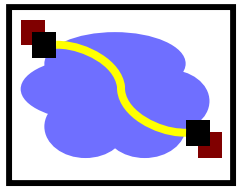


- An **autonomous system (AS)** is a region of the Internet that is administered by a single entity and that has a unified routing policy

- Each autonomous system is assigned an Autonomous System Number (**ASN**).
  - Rogers Cable Inc. (AS812)
  - Sprint (AS1239, AS1240, AS 6211, …)

- Interdomain routing is concerned with determining paths between autonomous systems **(interdomain routing**)
- Routing protocols for interdomain routing are called **exterior gateway protocols** (EGP)

AS Numbers (ASNs)



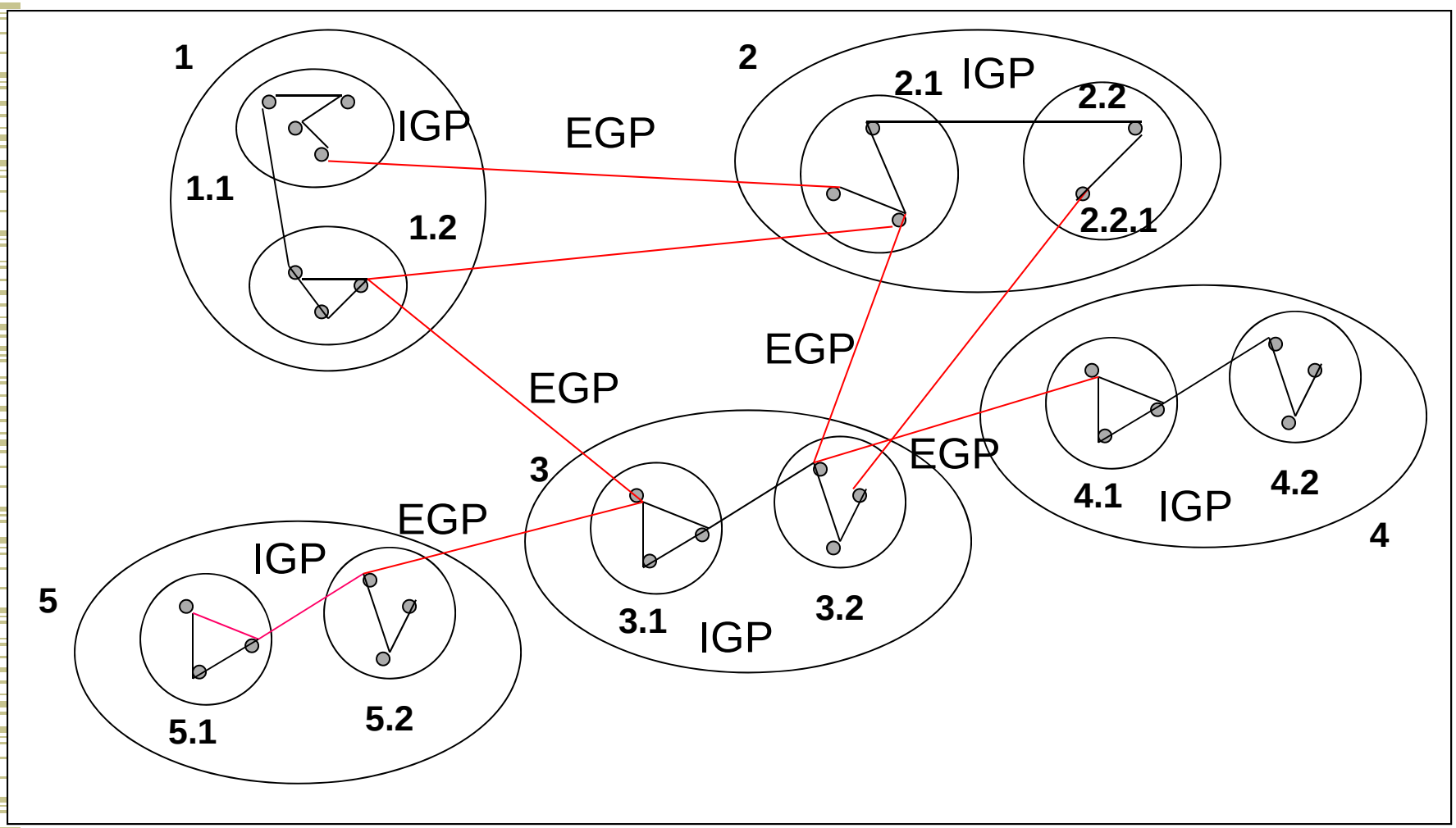
ASNs are 16 bit values 64512 through 65535 are “private”

Currently over 50,000 in use

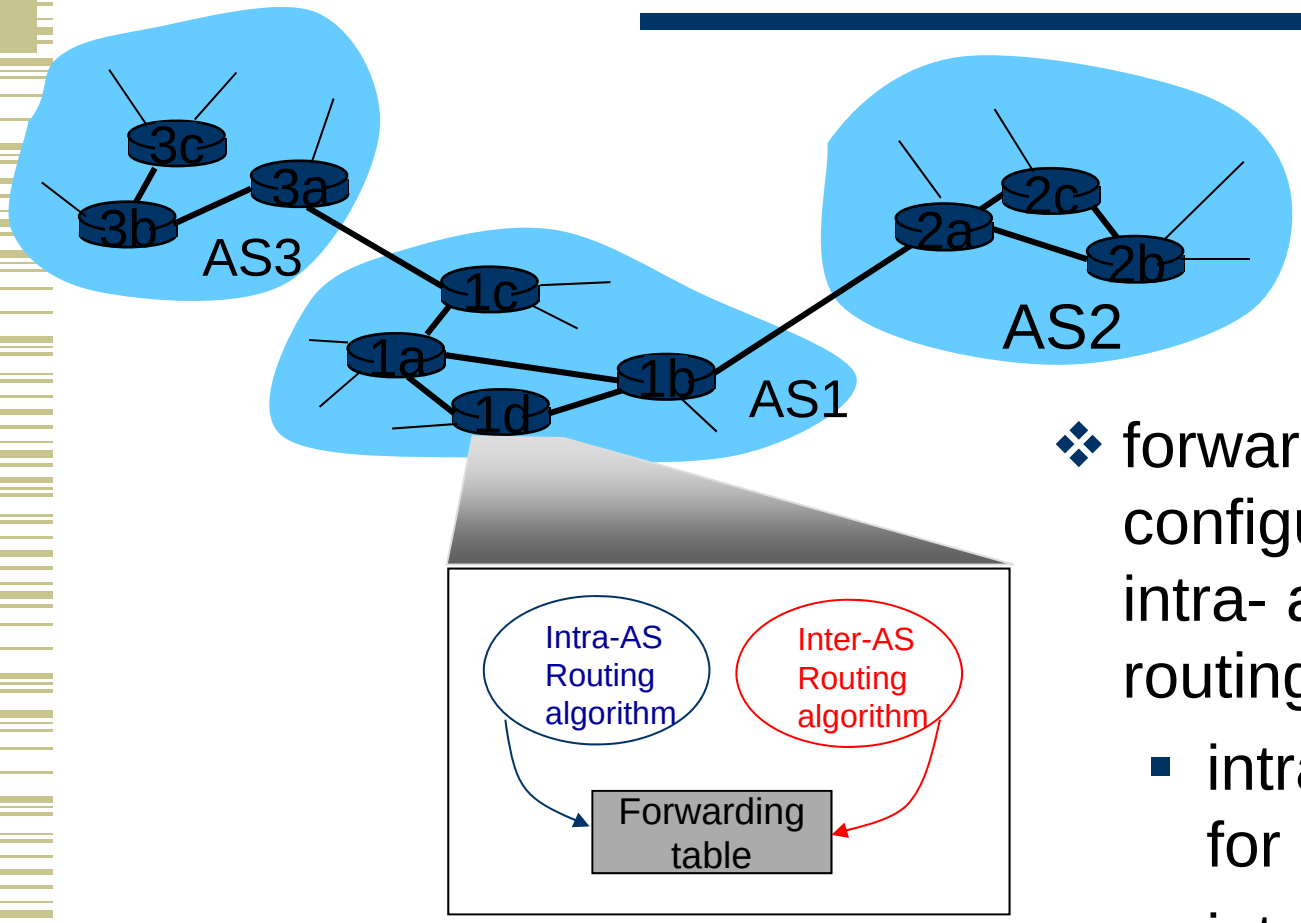
- Genuity: 1
- MIT: 3
- JANET: 786
- UC San Diego: 7377
- AT&T: 7018, 6341, 5074, …
- UUNET: 701, 702, 284, 12199, …
- Sprint: 1239, 1240, 6211, 6242, …
- …

ASNs represent units of routing policy

Example

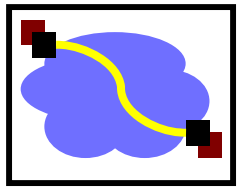

1
IGP
1.1
1.2
EGP
2
2.1
IGP
2.2
2.2.1
EGP
EGP
EGP
3
3.1
IGP
3.2
EGP
4
4.1
IGP
4.2
EGP
IGP
5
5.1
5.2

Interconnected ASes



- forwarding table configured by both intra- and inter-AS routing algorithm
  - intra-AS sets entries for internal dests
  - inter-AS & intra-AS sets entries for external dests

Inter-AS tasks

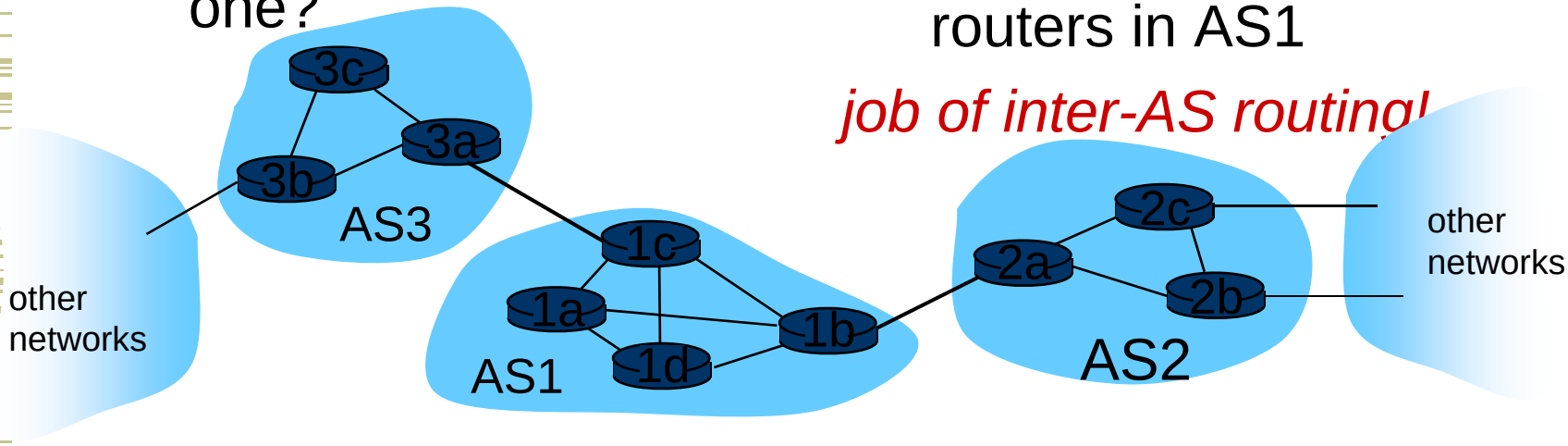


- suppose router in AS1 receives datagram destined outside of AS1:
  - router should forward packet to gateway router, but which one?

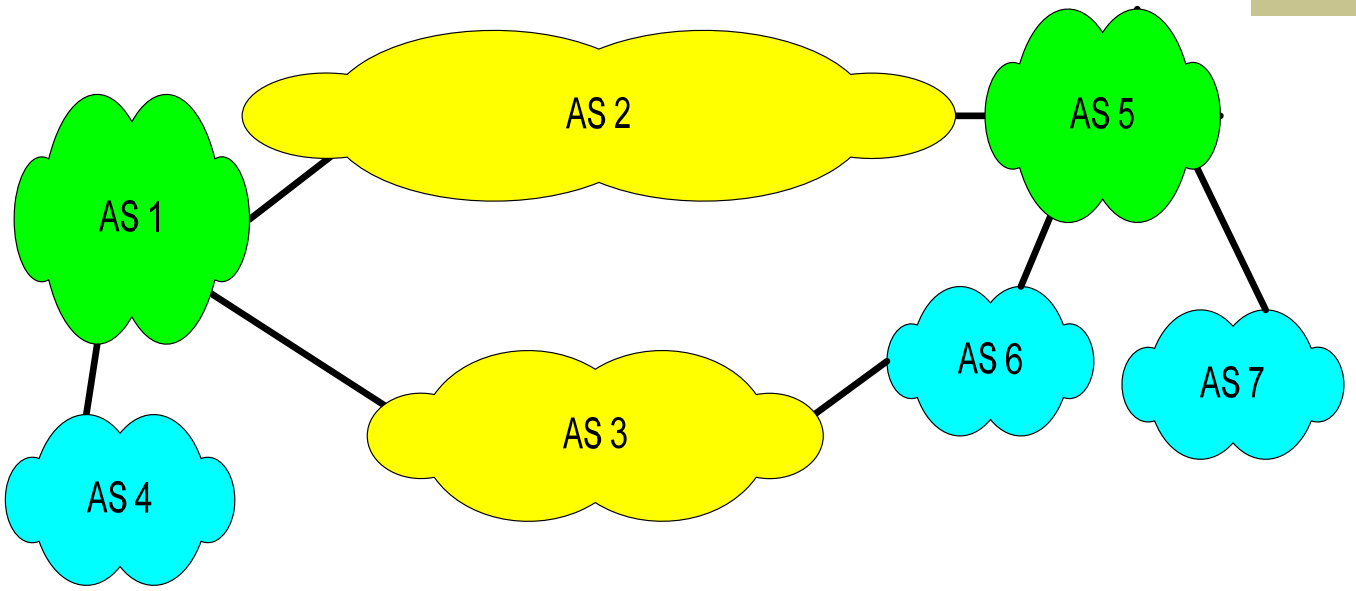
AS1 must:

1. learn which dests are reachable through AS2, which through AS3
2. propagate this reachability info to all routers in AS1

job of inter-AS routing!

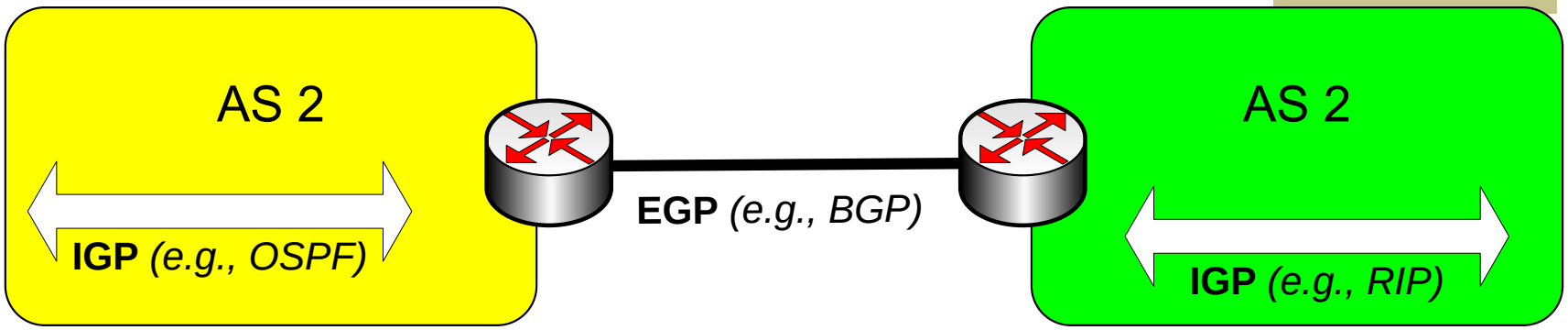


Interdomain and Intradomain Routing



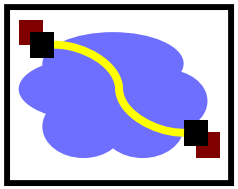
- Routing protocols for intradomain routing are called interior gateway protocols (IGP)
  - Objective: shortest path
- Routing protocols for interdomain routing are called exterior gateway protocols (EGP)
  - Objective: satisfy policy of the AS

Interdomain vs Intradomain

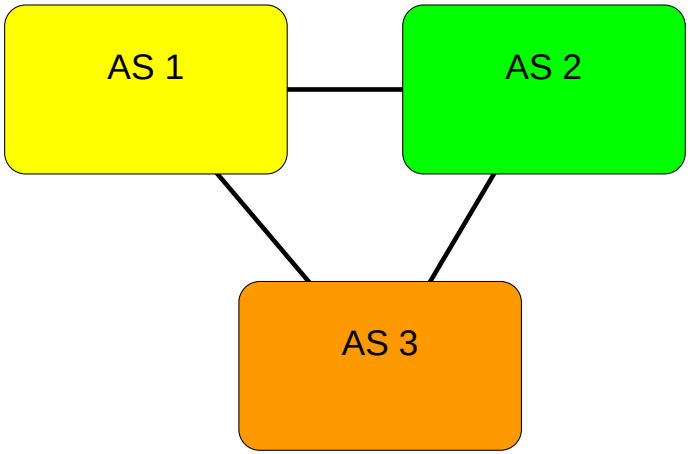


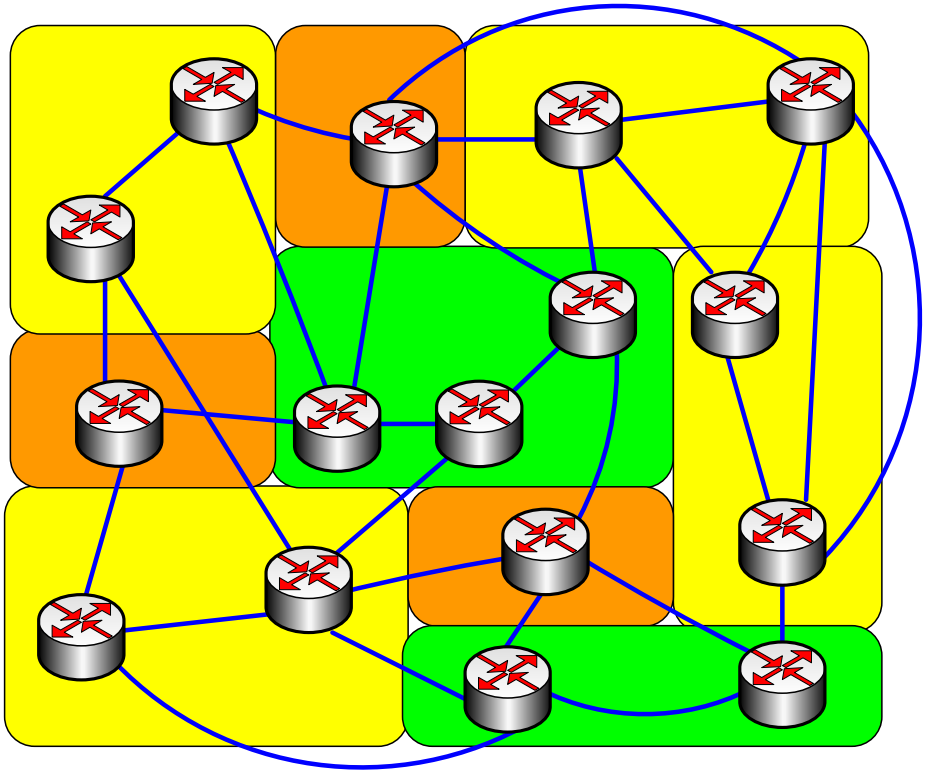
- Intradomain routing
  - Routing is done based on metrics
  - Routing domain is one autonomous system
- Interdomain routing
  - Routing is done based on policies
  - Routing domain is the entire Internet

Interdomain Routing

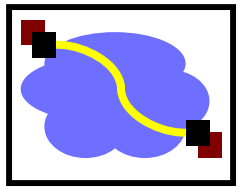


- Interdomain routing is based on connectivity between autonomous systems
- Interdomain routing can ignore many details of router interconnection

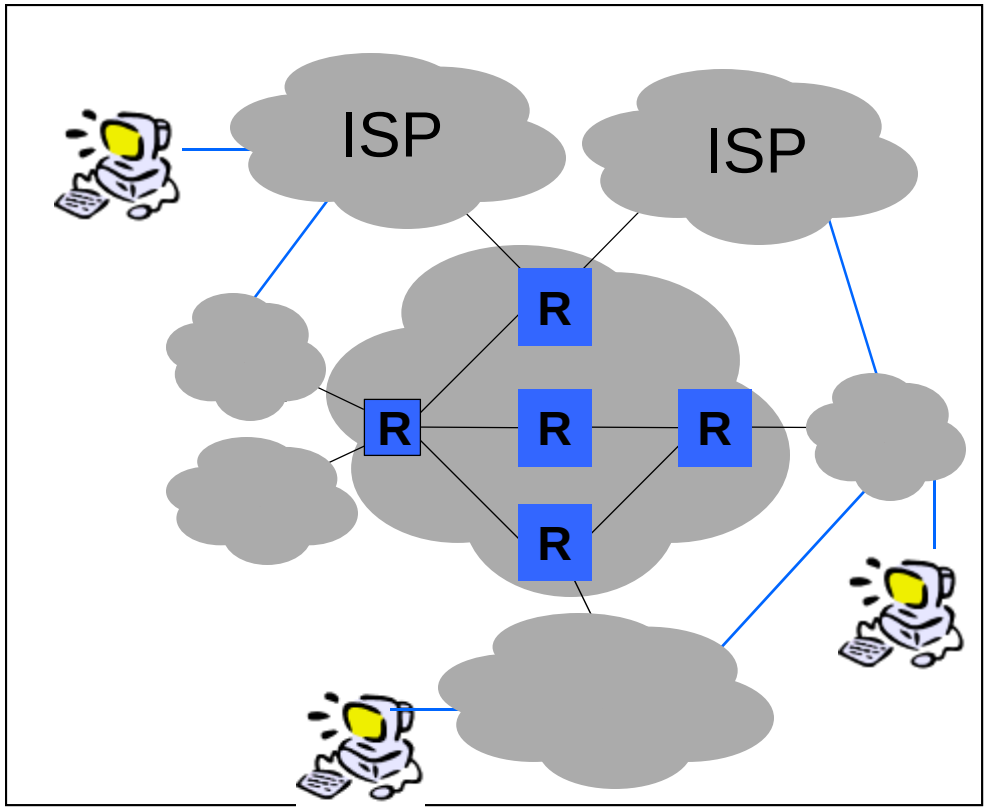




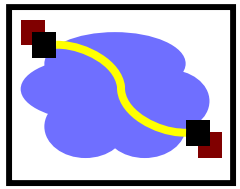
A Logical View of the Internet?



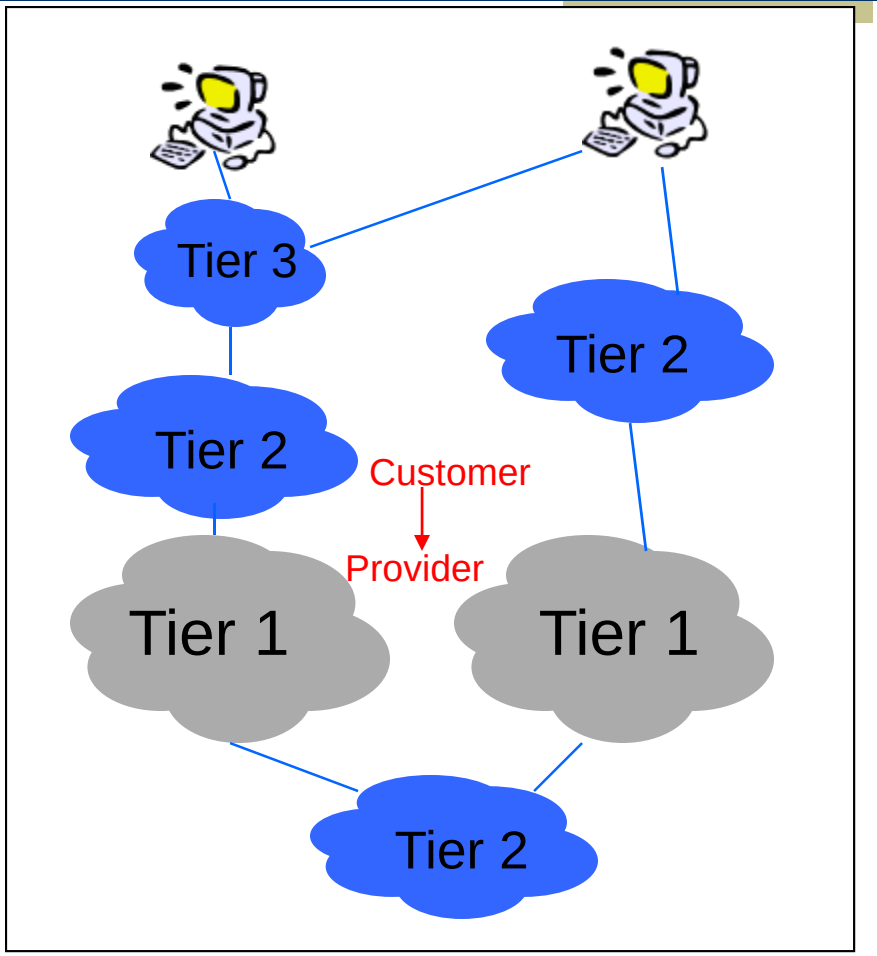
- RIP/OSPF not very scalable → area hierarchies
- NOT TRUE EITHER!
- ISP's aren't equal
  - Size
  - Connectivity



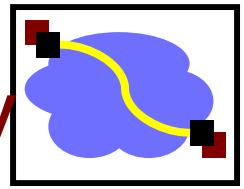
A Logical View of the Internet



- Tier 1 ISP
  - “Default-free” with global reachability info
- Tier 2 ISP
  - Regional or country-wide
- Tier 3 ISP
  - Local

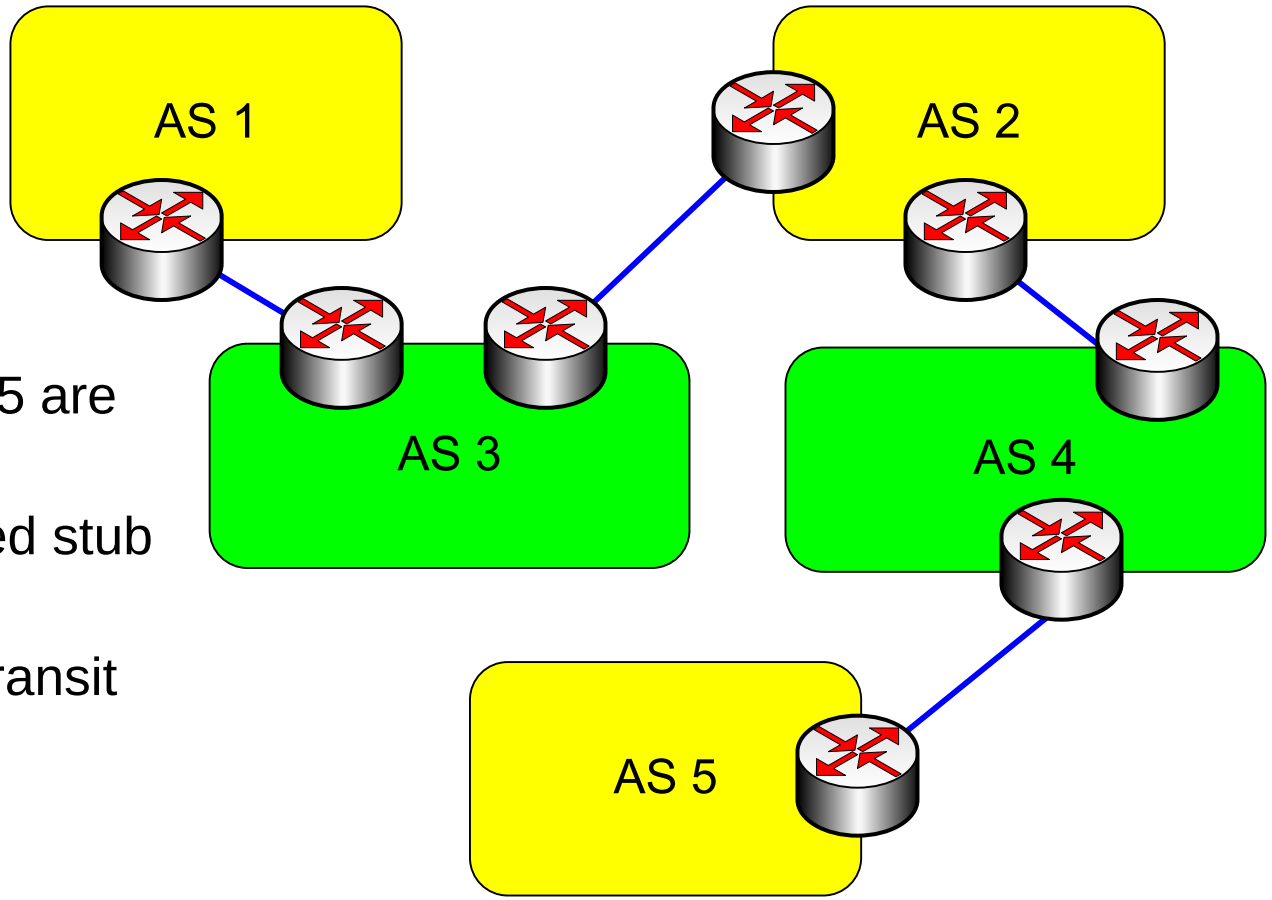


Autonomous Systems Terminology



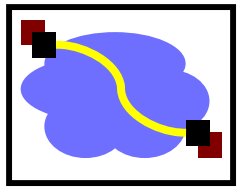
- **Local traffic** = traffic with source or destination in AS
- **Transit traffic** = traffic that passes through the AS
- **Stub AS** = has connection to only one AS, only carry local traffic
- **Multihomed AS** = has connection to >1 AS, but does not carry transit traffic
- **Transit AS** = has connection to >1 AS and carries transit traffic

Stub and Transit Networks



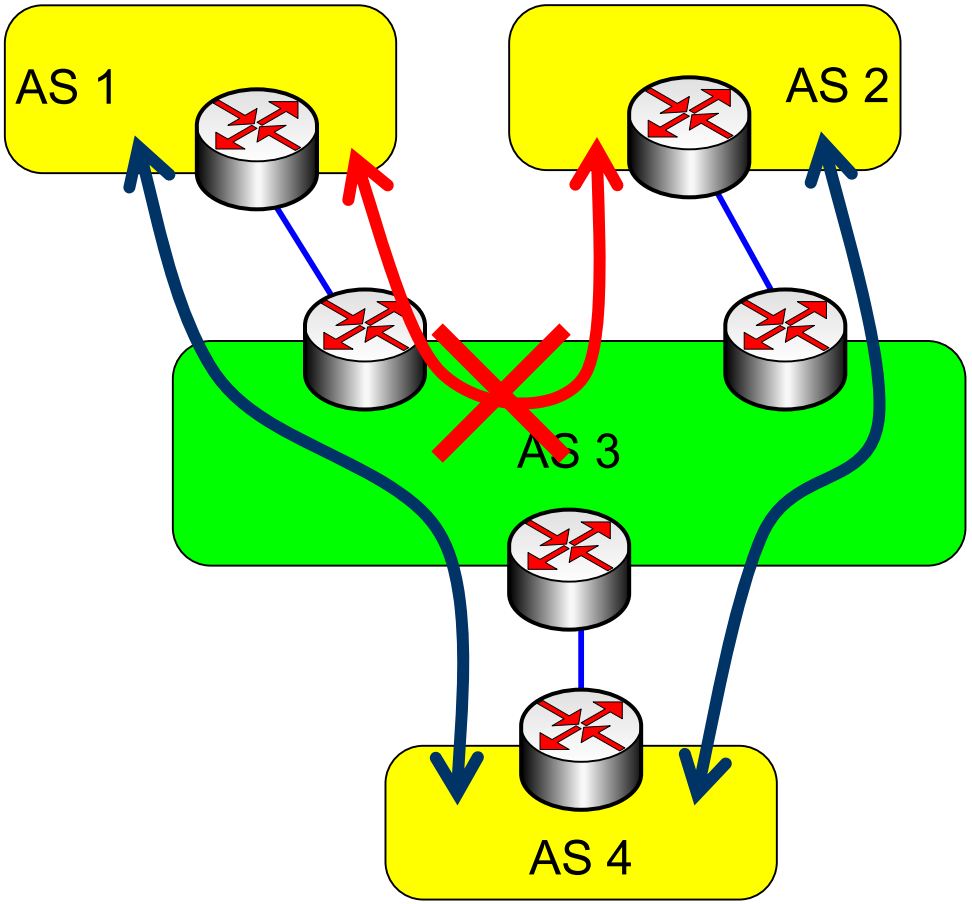
- AS 1, AS 2, and AS 5 are stub networks
- AS 2 is a multi-homed stub network
- AS 3 and AS 4 are transit networks

Selective Transit



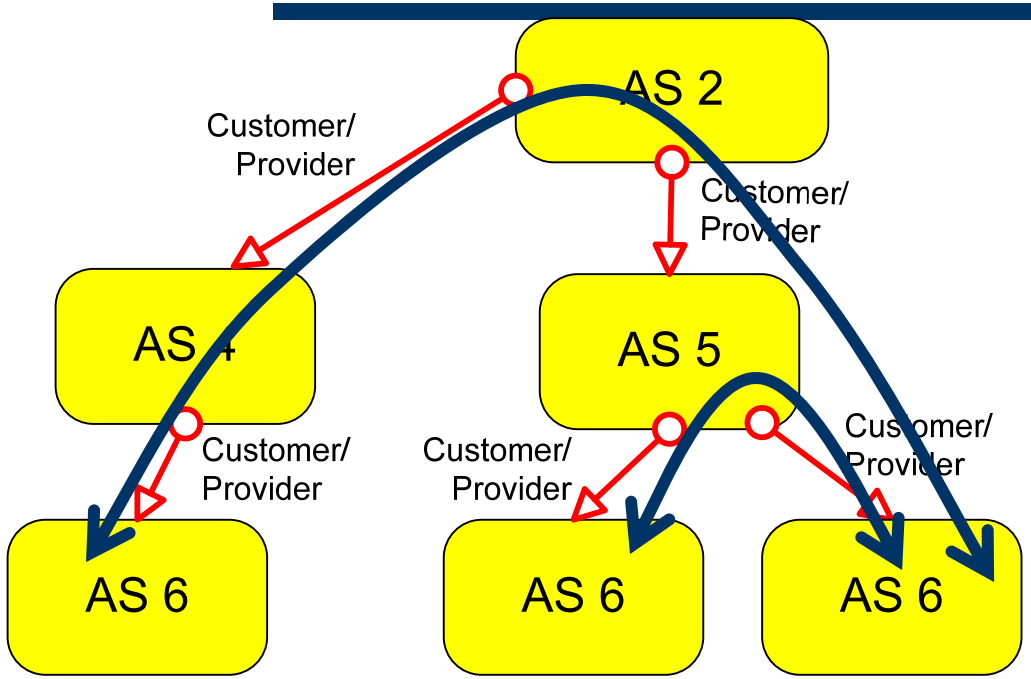
Example:

- Transit AS 3 carries traffic between AS 1 and AS 4 and between AS 2 and AS 4
- But AS 3 does **not** carry traffic between AS 1 and AS 2



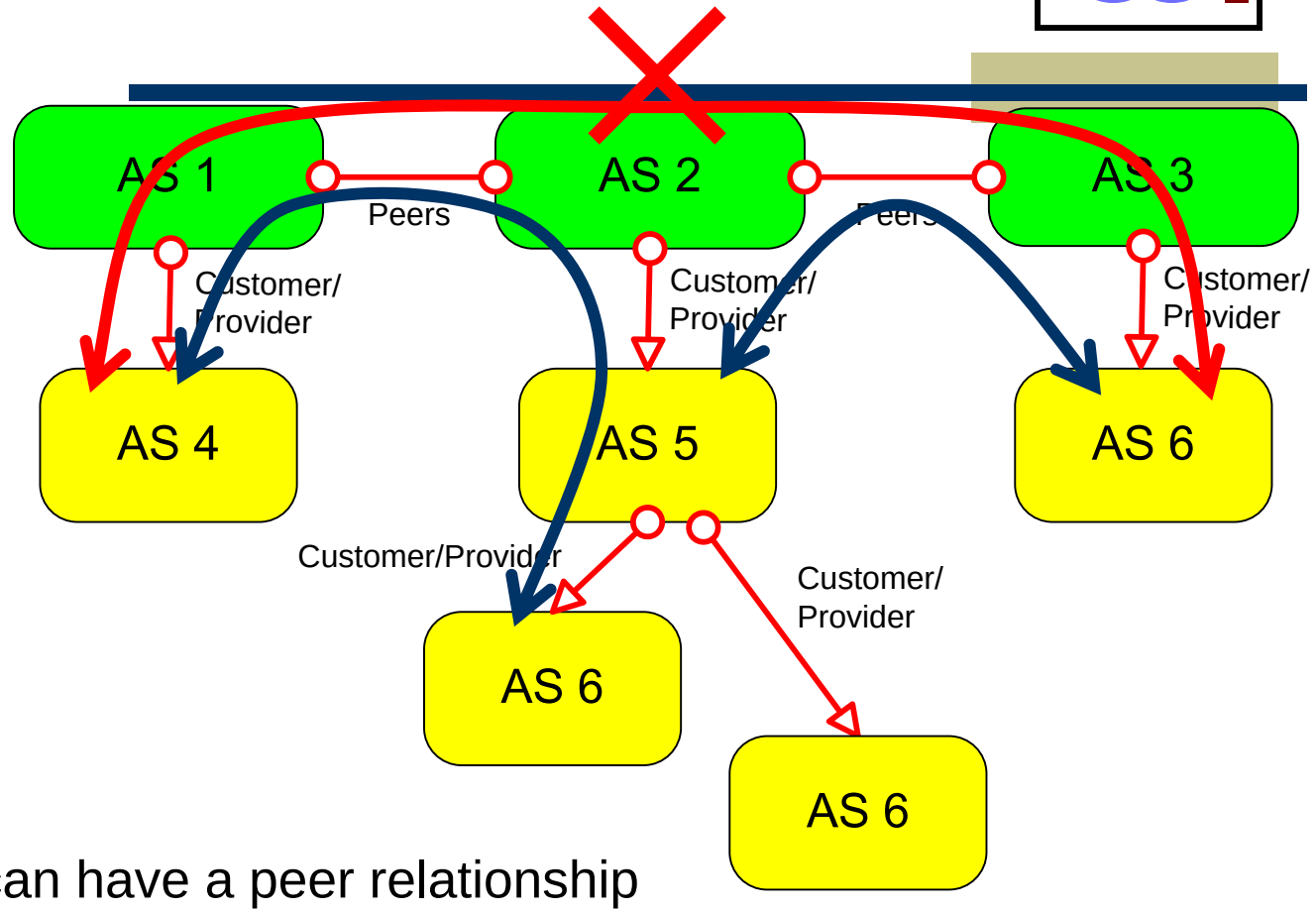
- The example shows a routing policy.

Customer/Provider



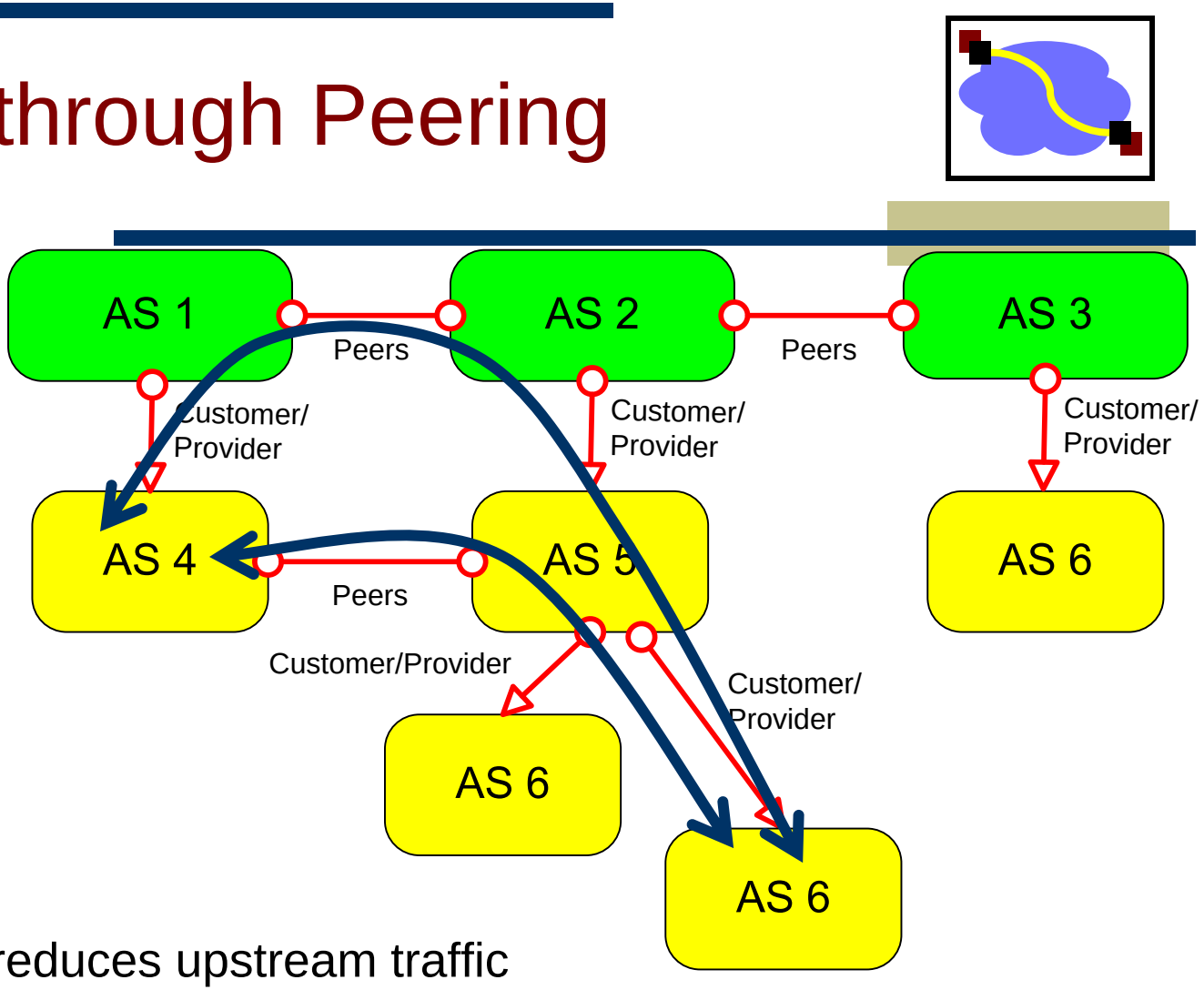
- A stub network typically obtains access to the Internet through a transit network.
- Transit network that is a provider may be a customer for another network
- Customer pays provider for service

Customer/Provider and Peers



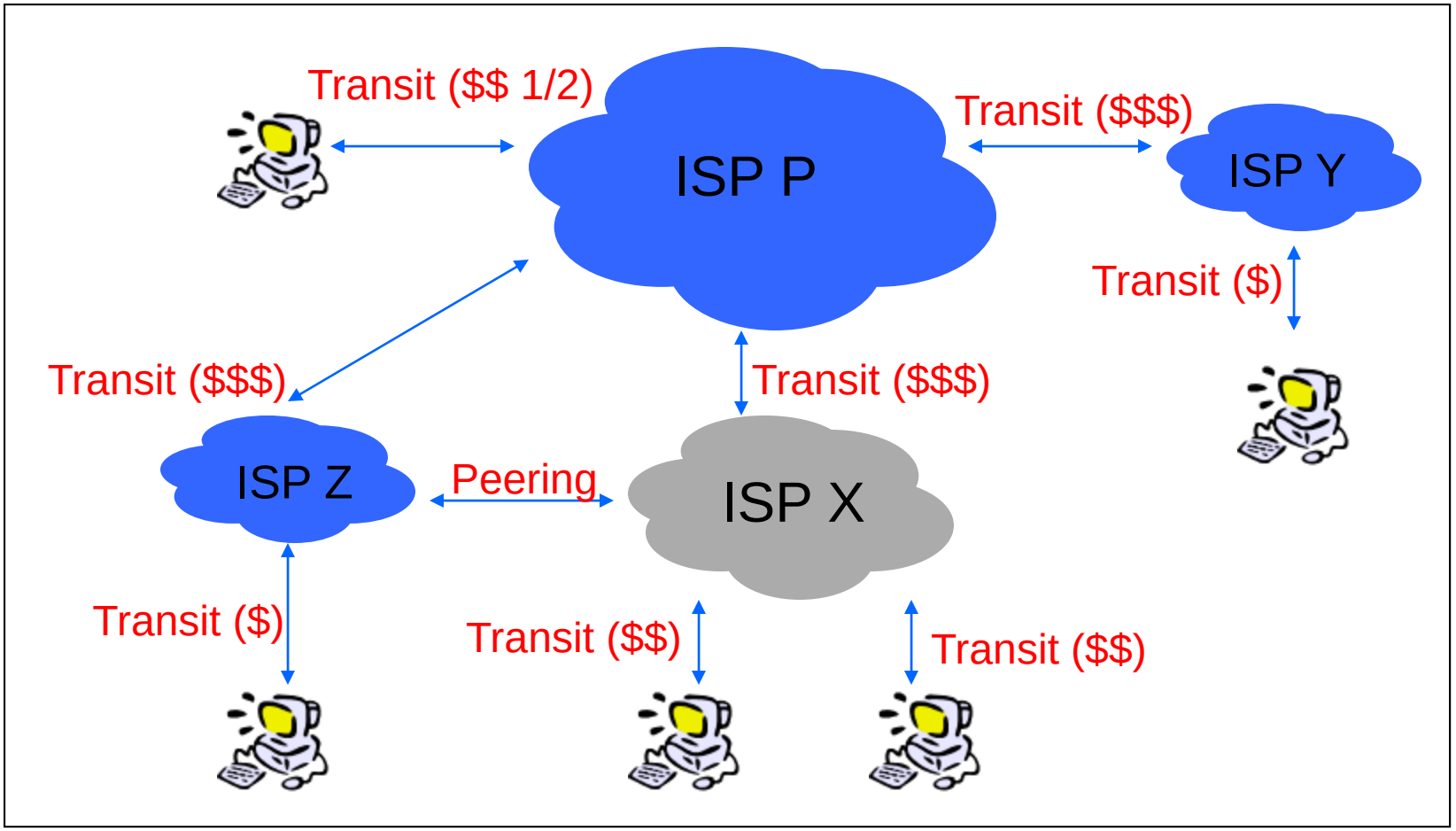
- Transit networks can have a peer relationship
- Peers provide transit between their respective customers
- Peers do not provide transit between peers
- Peers normally do not pay each other for service

Shortcuts through Peering

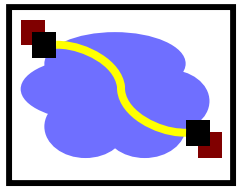


- Note that peering reduces upstream traffic
- Delays can be reduced through peering
- But: Peering may not generate revenue

Transit vs. Peering

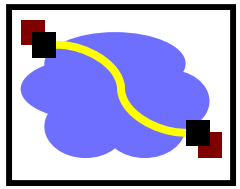


Policy Impact



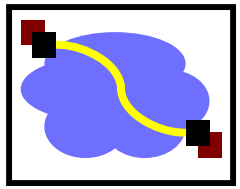
- "Valley-free" routing
  - Number links as (+1, 0, -1) for provider, peer and customer
  - In any path should only see sequence of +1, followed by at most one 0, followed by sequence of -1
- WHY?
  - Consider the economics of the situation

Border Gateway Protocol (BGP)



- Border Gateway Protocol is the interdomain routing protocol for the Internet for routing between autonomous systems
- Currently in version 4 (2006)
  - Network administrators can specify routing policies
  - BGP is a distance/path vector protocol (However, routing messages in BGP contain complete routes)
- Uses TCP to transmit routing messages

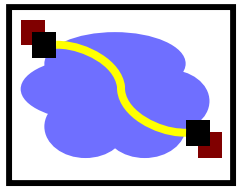
Border Gateway Protocol (BGP)



- An autonomous system uses BGP to advertise its network address(es) to other AS's

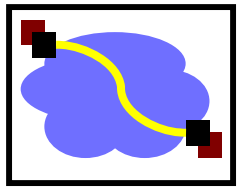
- BGP helps an autonomous system with the following:
  1. Collect information about reachable networks from neighboring AS's
  2. Disseminate the information about reachable networks to routers inside the AS and to neighboring AS's
  3. Picks routes if there are multiple routes available

Internet inter-AS routing: BGP

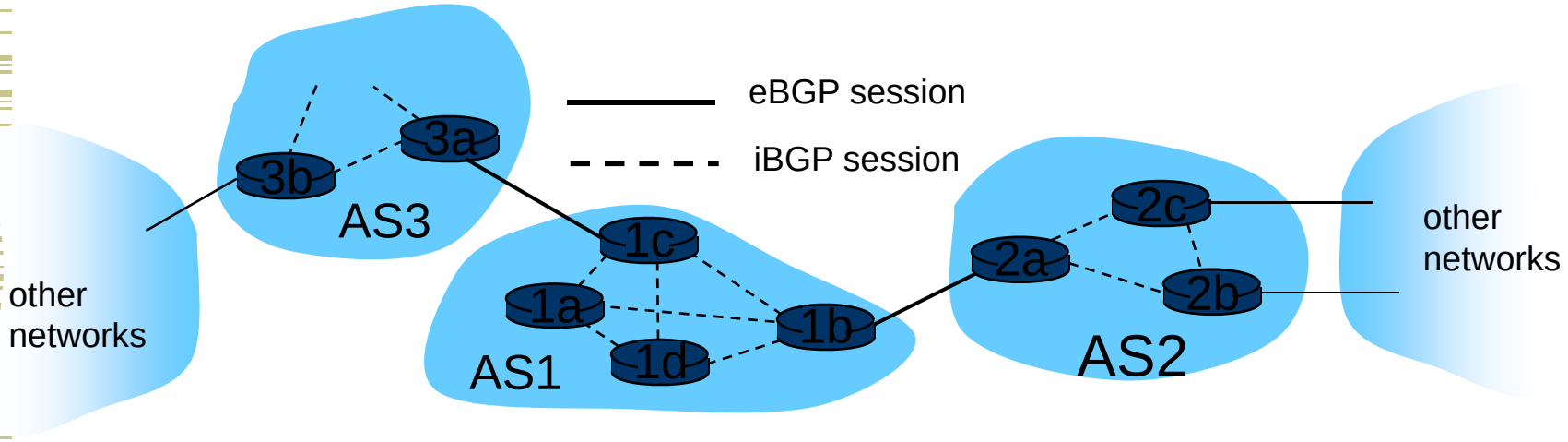


- BGP (Border Gateway Protocol): *the* de facto inter-domain routing protocol
  - "glue that holds the Internet together"
- BGP provides each AS a means to:
  - eBGP: obtain subnet reachability information from neighboring ASs.
  - iBGP: propagate reachability information to all AS-internal routers.
  - determine "good" routes to other networks based on reachability information and policy.
- allows subnet to advertise its existence to rest of Internet: *"I am here"*

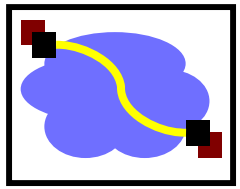
BGP basics: distributing path information



- using eBGP session between 3a and 1c, AS3 sends prefix reachability info to AS1.
  - 1c can then use iBGP do distribute new prefix info to all routers in AS1
  - 1b can then re-advertise new reachability info to AS2 over 1b-to-2a eBGP session
- when router learns of new prefix, it creates entry for prefix in its forwarding table.

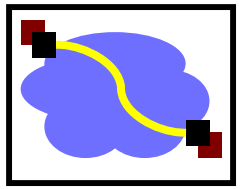


Outline



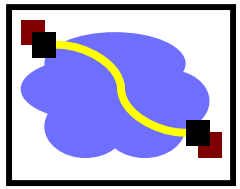
- Routing hierarchy
- Internet structure
- External BGP (E-BGP)
- Internal BGP (I-BGP)

Choices



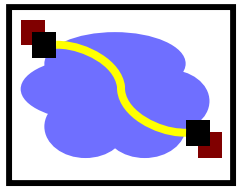
- Link state or distance vector?
  - No universal metric – policy decisions
- Problems with distance-vector:
  - Bellman-Ford algorithm may not converge
- Problems with link state:
  - Metric used by routers not the same – loops
  - LS database too large – entire Internet
  - May expose policies to other AS's

Solution: Distance Vector with Path



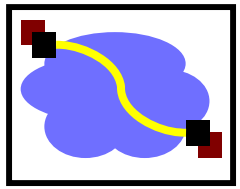
- Each routing update carries the entire path
- Loops are detected as follows:
  - When AS gets a route, check if AS already in path
    - If yes, reject route
    - If no, add self and (possibly) advertise route further
- Advantage:
  - Metrics are local - AS chooses path, protocol ensures no loops

Interconnecting BGP Peers



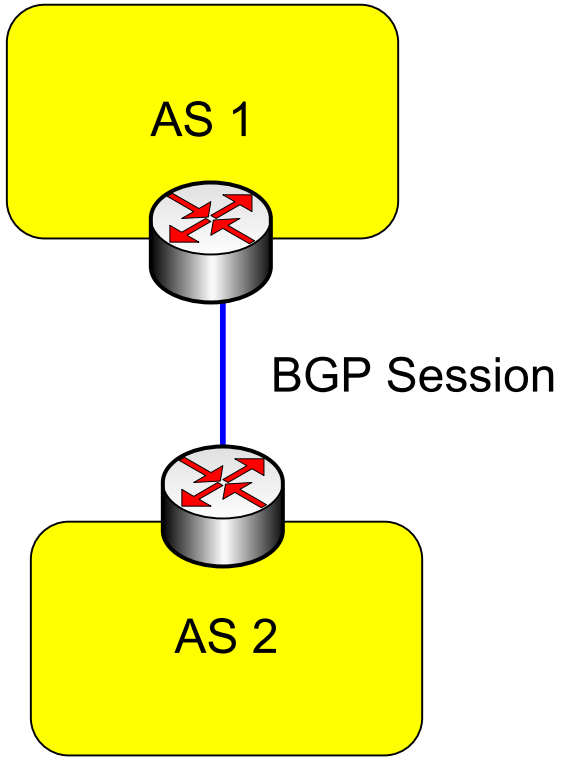
- BGP uses TCP to connect peers
- Advantages:
  - Simplifies BGP
  - No need for periodic refresh - routes are valid until withdrawn, or the connection is lost
  - Incremental updates
- Disadvantages
  - Congestion control on a routing protocol?
  - Poor interaction during high load

BGP Interactions

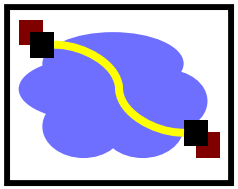


- Router establishes a TCP connection (TCP port 175)
- Routers exchange BGP routes
- Periodically send updates
- BGP is executed between two routers
  - BGP session
  - BGP peers or BGP speakers

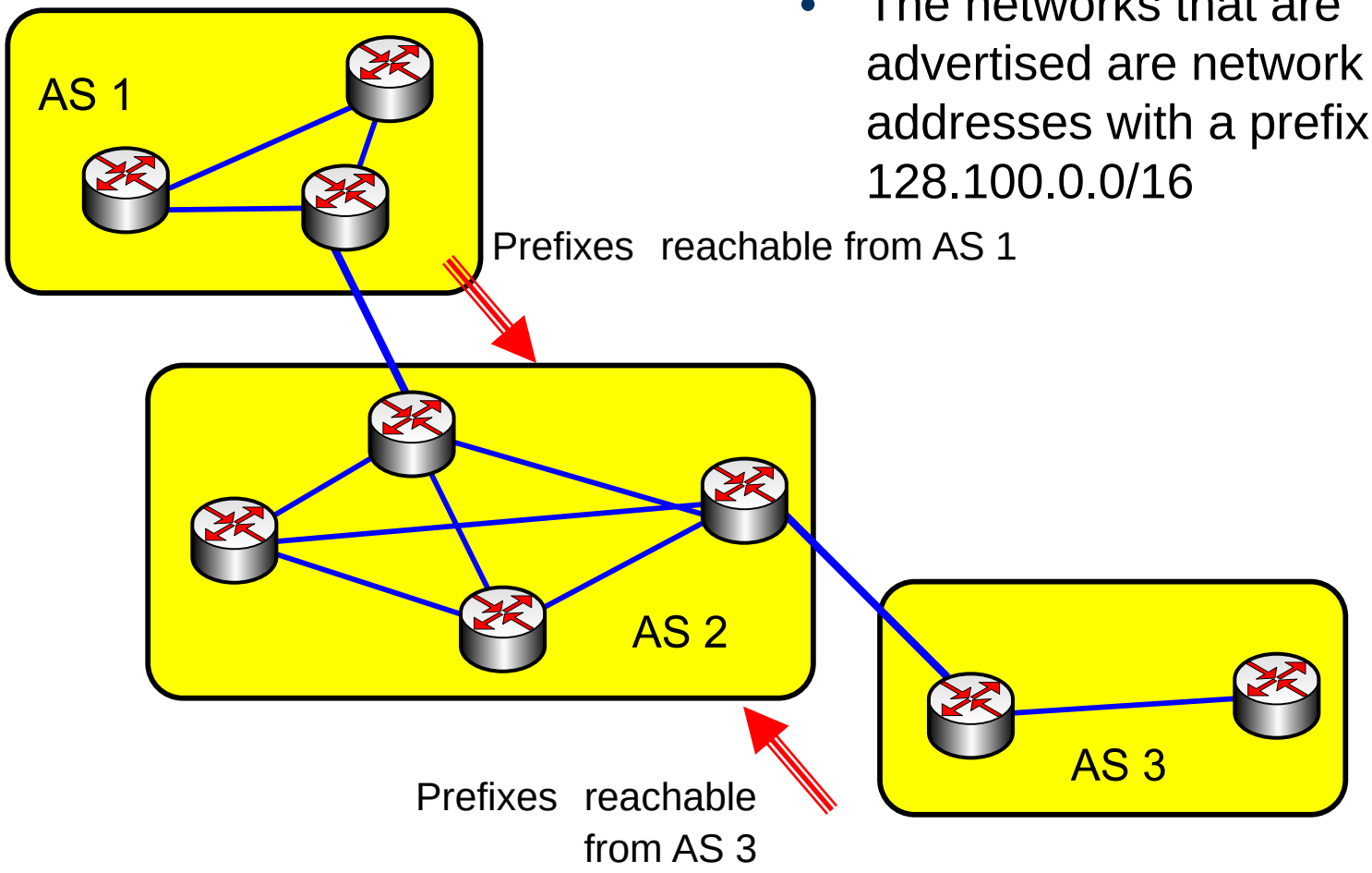
- **Note:** Not all autonomous systems need to run BGP. On many stub networks, the route to the provider can be statically configured



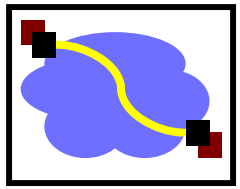
BGP Interactions



- The networks that are advertised are network IP addresses with a prefix, E.g., 128.100.0.0/16

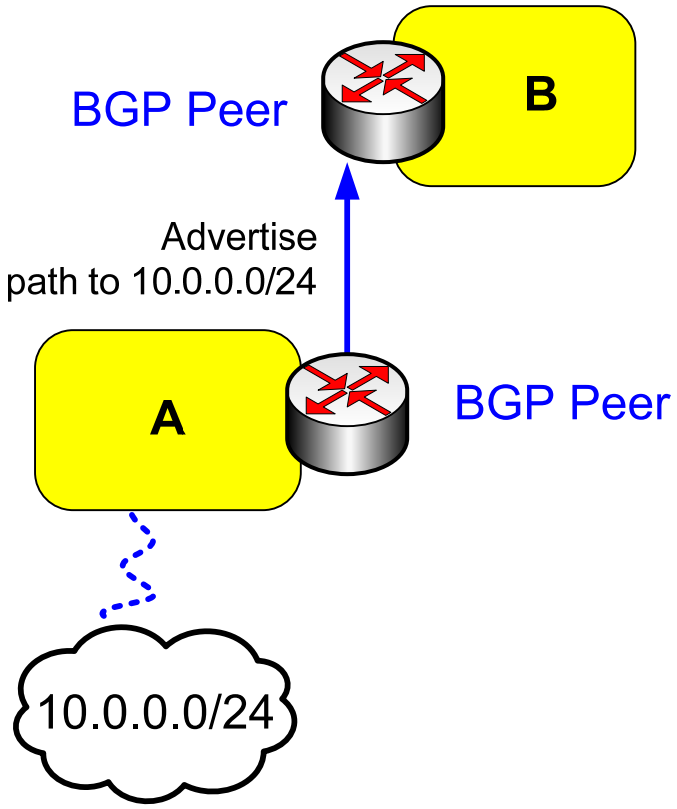


BGP Interactions

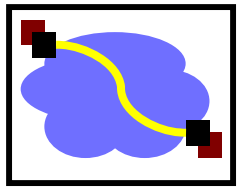


- BGP peers advertise reachability of IP networks
- A advertises a path to a network (e.g., 10.0.0.0/24) to B only if it is willing to forward traffic going to that network
- **Path-Vector:**
  - A advertises the complete path to the advertised network
  - Path is sent as a list of AS's

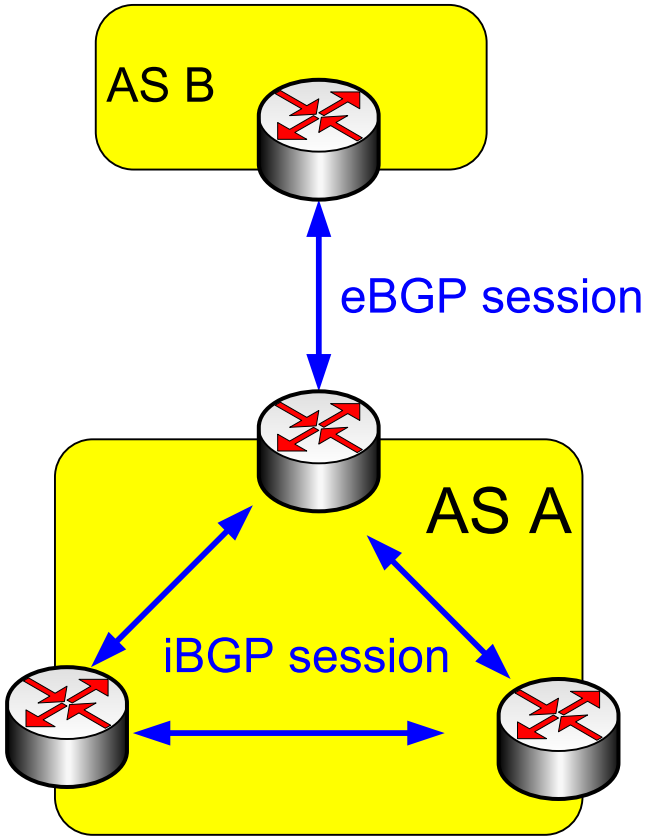
  → this avoids loops



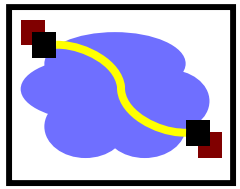
BGP Sessions



- **External BGP session (eBGP):**
  Peers are in different AS'es
- **Internal BGP session (iBGP)**
  Peers are in same AS
- Note that iBGP sessions are going over routes that are set up by an intradomain routing protocol!

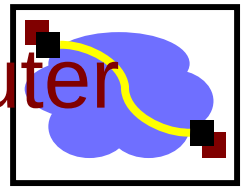


Hop-by-hop Model

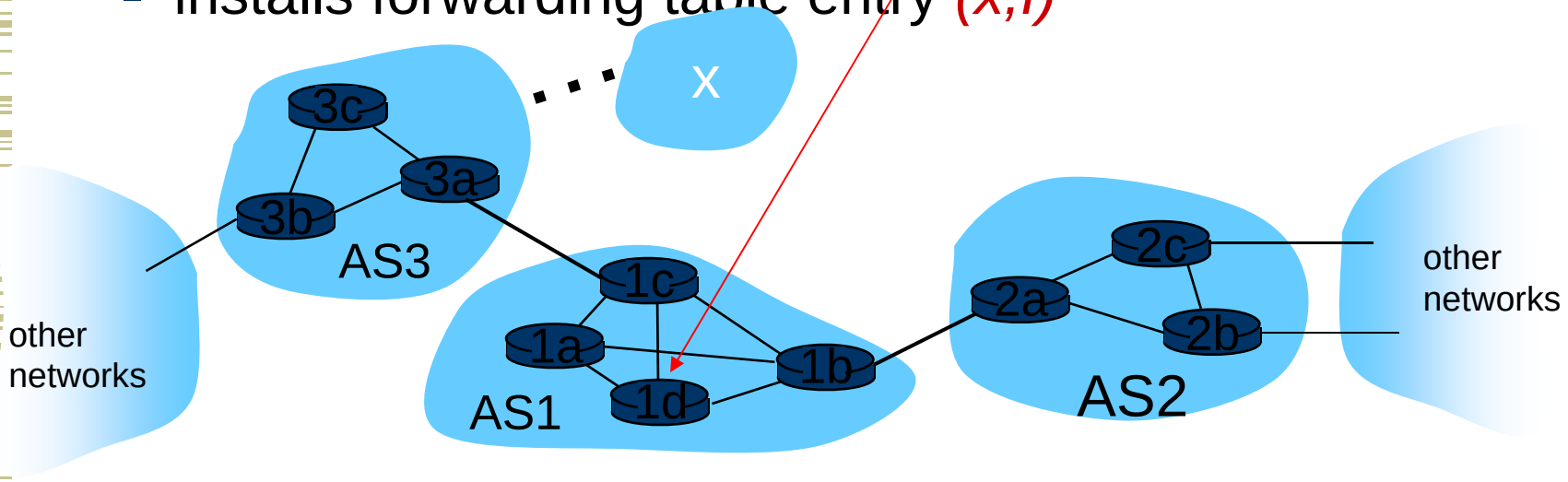


- BGP advertises to neighbors only those routes that it uses
  - Consistent with the hop-by-hop Internet paradigm
  - e.g., AS1 cannot tell AS2 to route to other AS's in a manner different than what AS1 has chosen (need source routing for that)

Example: setting forwarding table in router 1d

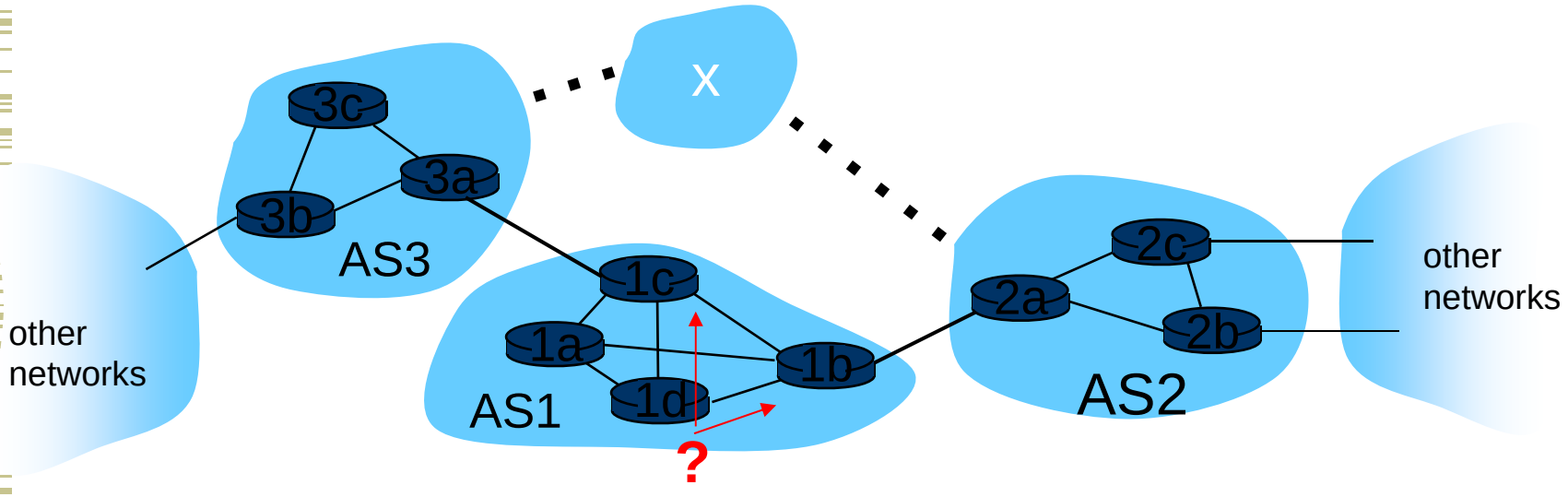


- suppose AS1 learns (via inter-AS protocol) that subnet *x* reachable via AS3 (gateway 1c), but not via AS2
  - inter-AS protocol propagates reachability info to all internal routers
- router 1d determines from intra-AS routing info that its interface *I* is on the least cost path to 1c
  - installs forwarding table entry *(x,I)*



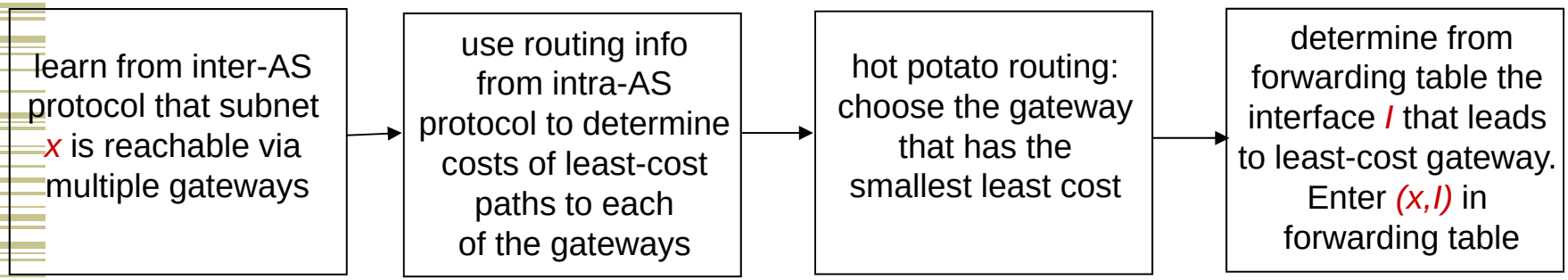
Example: choosing among multiple ASes

- now suppose AS1 learns from inter-AS protocol that subnet *x* is reachable from AS3 *and* from AS2.
- to configure forwarding table, router 1d must determine which gateway it should forward packets towards for dest x
  - this is also job of inter-AS routing protocol!

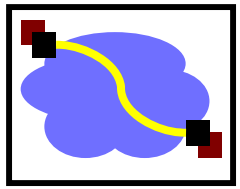


Example: choosing among multiple ASes

- now suppose AS1 learns from inter-AS protocol that subnet *x* is reachable from AS3 *and* from AS2.
- to configure forwarding table, router 1d must determine towards which gateway it should forward packets for dest x
  - this is also job of inter-AS routing protocol!
- *hot potato routing: send* packet towards closest of two routers.

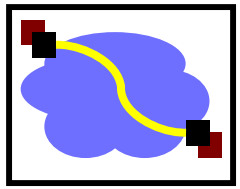


Path attributes and BGP routes



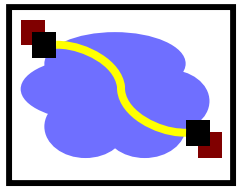
- advertised prefix includes BGP attributes
  - prefix + attributes = “route”
- two important attributes:
  - AS-PATH: contains ASs through which prefix advertisement has passed: e.g., AS 67, AS 17
  - NEXT-HOP: indicates specific internal-AS router to next-hop AS. (may be multiple links from current AS to next-hop-AS)
- gateway router receiving route advertisement uses import policy to accept/decline
  - e.g., never route through AS x
  - *policy-based* routing

BGP route selection



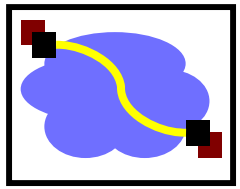
- router may learn about more than 1 route to destination AS, selects route based on:
  1. local preference value attribute: policy decision
  2. shortest AS-PATH
  3. closest NEXT-HOP router: hot potato routing
  4. additional criteria

Policy with BGP



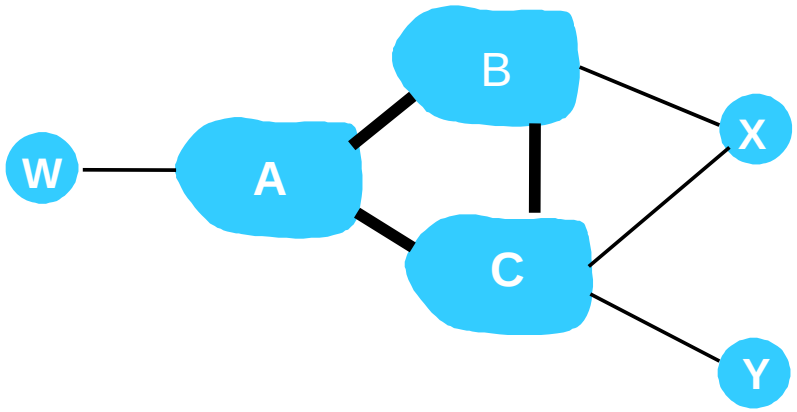
- BGP provides capability for enforcing various policies
- Policies are **not** part of BGP: they are provided to BGP as configuration information
- BGP enforces policies by choosing paths from multiple alternatives and controlling advertisement to other AS's

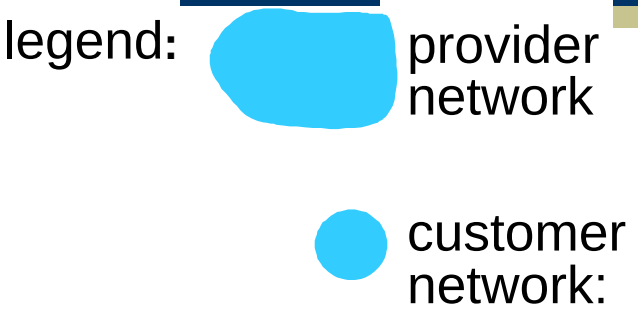
Examples of BGP Policies



- A multi-homed AS refuses to act as transit
  - Limit path advertisement
- A multi-homed AS can become transit for some AS's
  - Only advertise paths to some AS's
- An AS can favor or disfavor certain AS's for traffic transit from itself

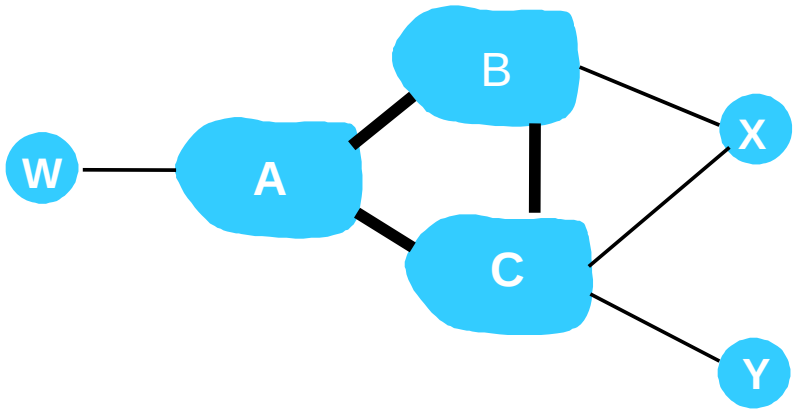
BGP routing policy

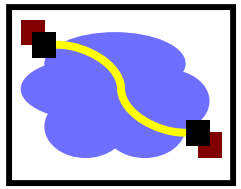




- A,B,C are *provider networks*
- X,W,Y are customer (of provider networks)
- X is *dual-homed:* attached to two networks
  - X does not want to route from B via X to C
  - .. so X will not advertise to B a route to C

BGP routing policy (2)





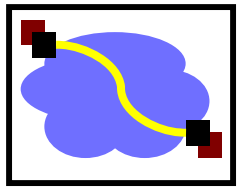
legend:

provider network

customer network:

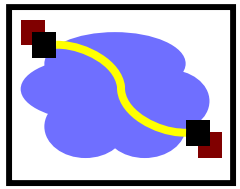
- A advertises path AW to B
- B advertises path BAW to X
- Should B advertise path BAW to C?
  - No way! B gets no "revenue" for routing CBAW since neither W nor C are B's customers
  - B wants to force C to route to w via A
  - B wants to route *only* to/from its customers!

BGP Messages



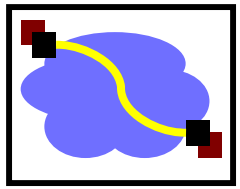
- Open
  - Announces AS ID
  - Determines hold timer – interval between keep_alive or update messages, zero interval implies no keep_alive
- Keep_alive
  - Sent periodically (but before hold timer expires) to peers to ensure connectivity.
  - Sent in place of an UPDATE message
- Notification
  - Used for error notification
  - TCP connection is closed *immediately* after notification

BGP UPDATE Message



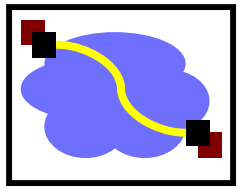
- List of withdrawn routes
- Network layer reachability information
  - List of reachable prefixes
- Path attributes
  - Origin
  - Path
  - Metrics
- All prefixes advertised in message have same path attributes

Path Selection Criteria



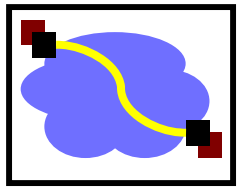
- Information based on path attributes
- Attributes + external (policy) information
- Examples:
  - Hop count
  - Policy considerations
    - Preference for AS
    - Presence or absence of certain AS
  - Path origin
  - Link dynamics

BGP Message Types



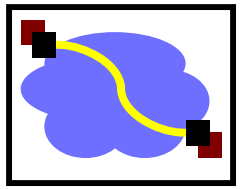
- **Open**: Establishes a peering session
- **Keep Alive**: Handshake at regular intervals to maintain peering session
- **Notification**: Closes a peering session
- **Update**: Advertises new routes or withdraws previously announced routes. Each announced route is specified as a network prefix with attribute values

Content of Advertisements

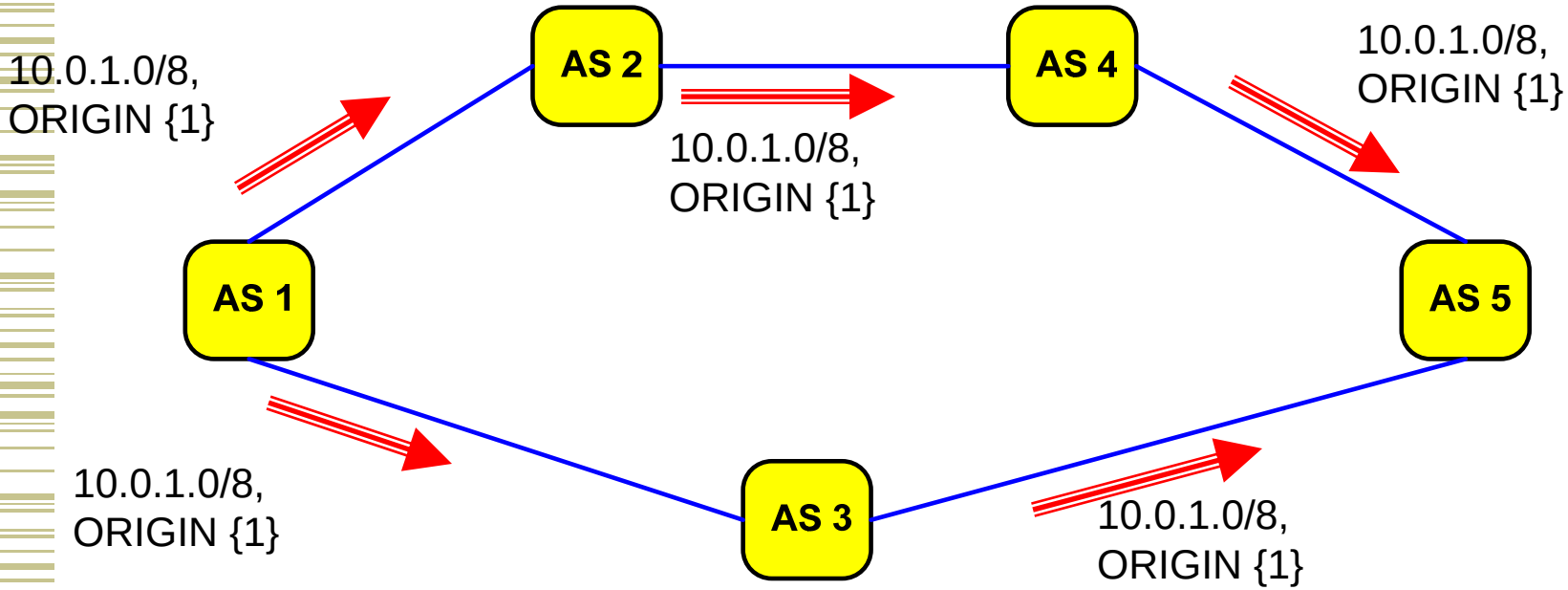


- BGP routers advertise **routes**
- Each route consists of a network prefix and a list of **attributes** that specify information about a route
- Mandatory attributes:
  - **ORIGIN**
  - **AS_PATH**
  - **NEXT_HOP**
- Many other attributes

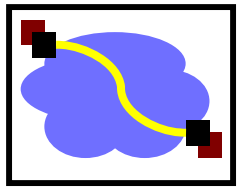
ORIGIN Attribute



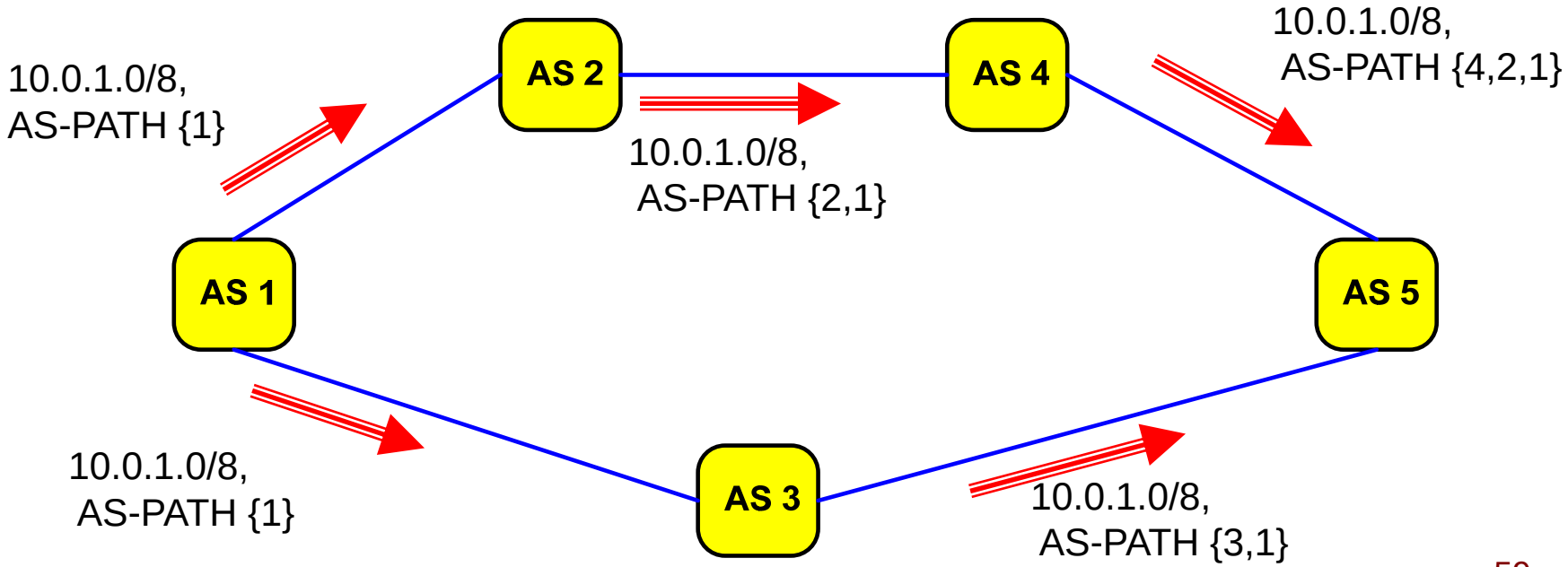
- Originating domain sends a route with ORIGIN attribute
- ORIGIN attributes also specifies if the origin is internal to the AS or not



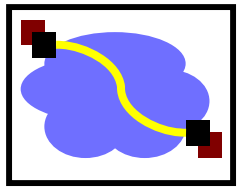
AS-PATH Attributes



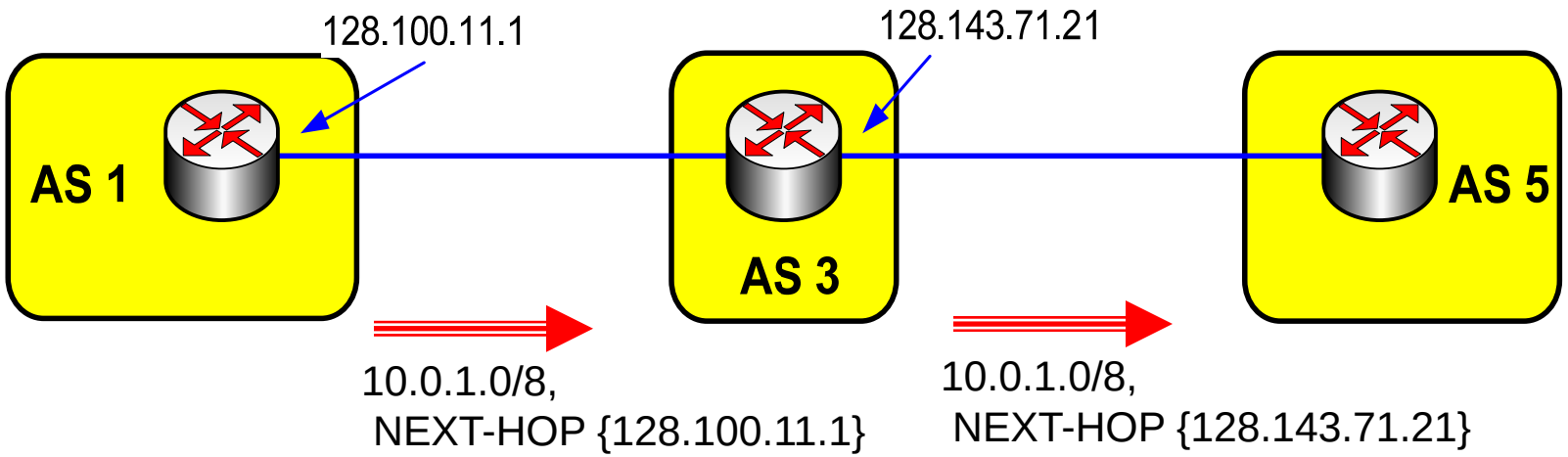
- Each AS that propagates a route prepends its own AS number
  - AS-PATH collects a path to reach the network prefix
- Path information prevents routing loops from occurring
- Path information also provides information on the length of a path (By default, a shorter route is preferred)
- Note: BGP aggregates routes according to CIDR rules



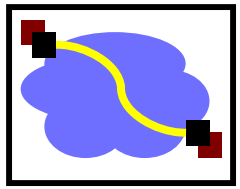
NEXT-HOP Attributes

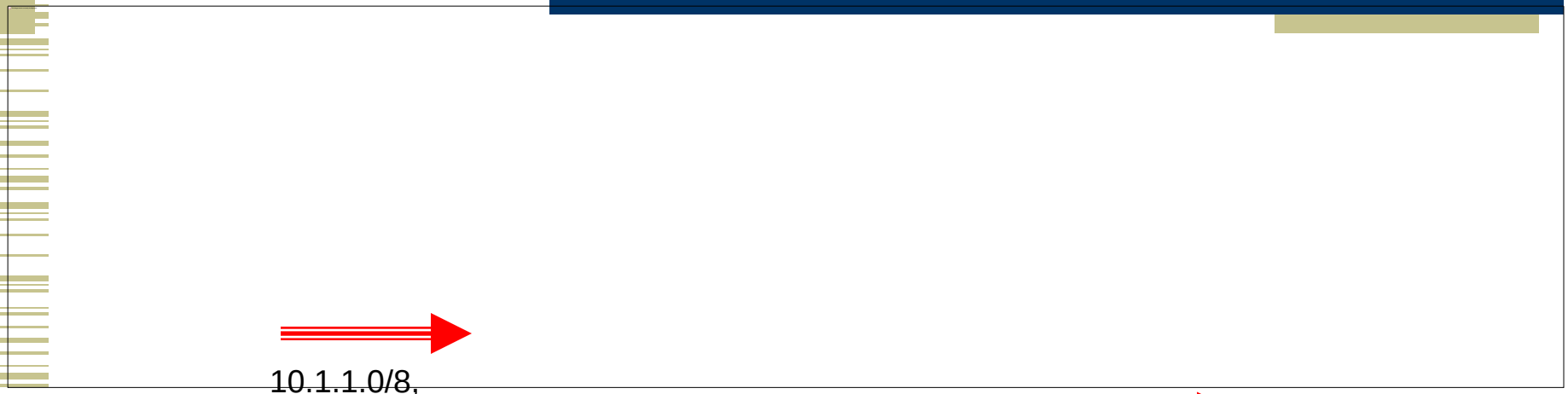


- Each router that sends a route advertisement it includes its own IP address in a NEXT-HOP attribute
- The attribute provides information for the routing table of the receiving router.



Connecting NEXT-HOP with IGP information





At R1:

Routing table

Dest.	Next hop
128.100.11.0/24	192.0.1.2

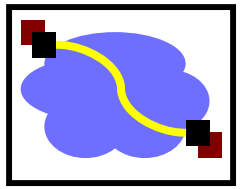
BGP info

Dest.	Next hop
10.1.1.0/8	128.100.11.1

Routing table

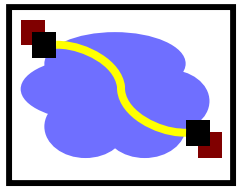
Dest.	Next hop
128.100.11.0/24	192.0.1.2
10.1.1.0/8	192.0.1.2

Route Selection



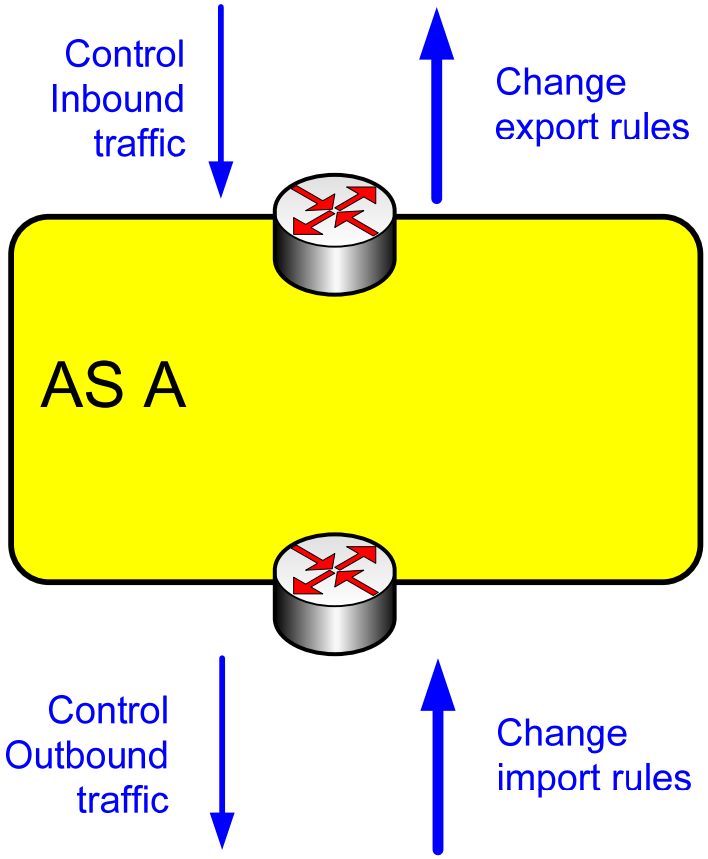
- Router may get more than one route to an address
- Rules for selecting a route (in order of priorities):
  - Preferences can be advertised as an attribute
  - Shorter routes are preferred
  - Close next-hop is preferred
- Router may not want to advertise some routes

Importing and Exporting Routes

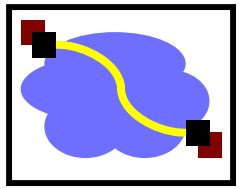


- An AS may not accept all routes that are advertised
- An AS may not advertise certain routes
- Route policies determines which routes are filtered

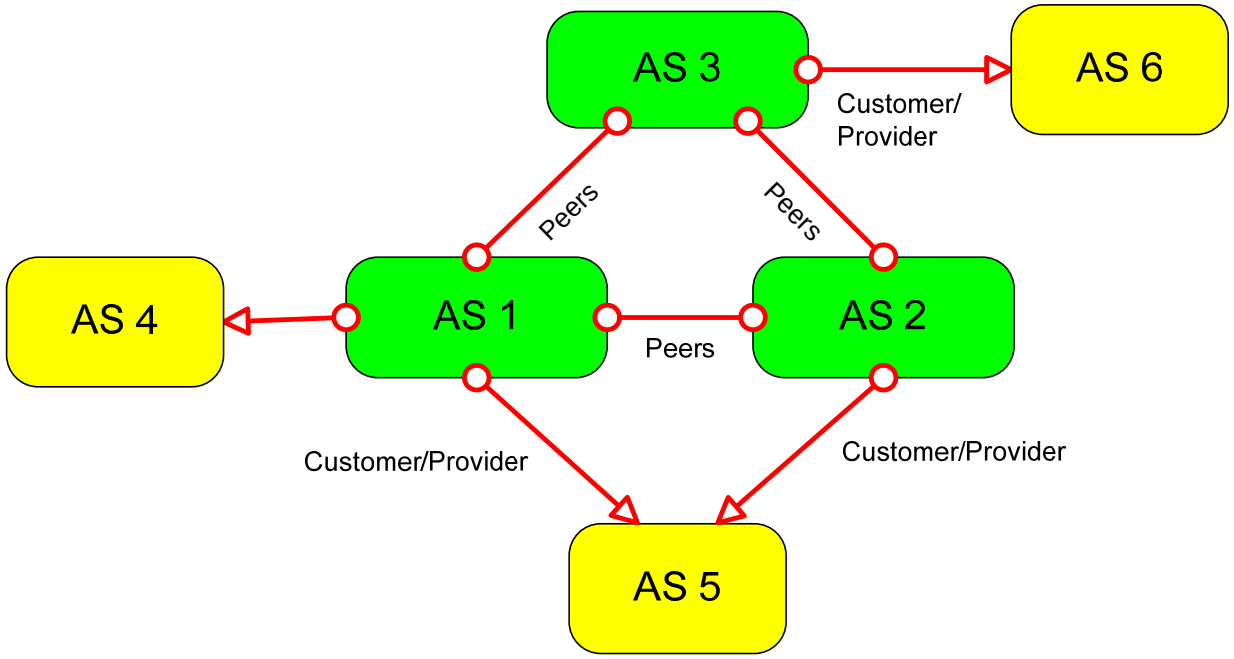
- If an AS wants to have less inbound traffic it should adapt its export rules
- If an AS wants to control its outbound traffic, it adapts its import rules



Routing Policies

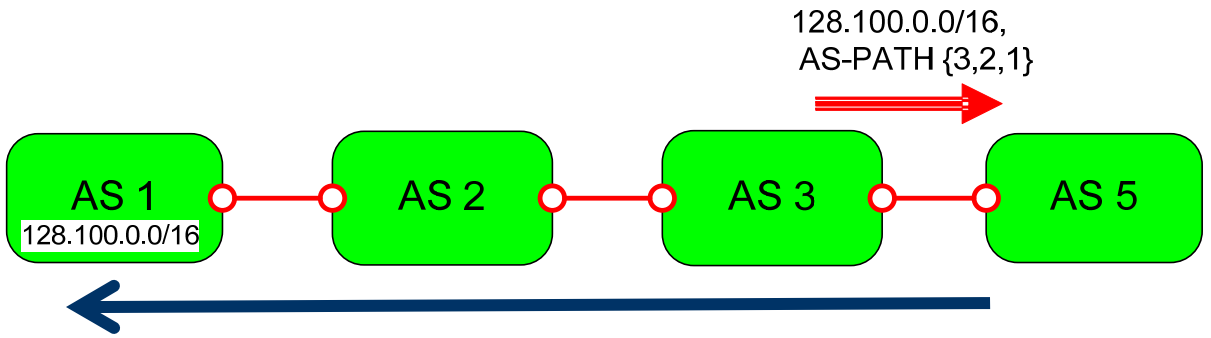


- Since AS 5 is a stub network it should not advertise routes to networks other than networks in AS 5
- When AS 3 learns about the path {AS1, AS4}, it should not advertise the route {AS3, AS1, AS4} to AS 2.



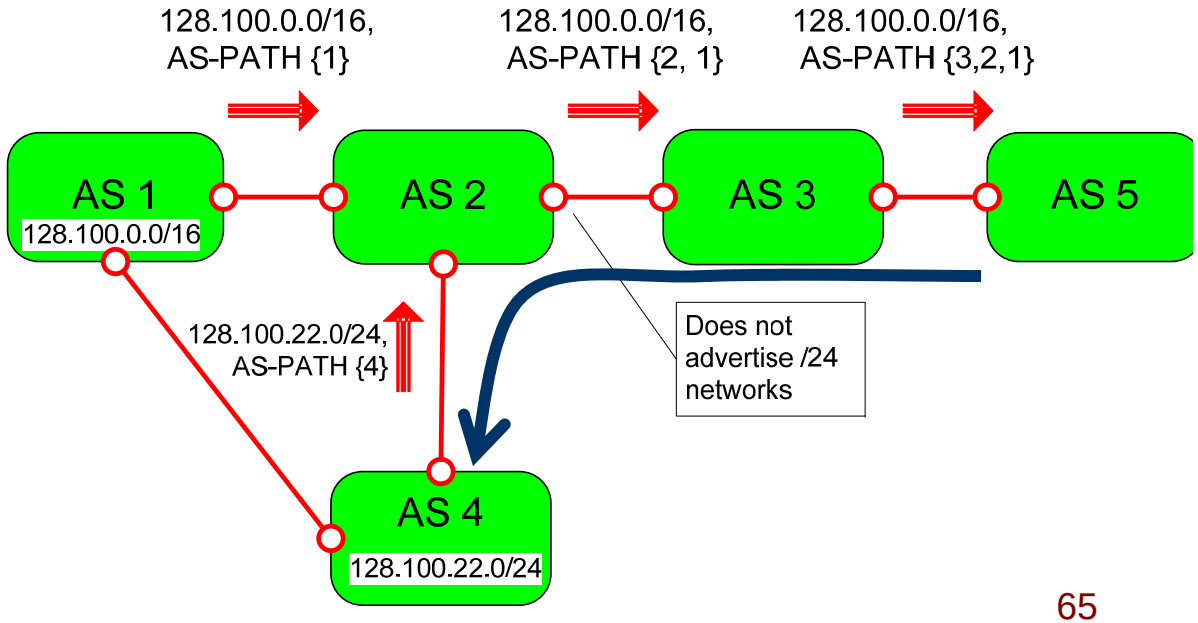
Traffic Often Follows ASPATH

- In many cases, packets are routed according to the AS-PATH



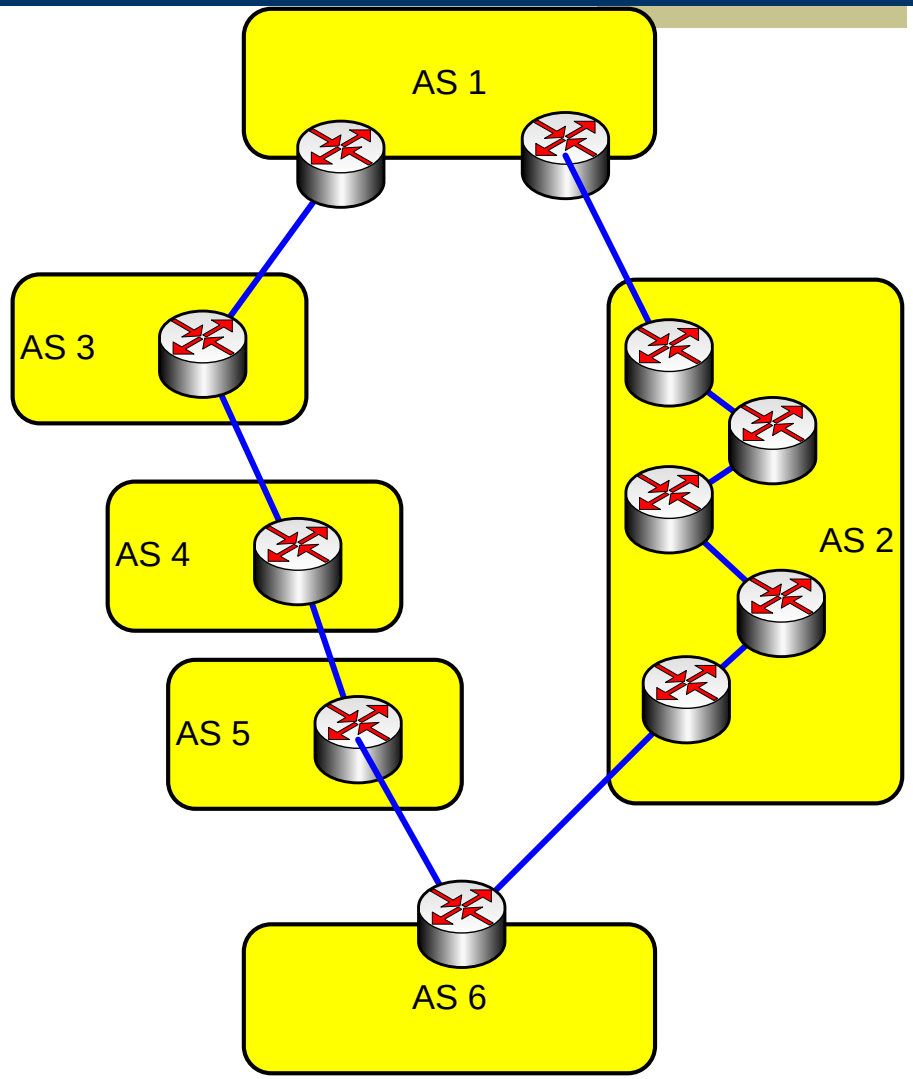
- However, in some cases this is not true

(Here: AS 2 filters routes with a long prefix)

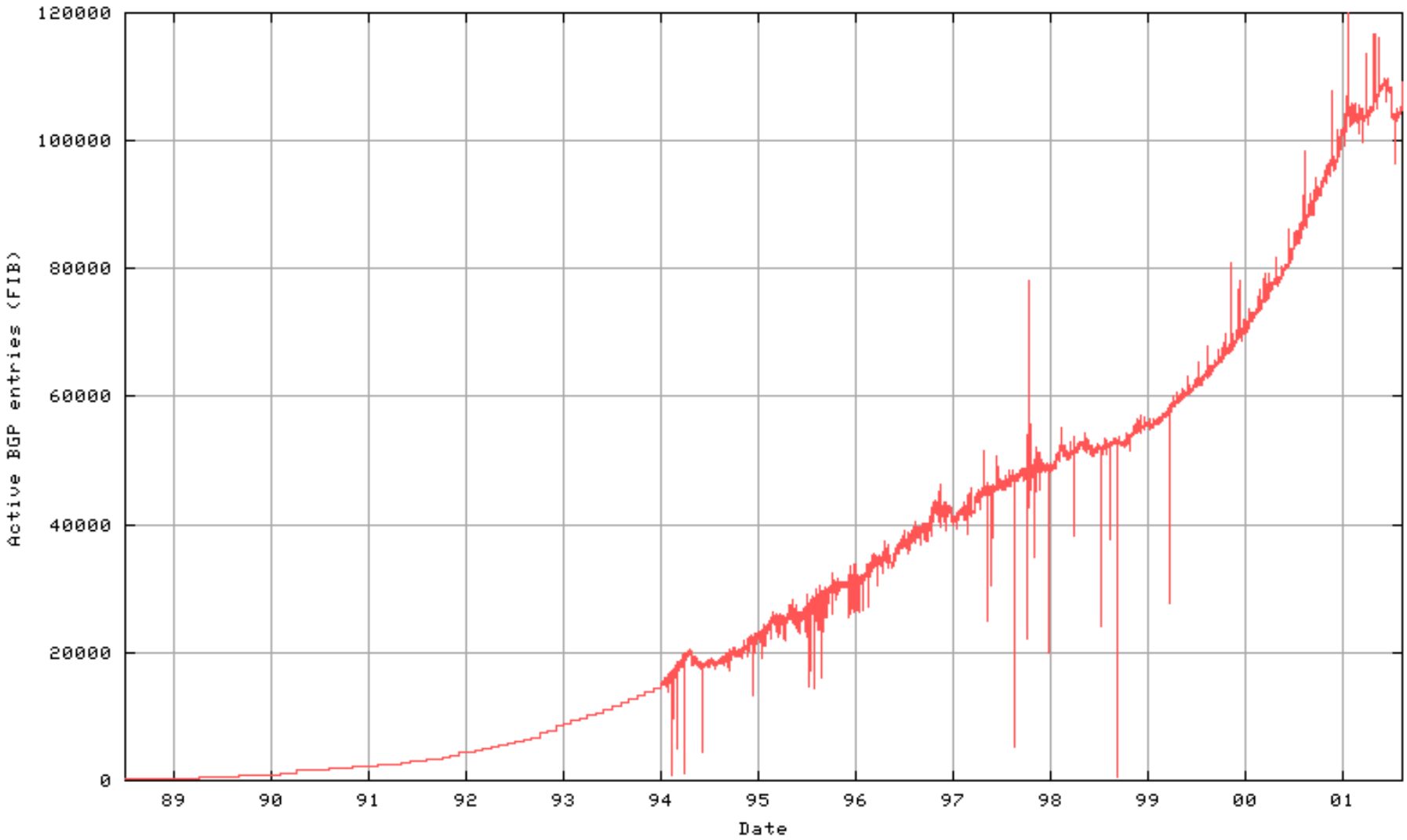


Short AS-PATH does not mean that route is short

- From AS 6's perspective
  - Path {AS2, AS1} is short
  - Path {AS5, AS4, AS3, AS1} is long
- But the number of traversed routers is larger when using the shorter AS-PATH

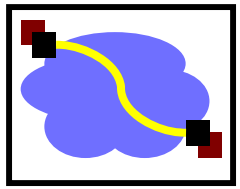


BGP Table Growth



Source: Geoff Huston. http://www.telstra.net/ops/bgptable.html on August 8, 2001

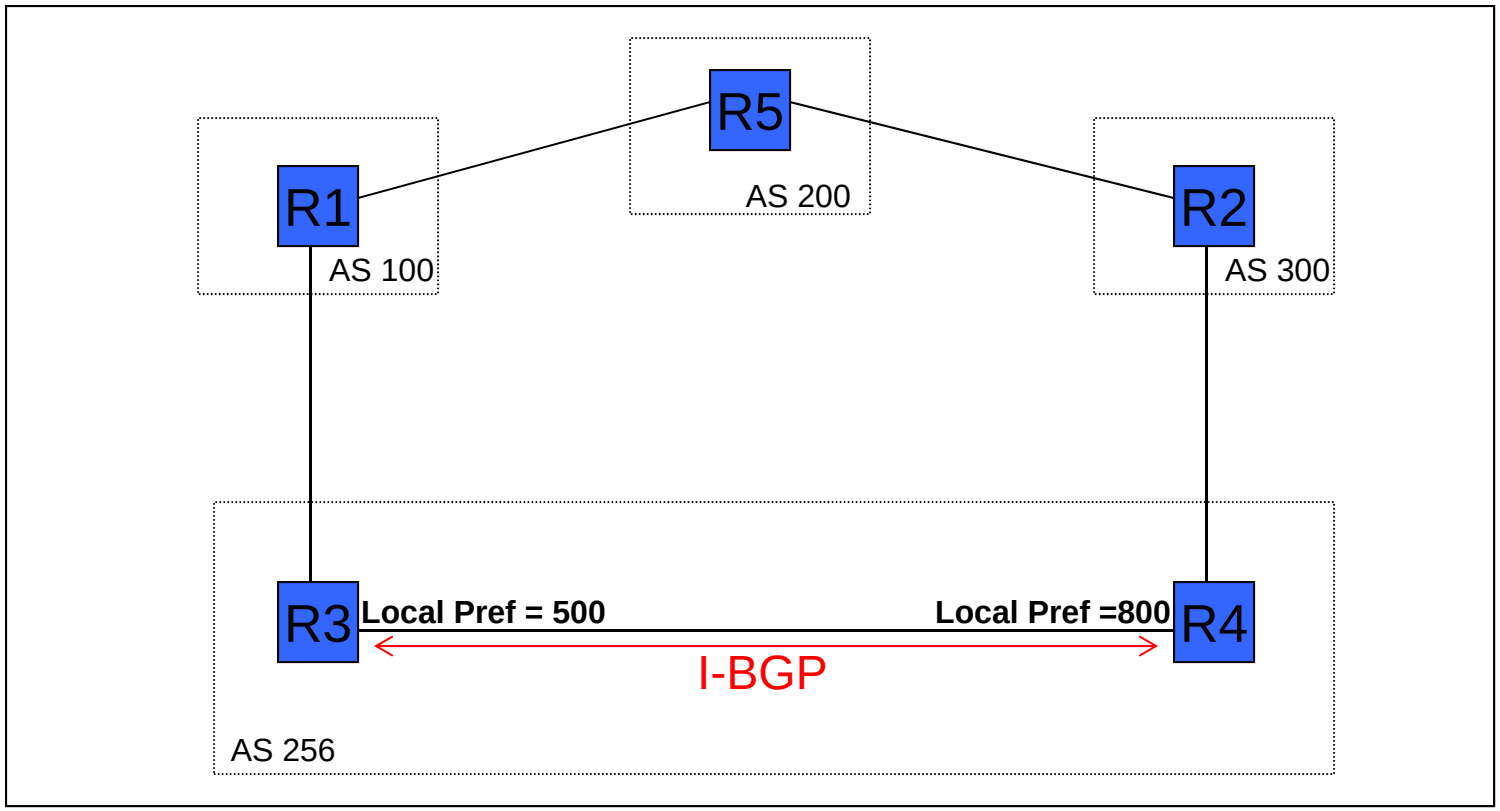
BGP Issues



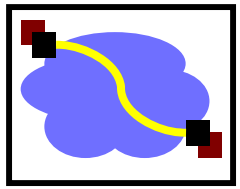
- BGP is a simple protocol but it is very difficult to configure
- BGP has severe stability issue due to policies → BGP is known to not converge
- As of July 2005, 39,000 AS numbers (of available 64,510) are consumed
- As of 2014, more than 47,000 Ases; move towards the use of 32-bit AS numbers

LOCAL PREF

- Local (within an AS) mechanism to provide relative priority among BGP routers



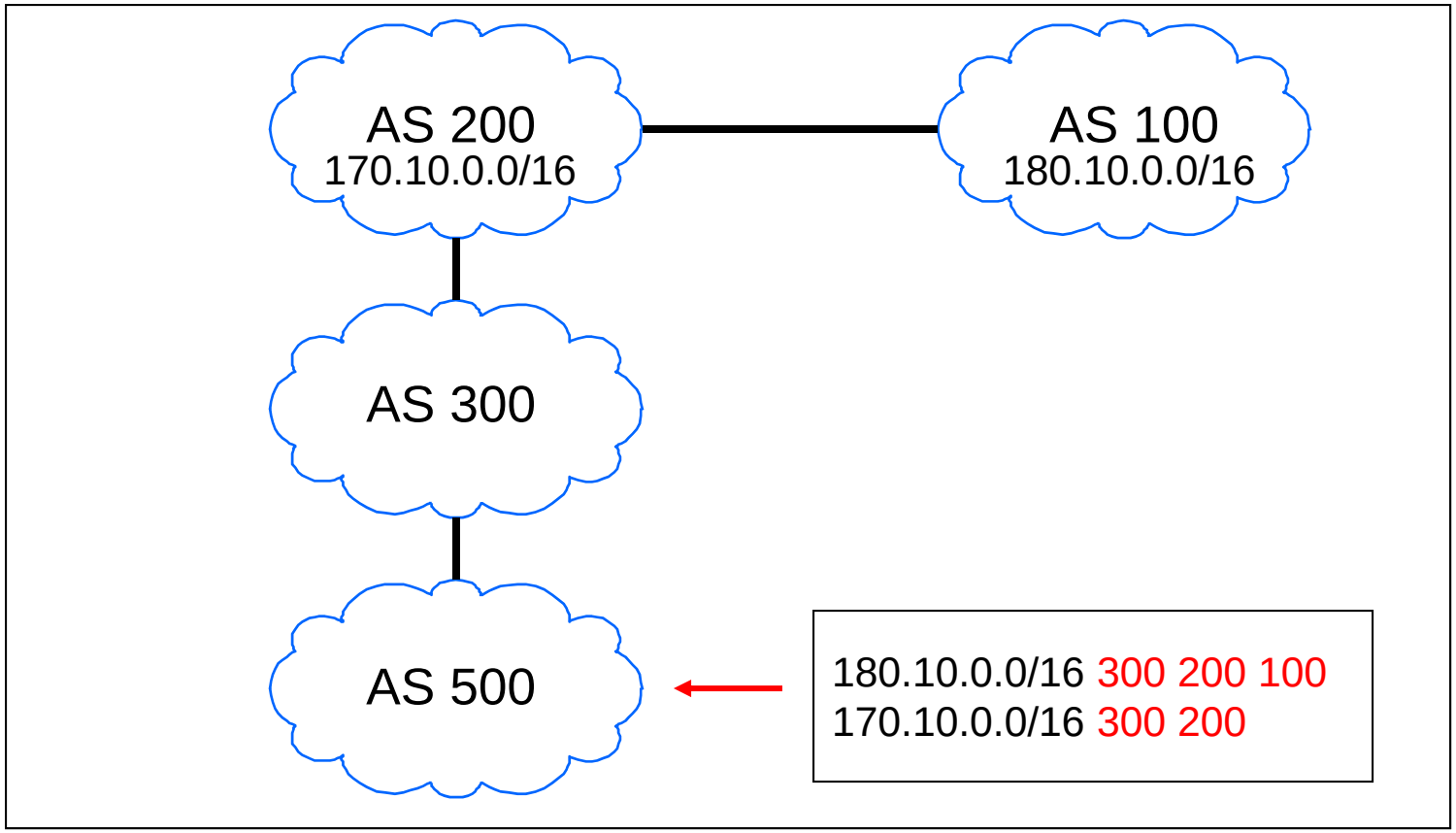
LOCAL PREF – Common Uses



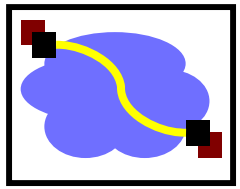
- Handle routes advertised to multi-homed transit customers
  - Should use direct connection
- Peering vs. transit
  - Prefer to use peering connection, why?
- In general, customer > peer > provider
  - Use LOCAL PREF to ensure this

AS_PATH

- List of traversed AS's



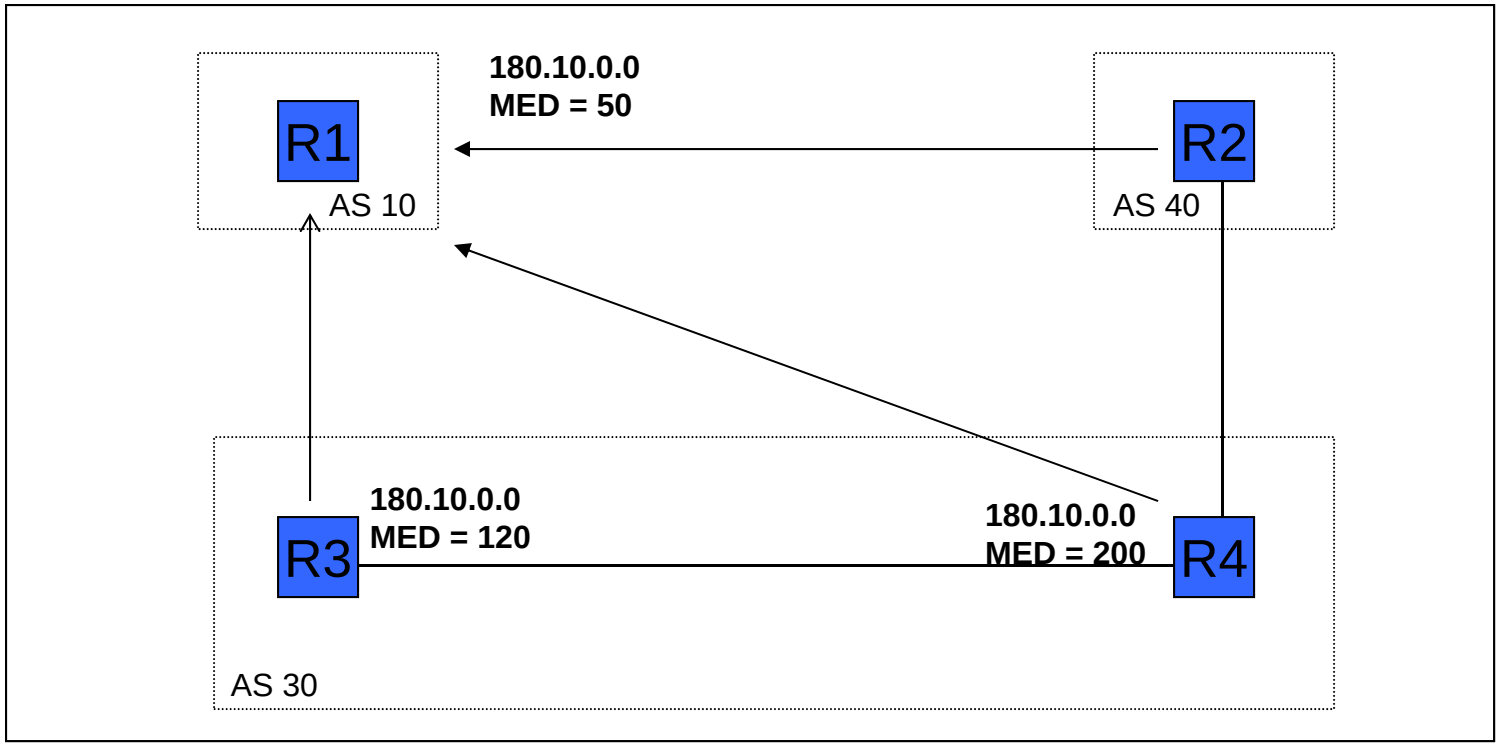
Multi-Exit Discriminator (MED)



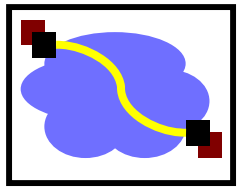
- Hint to external neighbors about the preferred path into an AS
  - Non-transitive attribute (we will see later why)
  - Different AS choose different scales
- Used when two AS's connect to each other in more than one place

MED

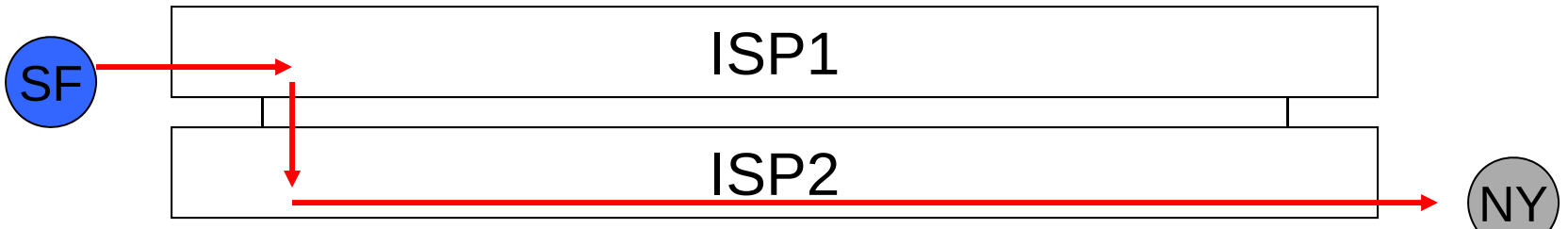
- Hint to R1 to use R3 over R4 link
- Cannot compare AS40's values to AS30's



MED

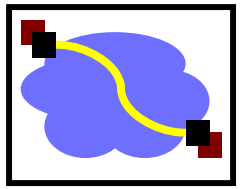


- MED is typically used in provider/subscriber scenarios
- It can lead to unfairness if used between ISP because it may force one ISP to carry more traffic:



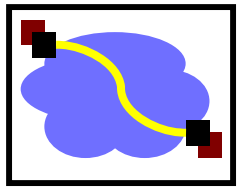
- ISP1 ignores MED from ISP2
- ISP2 obeys MED from ISP1
- ISP2 ends up carrying traffic most of the way

Decision Process



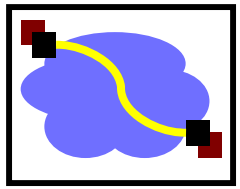
- Processing order of attributes:
  - Select route with highest LOCAL-PREF
  - Select route with shortest AS-PATH
  - Apply MED (if routes learned from same neighbor)

Outline

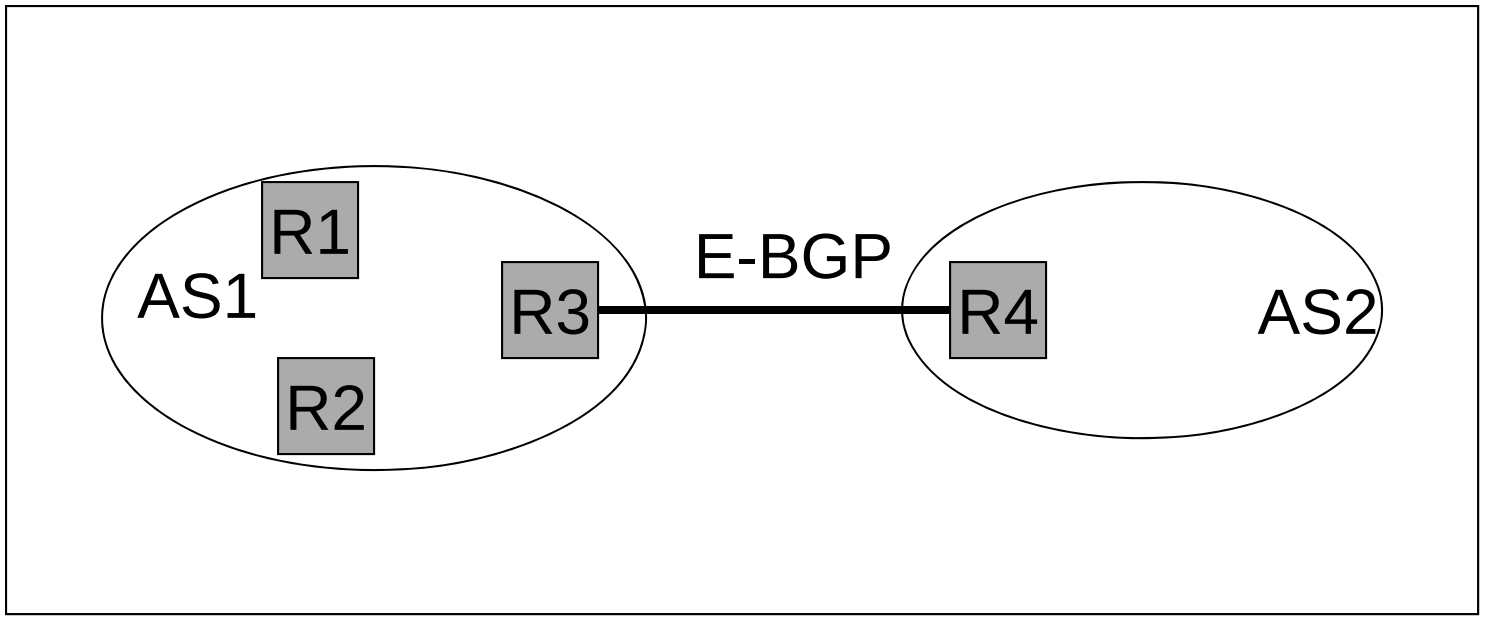


- Routing hierarchy
- Internet structure
- External BGP (E-BGP)
- Internal BGP (I-BGP)

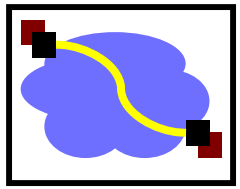
Internal vs. External BGP



- BGP can be used by R3 and R4 to learn routes
- How do R1 and R2 learn routes?

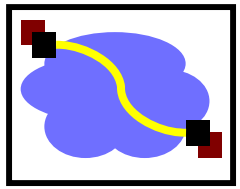


Internal BGP (I-BGP)



- Same messages as E-BGP
- Different rules about re-advertising prefixes:
  - Prefix learned from E-BGP can be advertised to I-BGP neighbor and vice-versa, but
  - Prefix learned from one I-BGP neighbor cannot be advertised to another I-BGP neighbor
  - Reason: no AS PATH within the same AS and thus danger of looping.

Internal BGP (I-BGP)

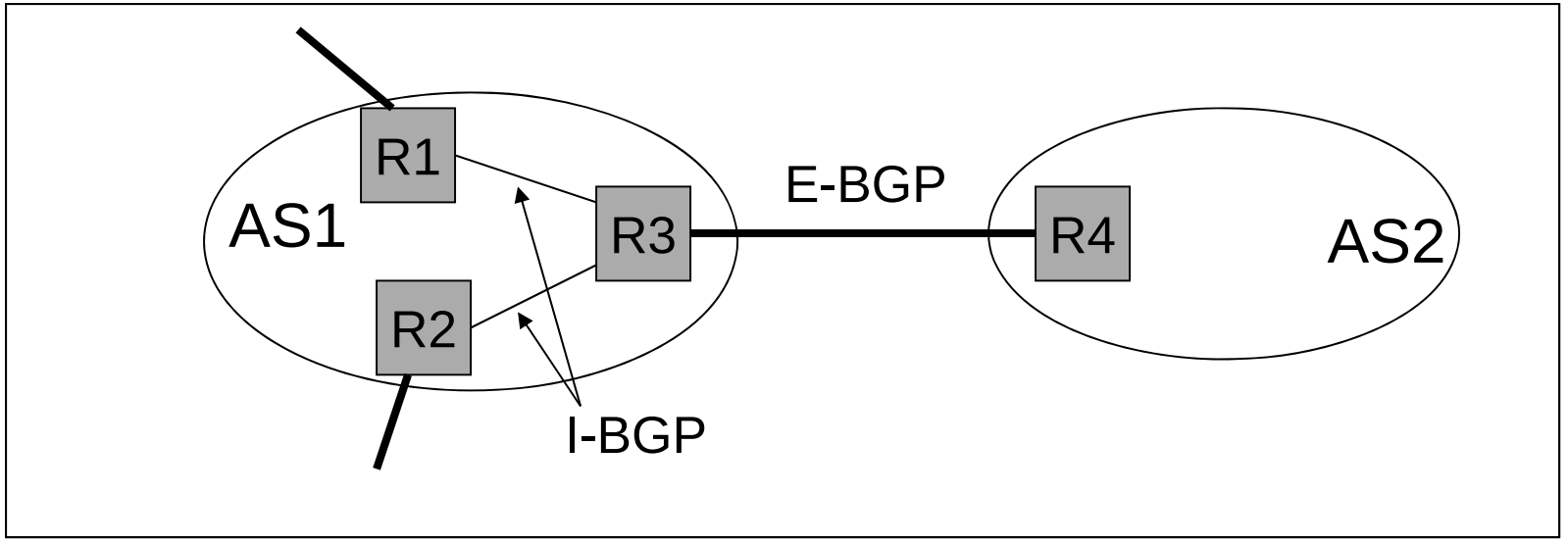


- R3 can tell R1 and R2 prefixes from R4
- R3 can tell R4 prefixes from R1 and R2
- R3 cannot tell R2 prefixes from R1

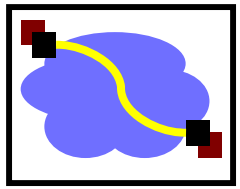
R2 can only find these prefixes through a *direct connection* to R1

Result: I-BGP routers must be fully connected (via TCP)!

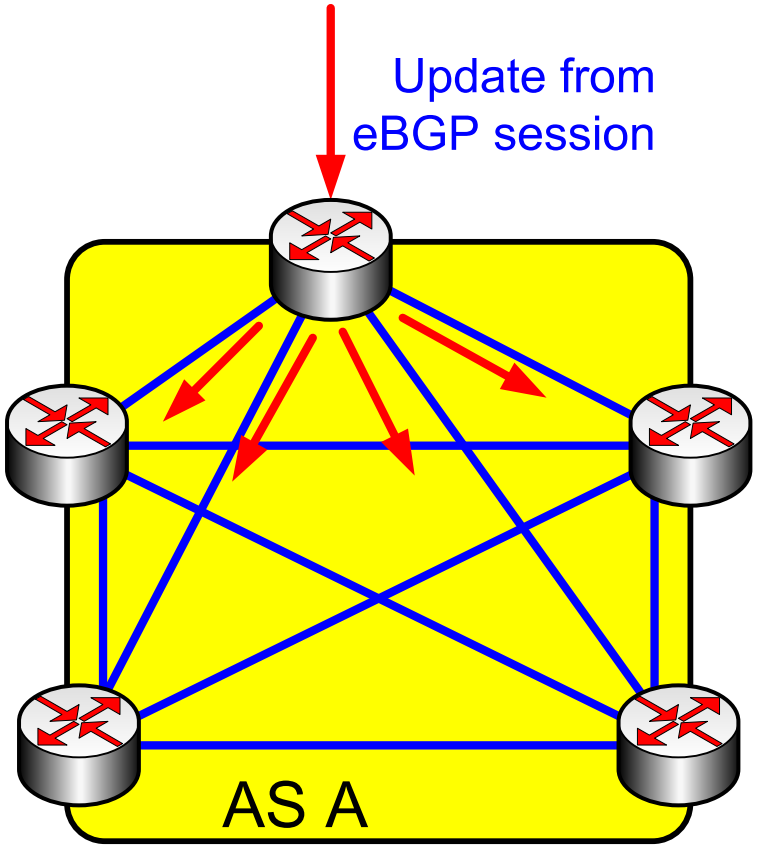
- contrast with E-BGP sessions that map to physical links



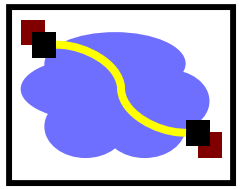
iBGP Sessions



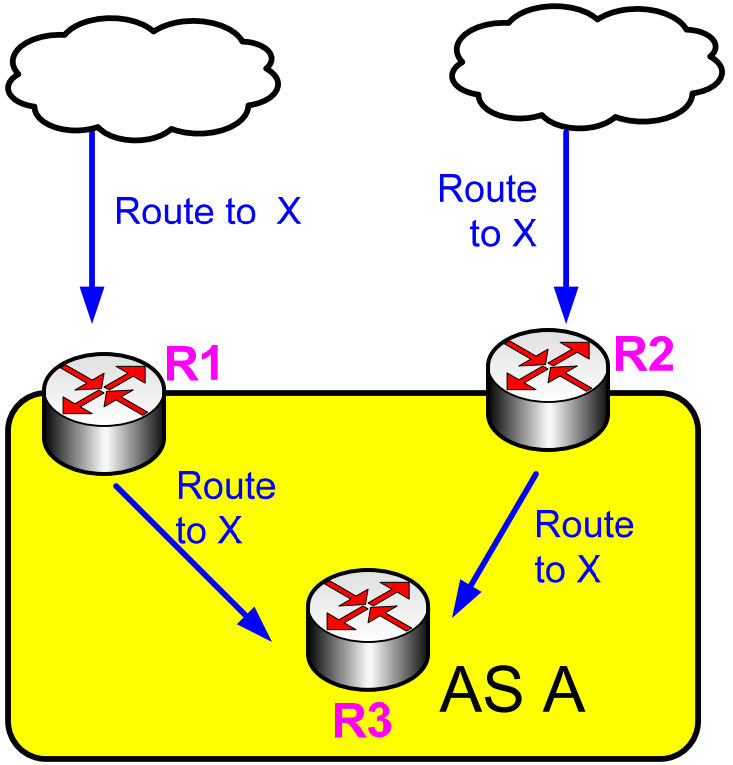
- All iBGP peers in the same autonomous system are fully meshed
- Peer announces routes received via eBGP to iBGP peers
- **But:** iBGP peers do not announce routes received via iBGP to other iBGP peers



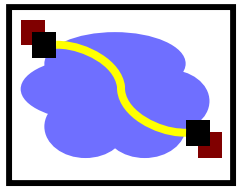
Hot Potato Routing



- Router R3 in autonomous system A receives two advertisements to network X
  - Which route should it pick?
- **Hot Potato Rule:** Select the iBGP peer that has the shortest IGP route
- *Analogy:* Get the packet out of one's own AS as quickly as possible, i.e., on the shortest path



Hot Potato Routing



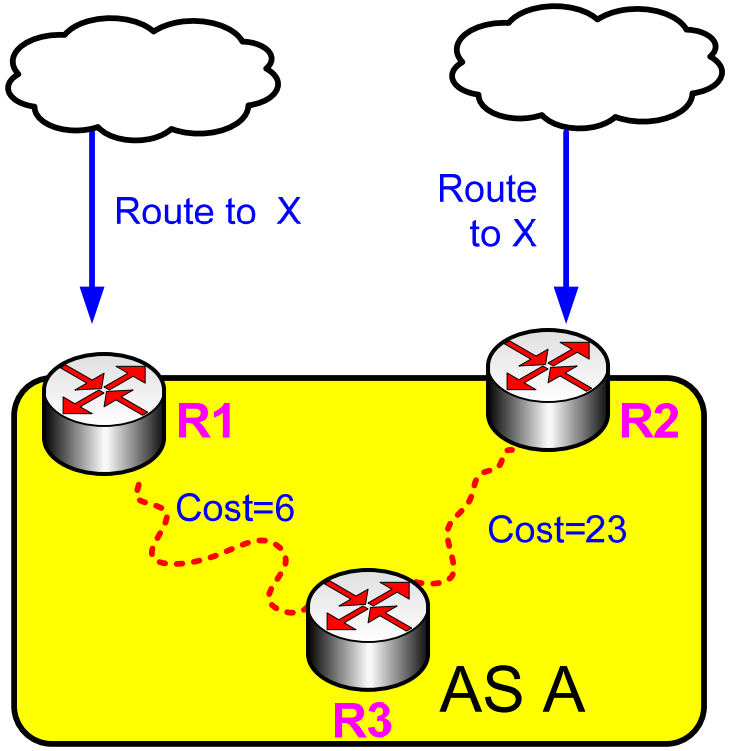
Finding the cheapest IGP route:

- Compare the cost of the two paths
  - R3→ R1
  - R3→ R2

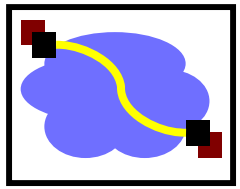
  according to the IGP protocol

- Here: R1 has the shortest path

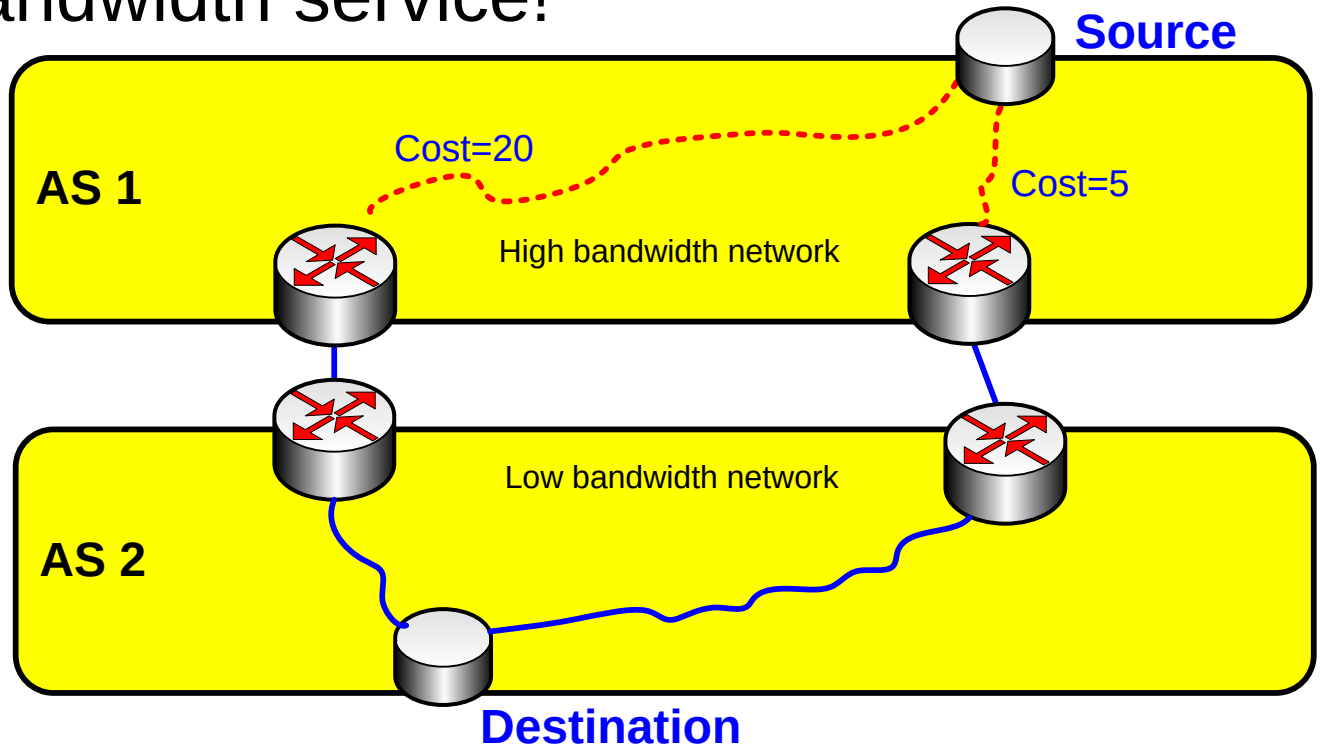
- Add a routing table entry for destination X



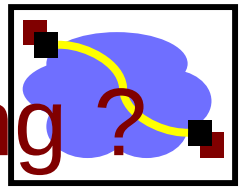
Hot Potato Routing can backfire!



- AS1 would serve its customer (source) better by not picking the shortest route to AS 2
- In fact, customer may have paid for a high-bandwidth service!



Why different Intra-, Inter-AS routing ?



policy:

- inter-AS: admin wants control over how its traffic routed, who routes through its net.
- intra-AS: single admin, so no policy decisions needed

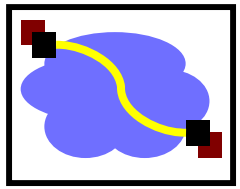
scale:

- hierarchical routing saves table size, reduced update traffic

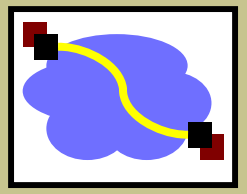
performance:

- intra-AS: can focus on performance
- inter-AS: policy may dominate over performance

Important Concepts



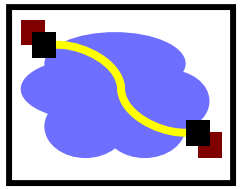
- Wide area Internet structure and routing driven by economic considerations
  - Customer, providers and peers
- BGP designed to:
  - Provide hierarchy that allows scalability
  - Allow enforcement of policies related to structure
- Mechanisms
  - Path vector – scalable, hides structure from neighbors, detects loops quickly
  - IBGP structure/requirements – reuse of BGP, need for a fully connected mesh



EXTRA SLIDES

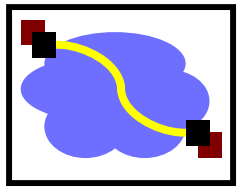
The rest of the slides are FYI

History



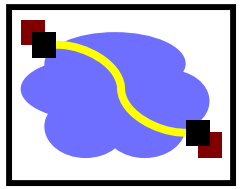
- Mid-80s: EGP
  - Reachability protocol (no shortest path)
  - Did not accommodate cycles (tree topology)
  - Evolved when all networks connected to NSF backbone
- Result: BGP introduced as routing protocol
  - Latest version = BGP 4
  - BGP-4 supports CIDR
  - Primary objective: connectivity not performance

Link Failures

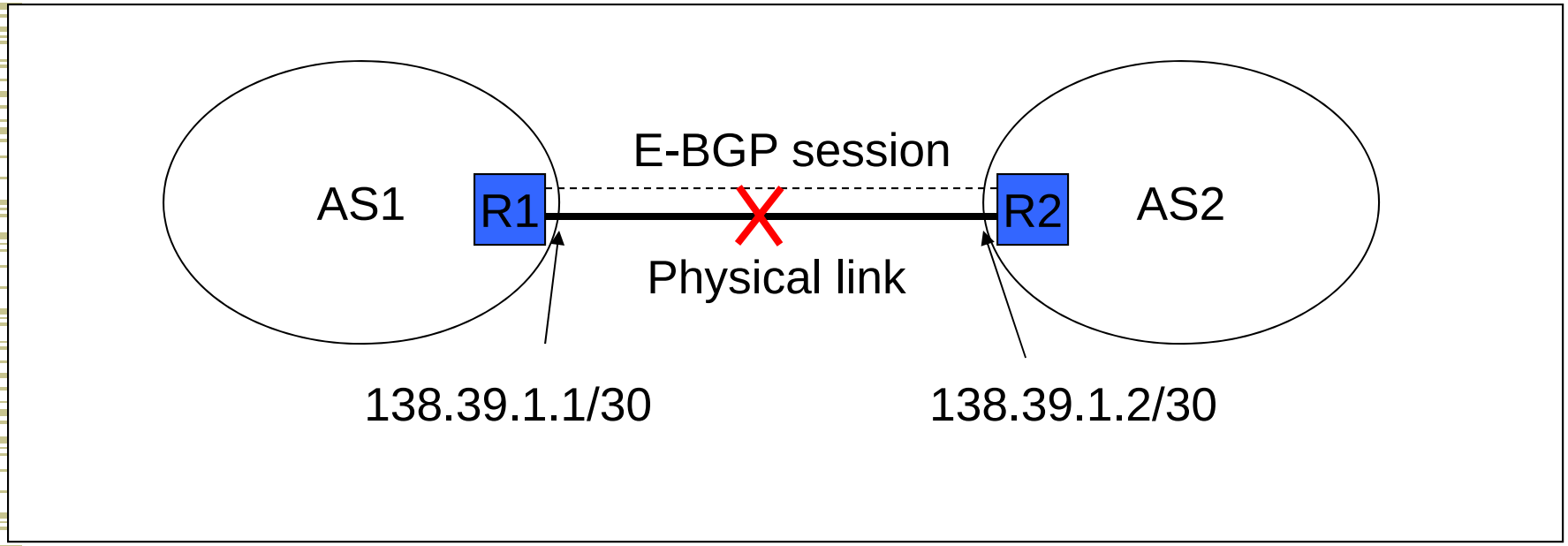


- Two types of link failures:
  - Failure on an E-BGP link
  - Failure on an I-BGP Link
- These failures are treated completely different in BGP
- Why?

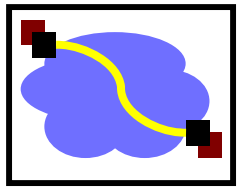
Failure on an E-BGP Link



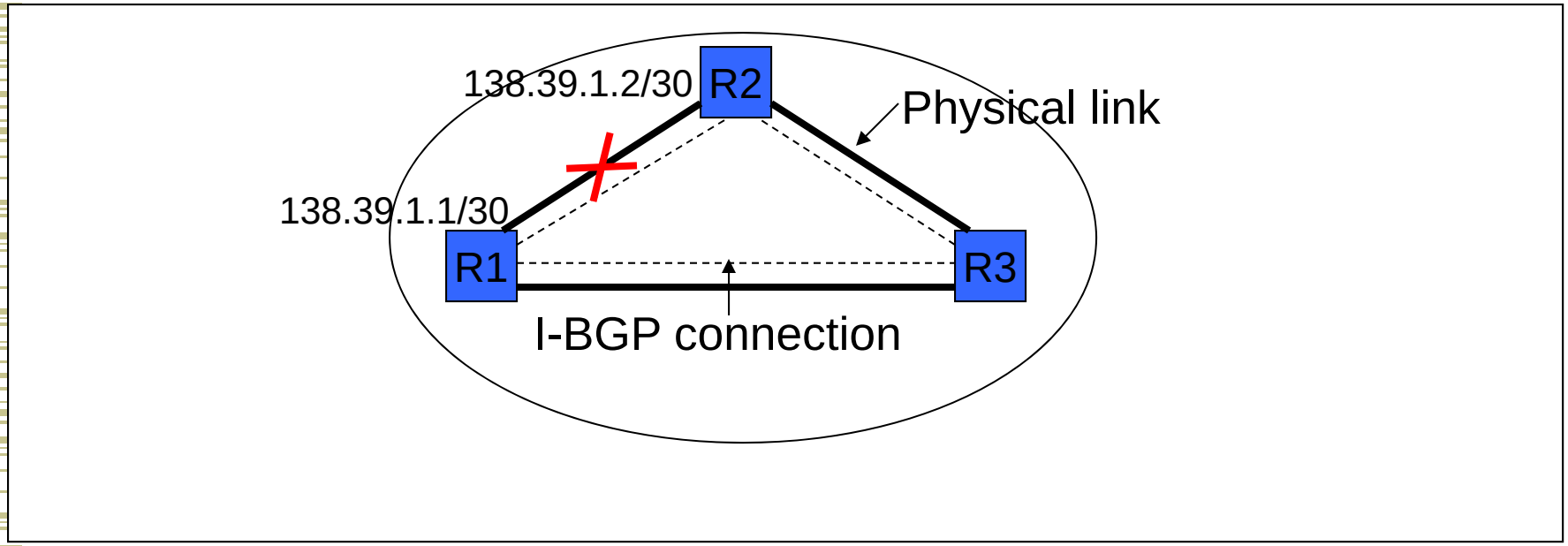
- If the link R1-R2 goes down
  - The TCP connection breaks
  - BGP routes are removed
- This is the *desired* behavior



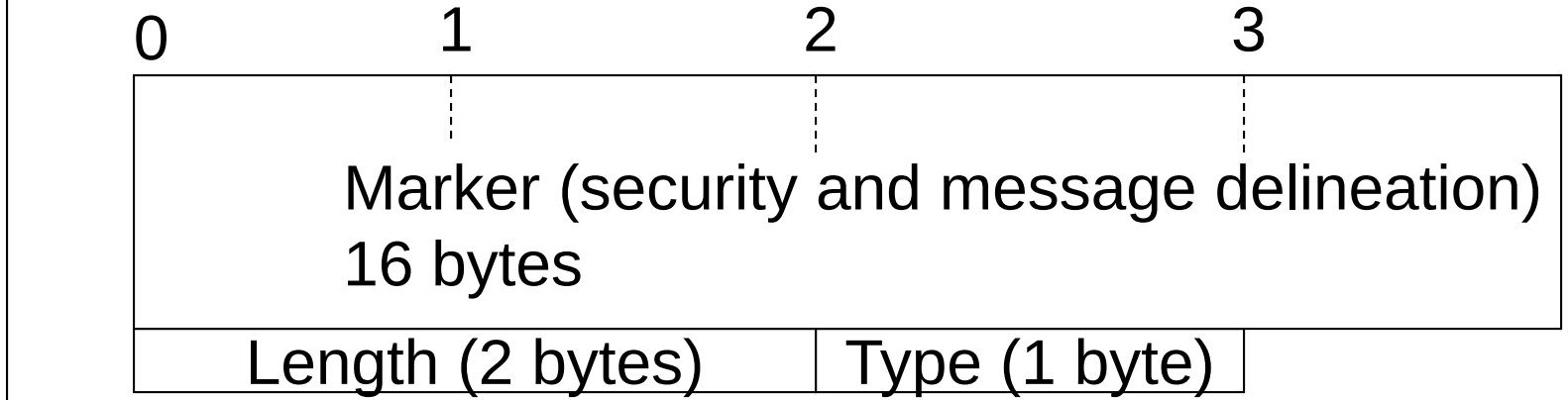
Failure on an I-BGP Link



- If link R1-R2 goes down, R1 and R2 should still be able to exchange traffic
- The indirect path through R3 must be used
- Thus, E-BGP and I-BGP must use *different conventions* with respect to TCP endpoints

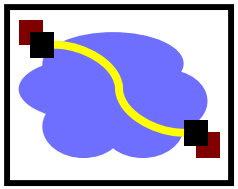


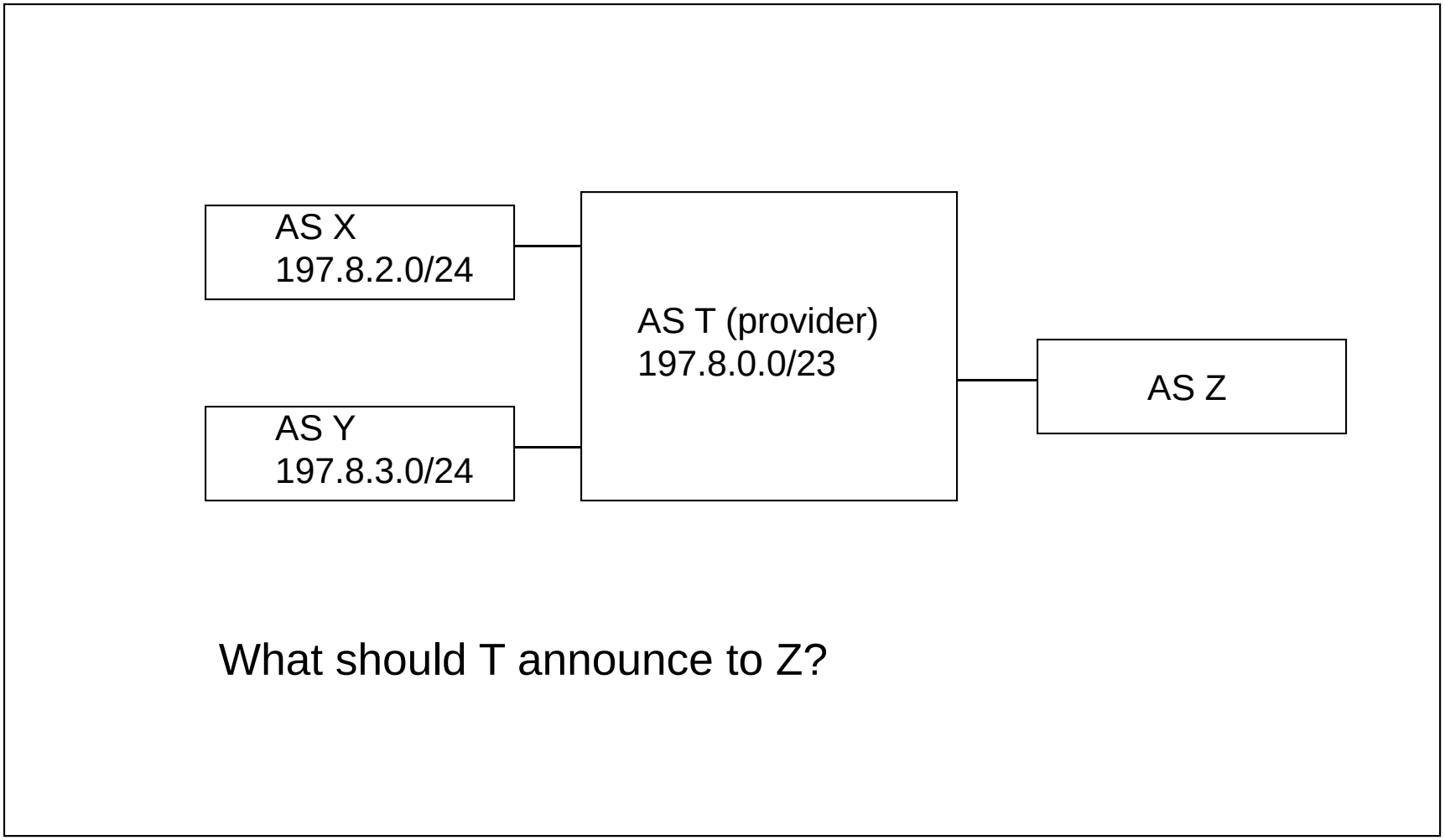
BGP Common Header



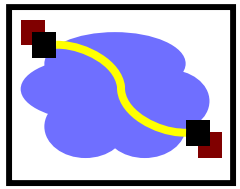
Types: OPEN, UPDATE, NOTIFICATION, KEEPALIVE

CIDR and BGP



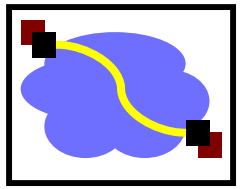

AS X
197.8.2.0/24
AS Y
197.8.3.0/24
AS T (provider)
197.8.0.0/23
AS Z
What should T announce to Z?

Options



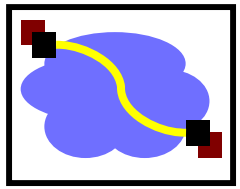
- Advertise all paths:
  - Path 1: through T can reach 197.8.0.0/23
  - Path 2: through T can reach 197.8.2.0/24
  - Path 3: through T can reach 197.8.3.0/24
- But this does not reduce routing tables! We would like to advertise:
  - Path 1: through T can reach 197.8.0.0/22

Sets and Sequences



- Problem: what do we list in the route?
  - List T: omitting information not acceptable, may lead to loops
  - List T, X, Y: misleading, appears as 3-hop path
- Solution: restructure AS Path attribute as:
  - Path: (Sequence (T), Set (X, Y))
  - If Z wants to advertise path:
    - Path: (Sequence (Z, T), Set (X, Y))
  - In practice used only if paths in set have same attributes

Other Attributes



- ORIGIN
  - Source of route (IGP, EGP, other)
- NEXT_HOP
  - Address of next hop router to use
- Check out http://www.cisco.com for full explanation