BITS Pilani
Pilani Campus

Advance Computer Networks (CS G525)

Virendra S Shekhawat
Department of Computer Science and Information Systems

First Semester 2018-2019
Lecture: 21-22 [1-3 Oct 2018]

Agenda

- Internet Routing Basics
  - Inter-domain
  - Intra-domain

- Inter-domain Routing with Border Gateway Protocol (BGP)

Forwarding vs. Routing

- Forwarding: Data plane
  - Directing a data packet to an outgoing link
  - Individual router using a forwarding table

- Routing: Control plane
  - *Computing the paths* the packets will follow
  - Routers *talking* amongst themselves
  - Individual router creating a *forwarding table*

Routing Sub Functions

- Topology Update: Characterize and maintain connectivity
  - Discover neighbors, Measure “distance” (one or more metric), Disseminate
- Route Computation:
  - Kind of path: Multicast, Unicast
  - Centralized or Distributed Algorithm
  - Policy
  - Hierarchy
- Switching: Forward the packets at each node

Datagram v/s Virtual Circuit

- Datagram Routing
  - Each packet to be forwarded independently
- Virtual Circuit
  - Each packet from same s-d uses same route
  - More state (pick the “right” granularity)
- QoS sensitive networks use VC’s and signaling
  - Find a route with resources available for the connection
  - “Reserve” the resources before sending data packets

Internet Routing Protocols

	Link State	Distance Vector	Path Vector
Information Dissem-ination	Flood link state advertisements to all routers	Update distances from neighbors' distances	Update paths based on neighbors' paths
Algorithm	Dijsktra's shortest path	Bellman-Ford shortest path	Local policy to rank paths
Converge	Fast due to flooding	Slow, due to count-to-infinity	Slow, due to path exploration
Protocols	OSPF, IS-IS	RIP, EIGRP	BGP

Link State vs. Distance Vector

- Disadvantages of LS
  - Need consistent computation of shortest paths
    - Same view of topology
    - Same metric in computing routes
  - Slightly more complicated protocol [rfc 2328 of 244 pages]
- Advantages of LS
  - Faster convergence
  - Global information allows optimal route computation
  - Gives unified global view
    - Useful for other purposes, e.g., building MPLS tables
- Question: Can link state have forwarding loops?

Link State Variant: Source Routing

- Algorithm:
  - Broadcast the entire topology to everyone
  - Forwarding at source:
    - Compute shortest path (Dijkstra's algorithm)
    - Put path in packet header
  - Forwarding at source and remaining hops:
    - Follow path specified by source
    - Router looks up next hop in packet header, strips it off and forwards remaining packet
    - Used in Adhoc networks (e.g. DSR protocol)
- Question: Can this result in forwarding loops?

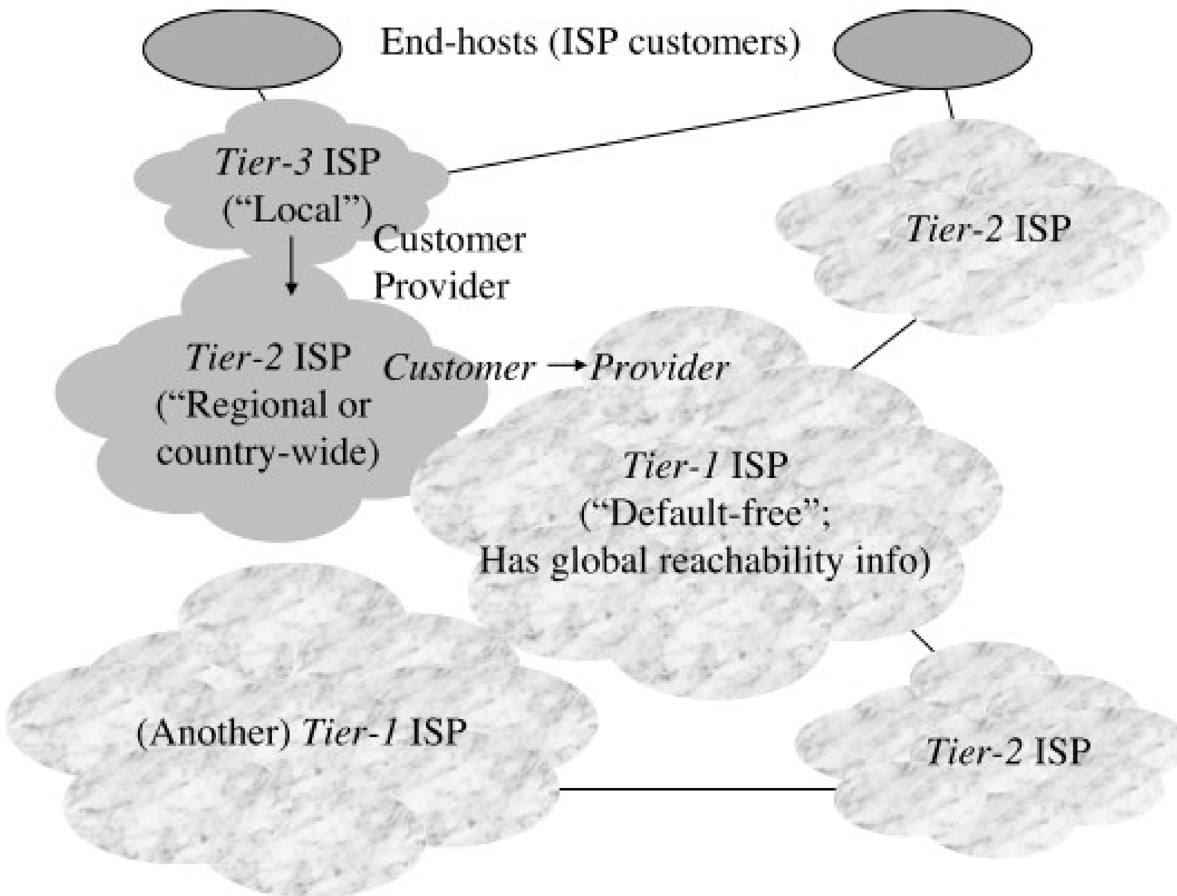
Internet Routing System: Two Tier

- Interdomain routing: Between ASes
  - Routing policies based on *business relationships*
  - No common metrics, and limited cooperation
  - BGP: policy-based, path-vector routing protocol
- Intradomain routing: Within an AS
  - Shortest-path routing based on *link metrics*
  - Routers are managed by a single institution
  - OSPF and IS-IS: link-state routing protocol
  - RIP and EIGRP: distance-vector routing protocol

Next…

- BGP
  - ASes, Policies
  - BGP Attributes
  - BGP Path Selection
  - I-BGP vs. E-GBP
- Reference
  - BGP Routing Policies in ISP Networks
    - By Matthew Caesar, University of California at Berkeley & Jennifer Rexford, Princeton University

The BIG Picture



Autonomous Systems (ASes)

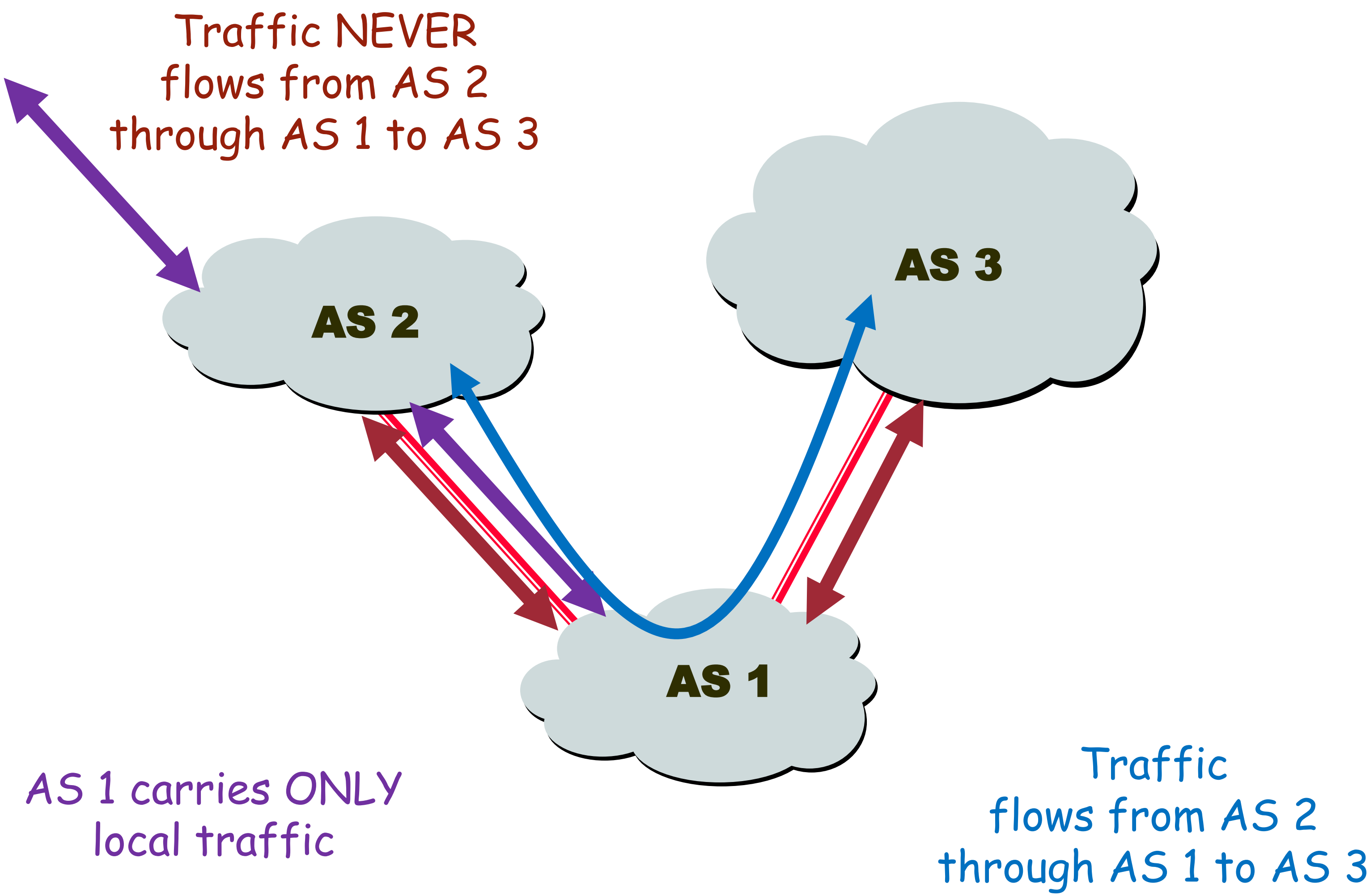
- Autonomous system
  - AS is an actual entity that participates in routing
  - Has an unique 16 bit ASN (now 32 bit [RFC 4893 @ 2007]) assigned to it and typically participates in inter-domain routing

- Examples:
  - MIT: 3, CMU: 9
  - AT&T: 7018, 6341, 5074, …
  - UUNET: 701, 702, 284, 12199, …
  - Sprint: 1239, 1240, 6211, 6242, …

Let's Find out…

- How do ASes interconnect to provide global connectivity?
- How does routing information get exchanged?

AS Categories [Stub/Multi-homed/Transit]



Peering Relationship

- Customer
- Provider
- Peer
- Sibling

Inter-domain Routing in the Internet

- Link state or distance vector?

- 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 route, check if AS already in path
    - If yes, reject route
    - If no, add self and (possibly) advertise route further

BGP-4

- BGP = Border Gateway Protocol
- Is a Policy-Based routing protocol
- It is the EGP of today's global Internet
- Relatively simple protocol, but configuration is complex

1989 : BGP-1 [RFC 1105]

- **Replacement for EGP (1984, RFC 904)**

1990 : BGP-2 [RFC 1163]

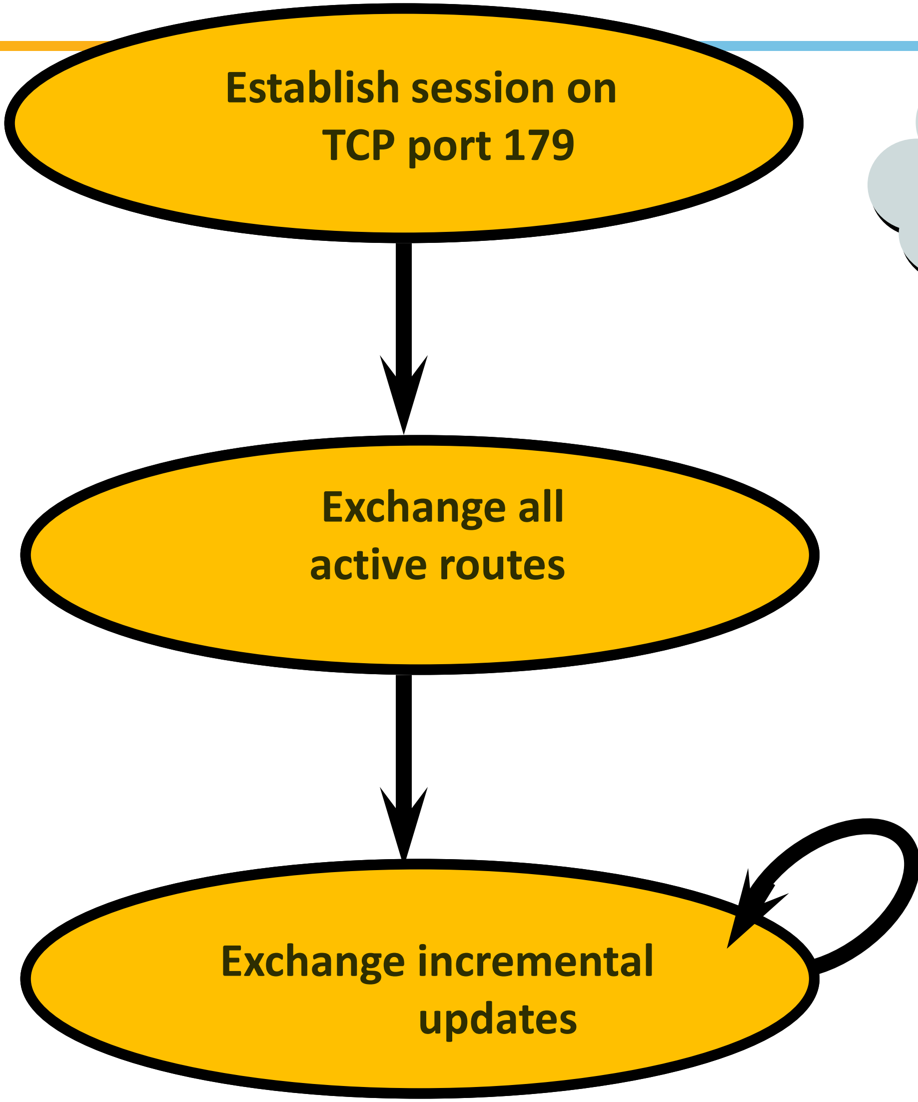
1991 : BGP-3 [RFC 1267]

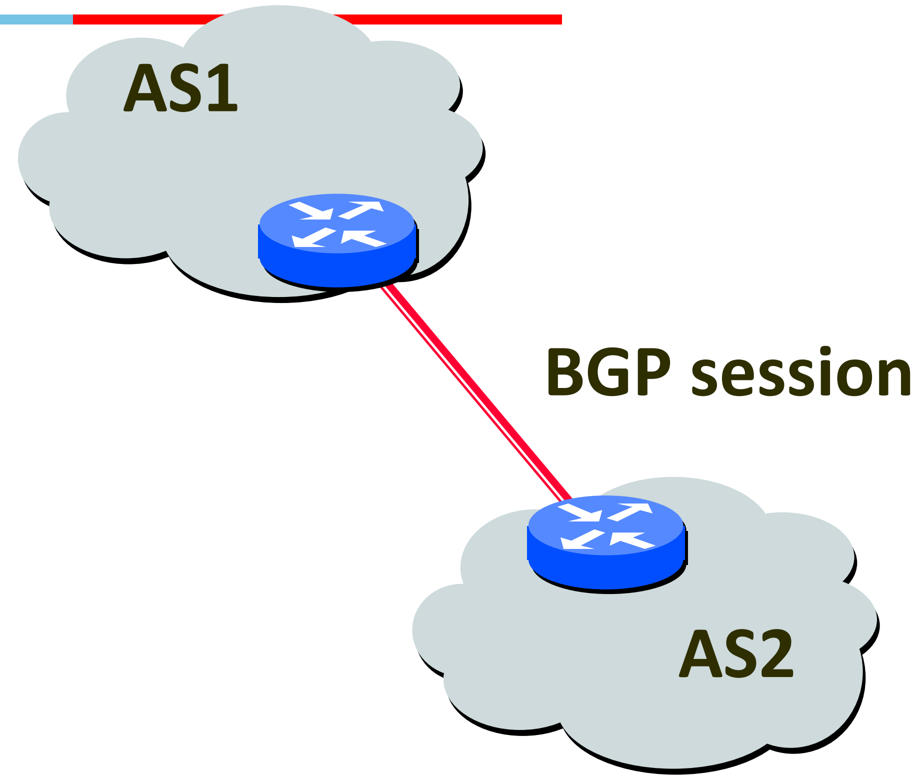
1995 : BGP-4 [RFC 1771]

2006: BGP-4 [RFC 4271]

- **Support for Classless Interdomain Routing (CIDR) , Route Aggregation**

BGP Operations





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 + <u>attributes values</u>

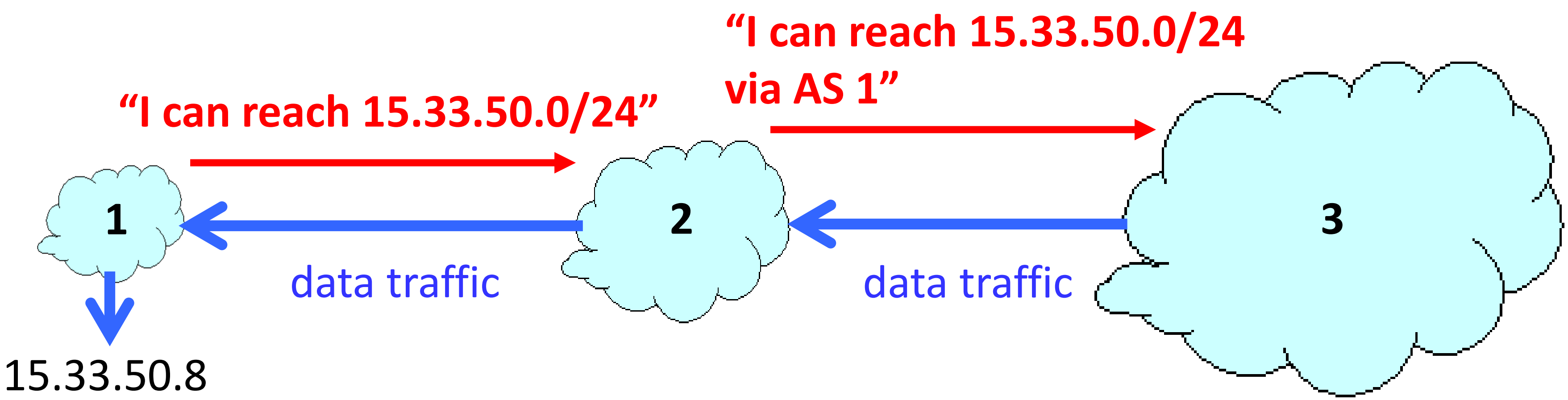
Fundamental Rules: BGP

- BGP advertises to neighbors only those routes that it uses
  – Consistent with the hop-by-hop Internet paradigm
- No need for periodic refresh - routes are valid until withdrawn, or the connection is lost
- Incremental updates are possible

Policy Decisions

- BGP provides capability for enforcing various policies
- BGP enforces policies by **choosing paths from multiple alternatives** and controlling advertisement to other AS's
- Import policy
  - What to do with routes learned from neighbors?
- Export policy
  - What routes to announce to neighbors?
  - Depends on relationship with neighbors

BGP Route Export Example



BGP Policy Taxonomy

- Business Relationship
  – Governs economic and political relationship
- Traffic Engineering
  – Controls traffic flow within ISP and across peering points and maintain QoS
- Scalability
  – Reduce control traffic and avoid overloading routes
- Security
  – Protect ISP against malicious or accidental attacks

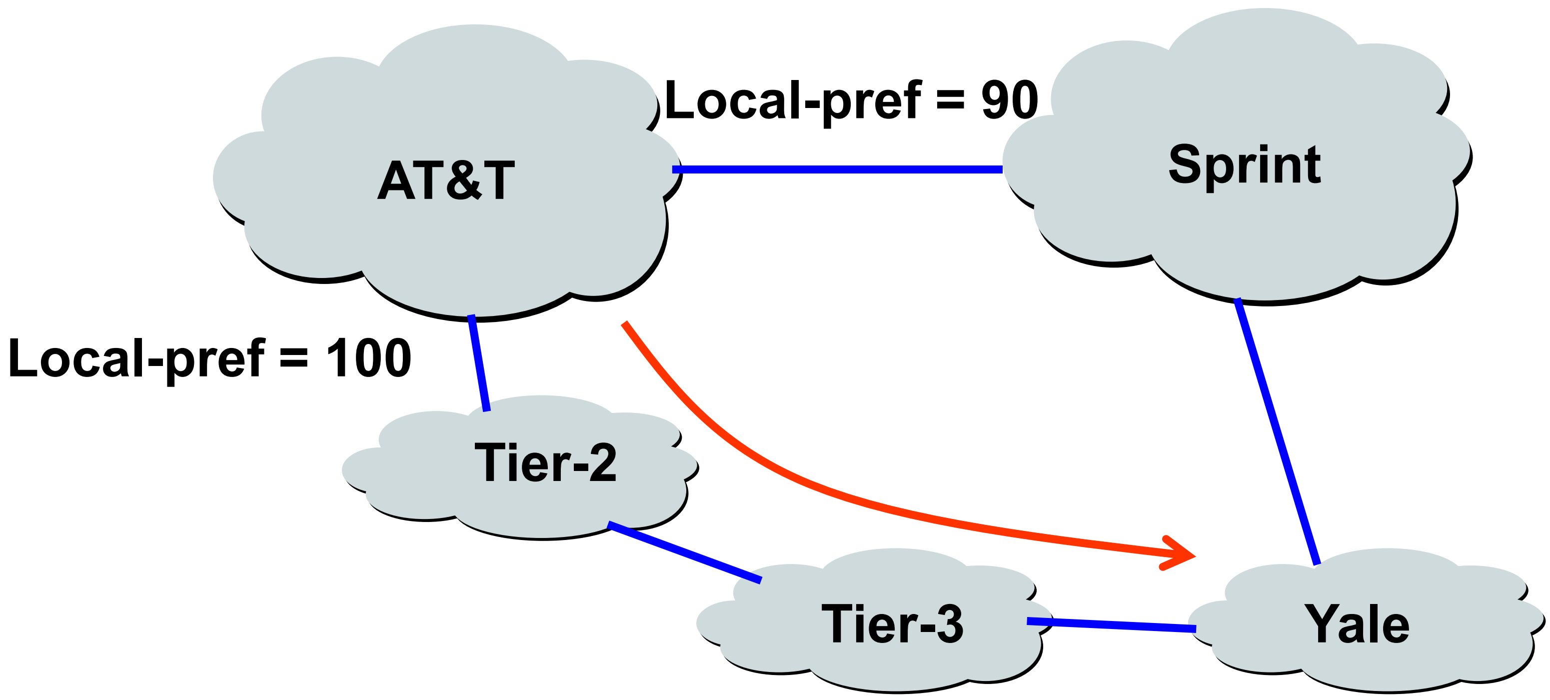
Example: Export Policy

- Once the route is announced the AS is willing to transit traffic on that route
- To Customers
  - Announce all routes learned from **peers**, **providers** and **customers**, and **self-origin routes**
- To Providers
  - Announce routes learned from **customers** and **self-origin routes**
- To Peers
  - Announce routes learned from **customers** and **self-origin routes**

How to implement export policies?

- BGP Attributes
  - Local Preference
  - AS-Path Length
  - MED (Multi Exit Discriminator)
  - Next hop

Example: Local Preference

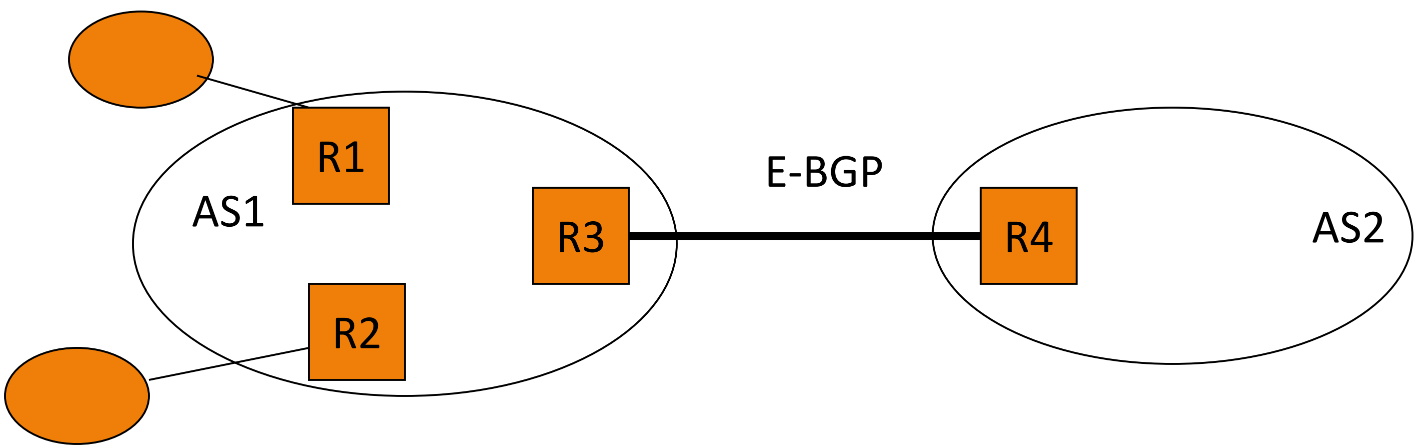


Local Pref. Vs. MED

- Use Local Pref., if you have multiple exit points to a neighbor and want to *tell your routers* where to direct traffic
  - Intra-AS policy
- Use MED, if you have multiple links with a neighbor and want to *tell your neighbor* where to send traffic to you
  - Inter-AS policy

Internal BGP vs. External BGP

- R3 and R4 can learn routes by using BGP
- How do R1 and R2 learn routes?
- Option 1: Inject routes in IGP
  - Only works for small routing tables
- Option 2: Use I-BGP



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.

Route Selection Process

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**

Traffic Engineering Goals

- Load balancing
  - Making good use of network resources
  - Alleviating network congestion
- End-to-end performance
  - Avoiding paths with downstream congestion
  - By moving traffic to alternate paths
- Mechanisms
  - Preferring some paths over other paths
  - E.g., by setting local-preference attribute
  - Among routes within the same business class

Next Class…

- Fundamental Problems with BGP

Agenda

- Fundamental Problems with BGP
- Reference
  - Some Foundational Problems in Inter-domain Routing [Nick Feamster 2004]

- BGP Mis-configurations
- Reference
  - Understanding BGP Mis-configuration [Ratul Mahajan, 2002]

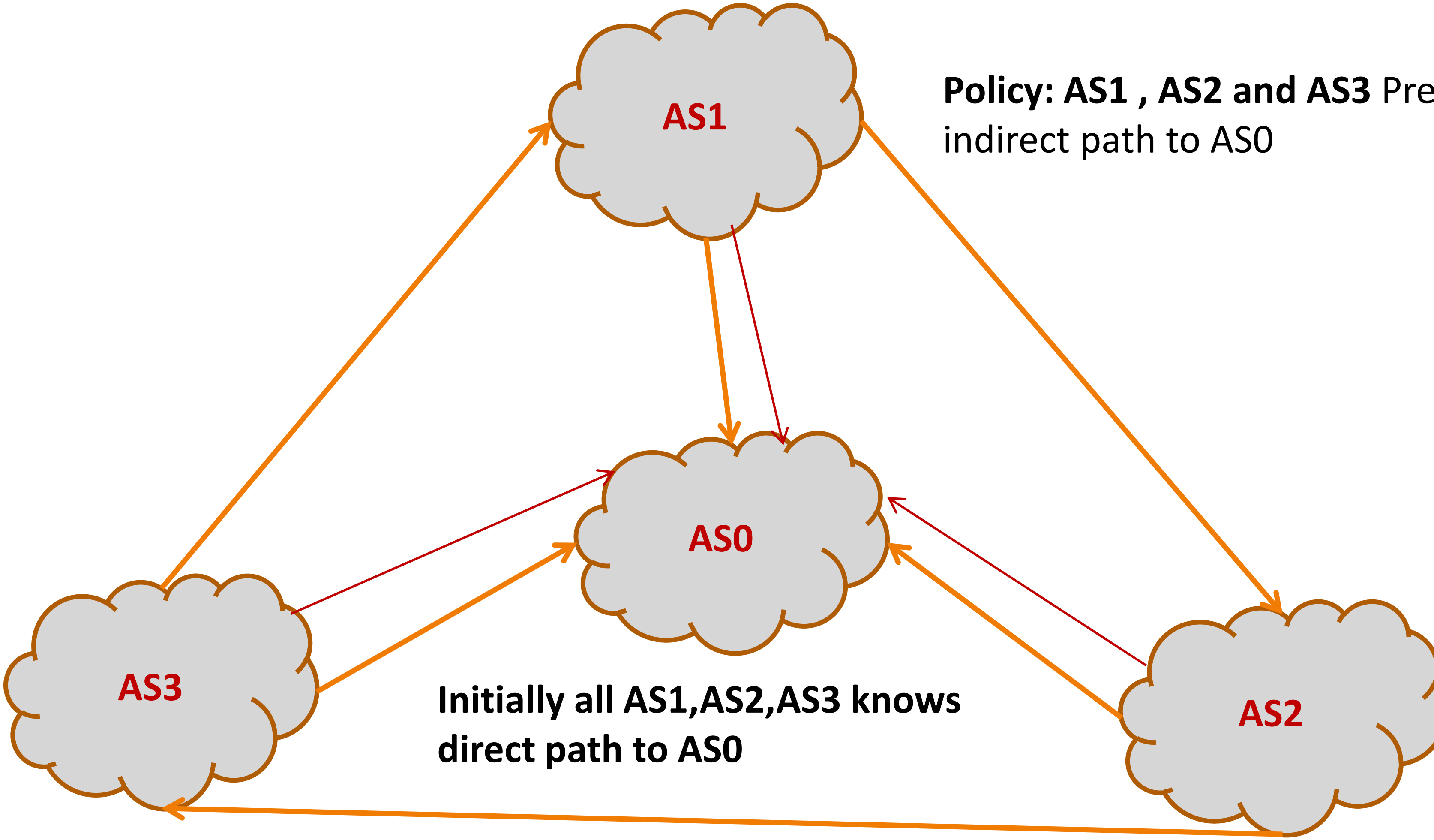
Fundamental Problems with BGP

- Protocol Oscillations
  - Policy Disputes
  - Non monotonic ranking
- Weak Security
  - Control Plane Security
  - Data plane security
- Scalability Induced problems
  - Prefix aggregation

Oscillations: Policy Disputes

- Routing policies are used to implement traffic agreements between peering AS
  - BGP allows each AS to define its own routing policies independently
  - No global coordination exists for configuring the AS routing policies
    - This can lead to protocol oscillations i.e. non-stable routing state

Example: Policy Dispute Oscillations

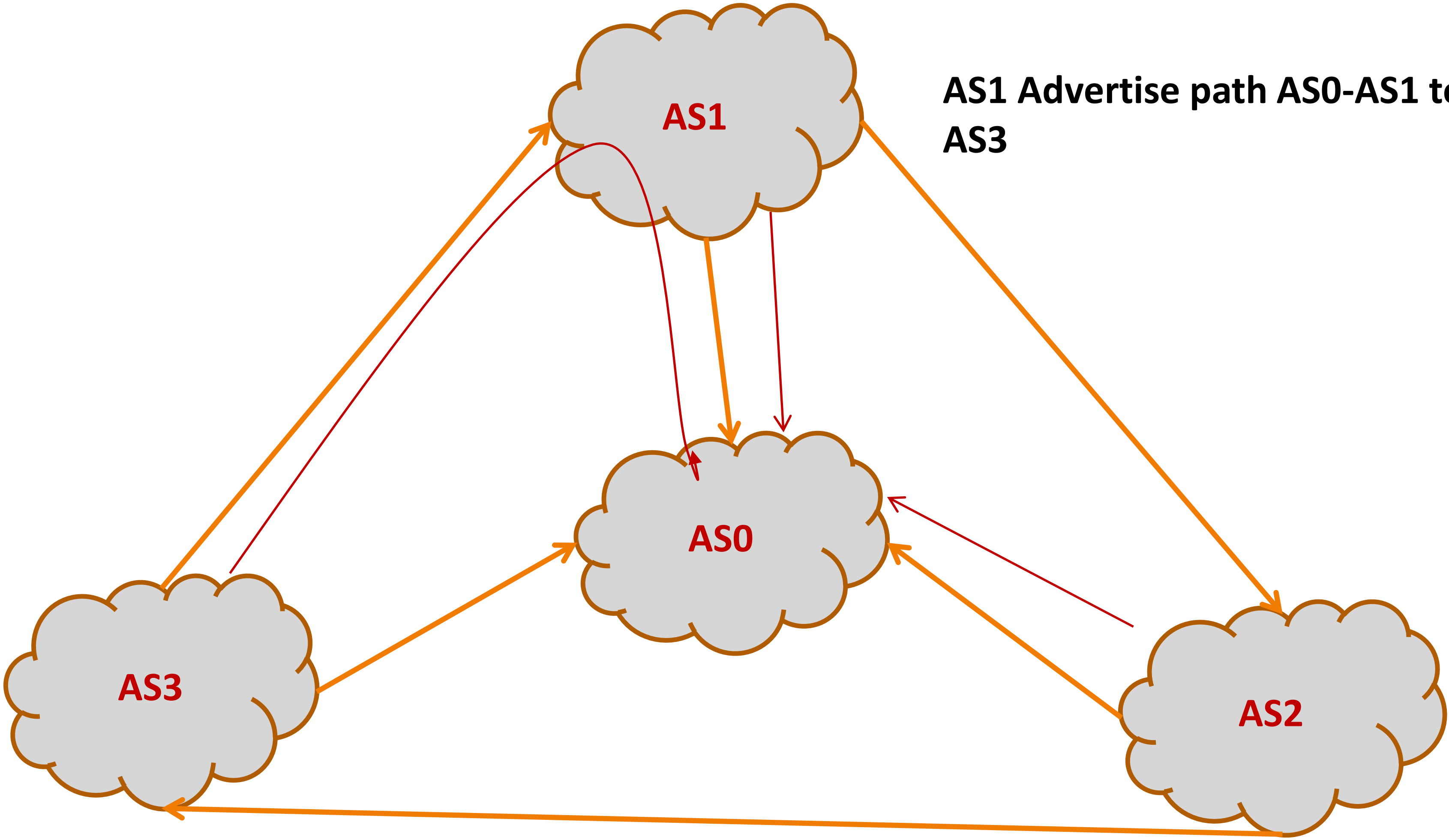


Policy: AS1 , AS2 and AS3 Prefers indirect path to AS0

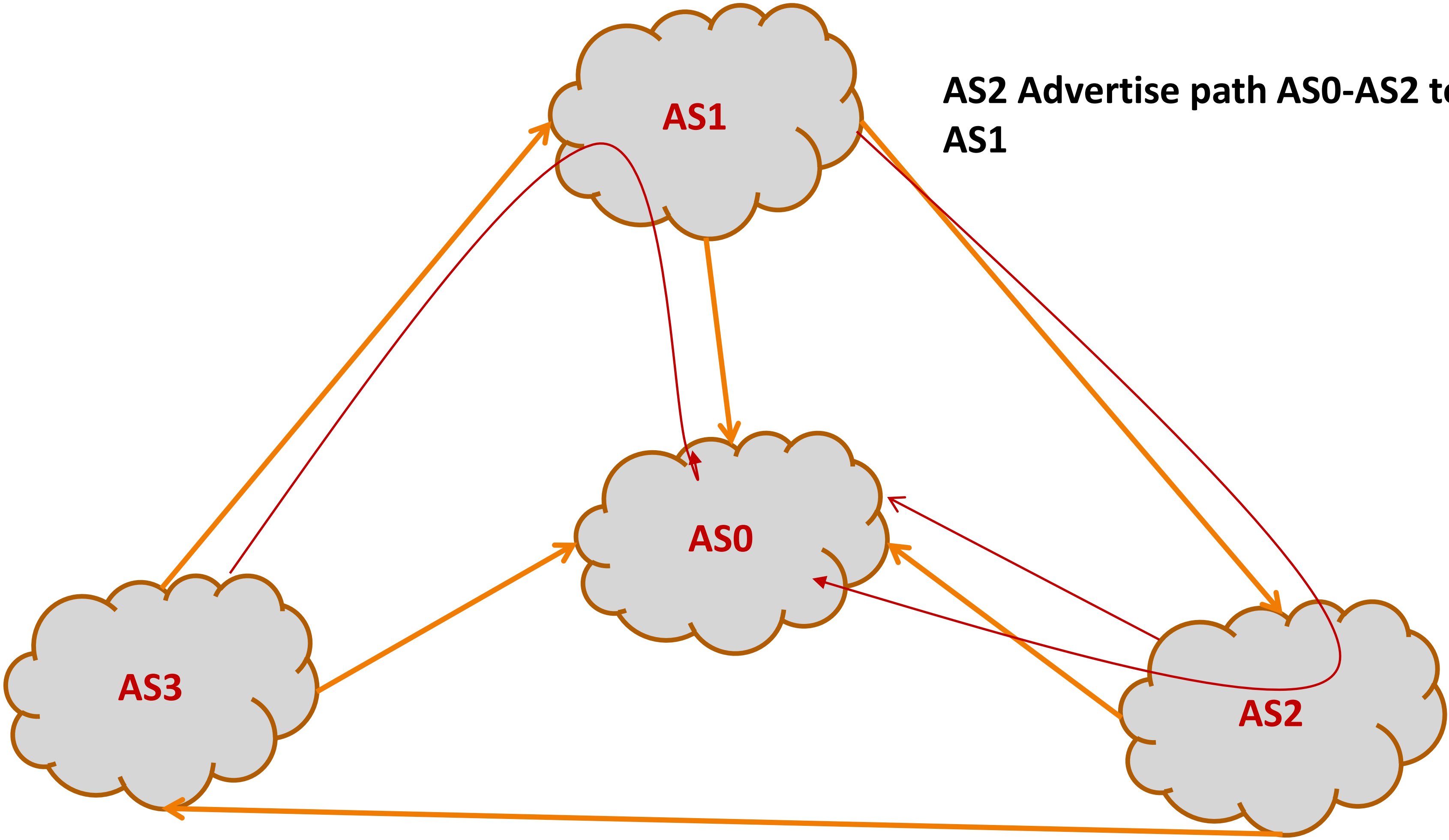
Initially all AS1,AS2,AS3 knows direct path to AS0

Policy Dispute Oscillations

AS1 Advertise path AS0-AS1 to AS3

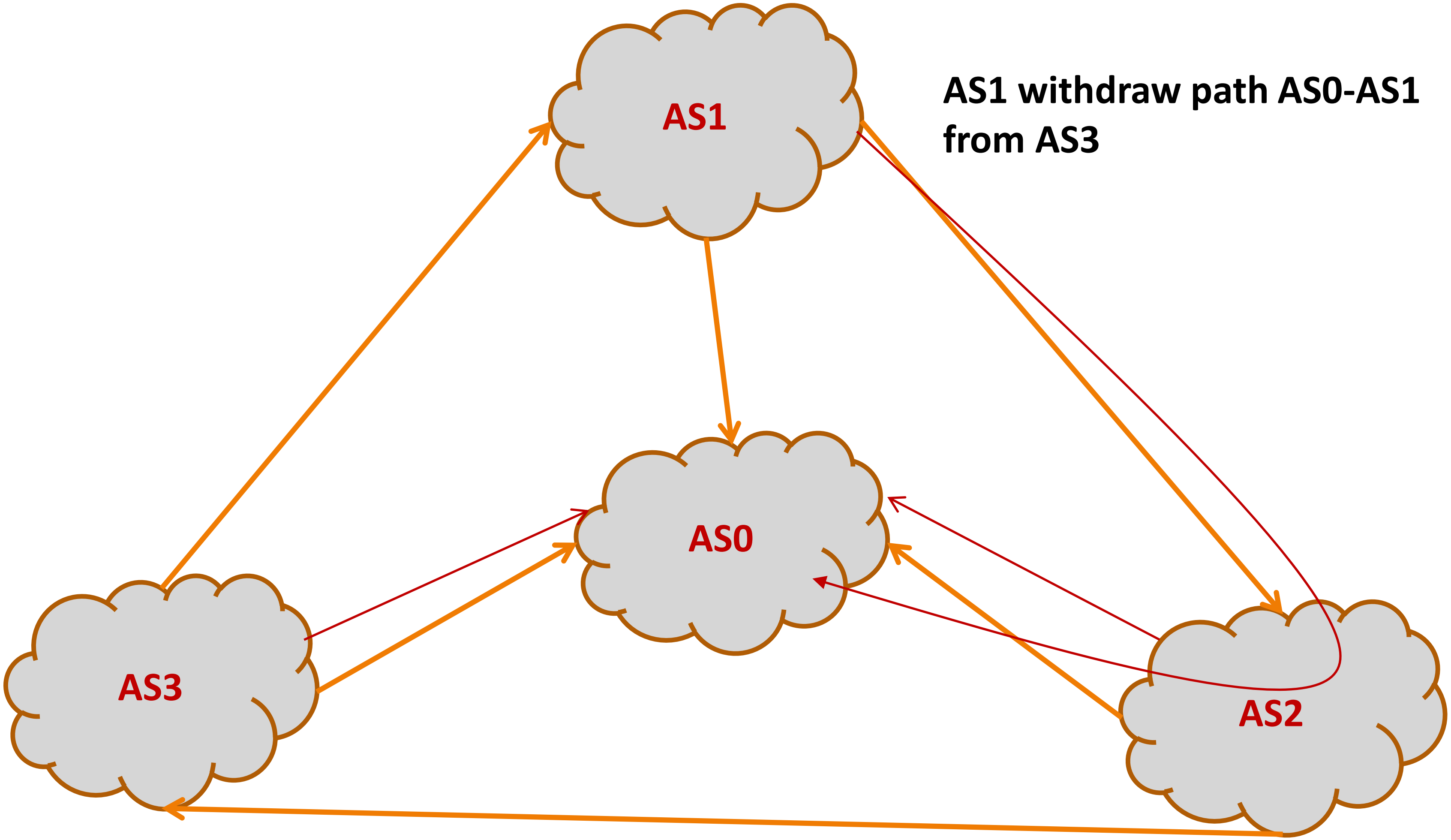


Policy Dispute Oscillations



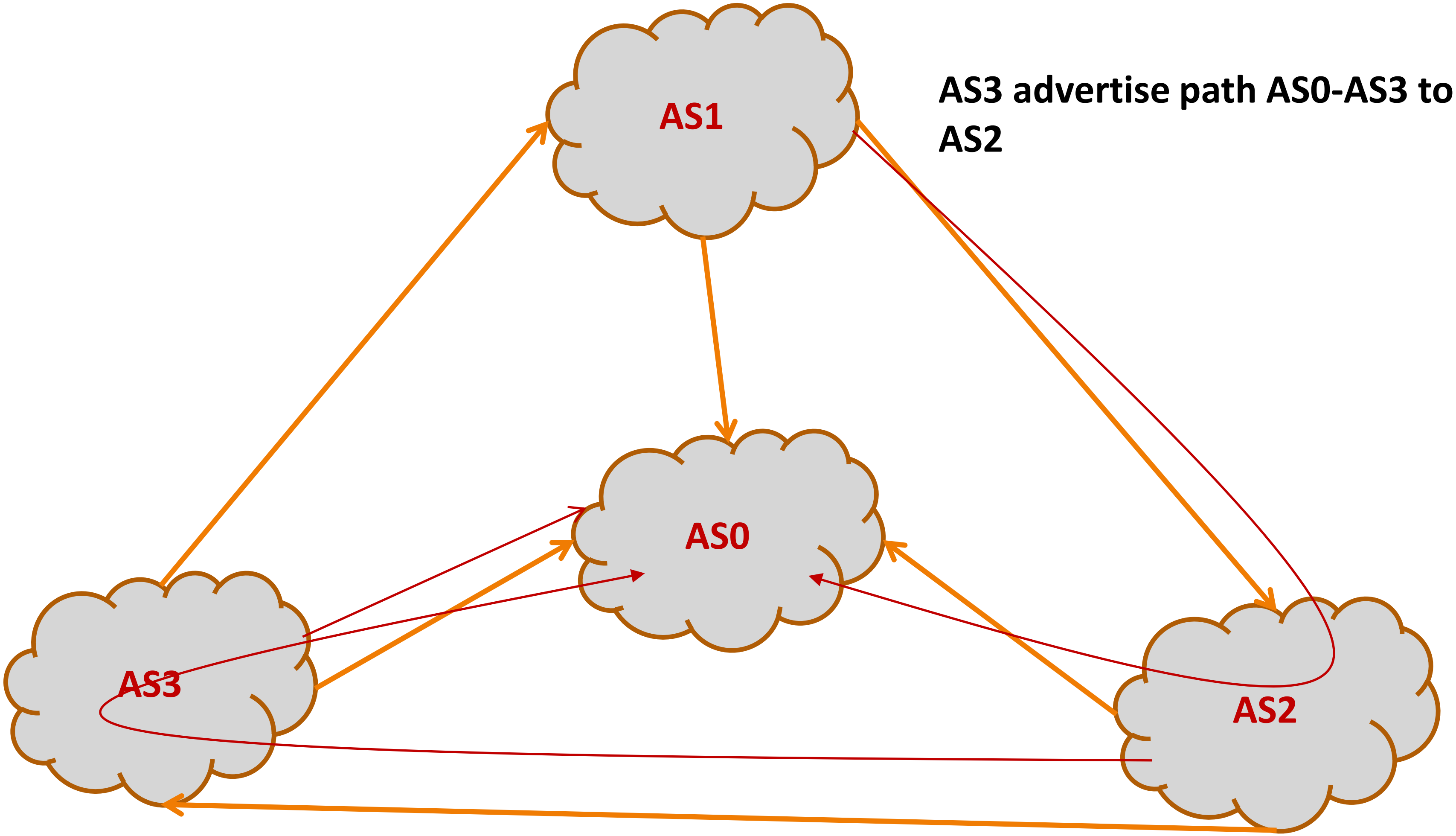
AS2 Advertise path AS0-AS2 to AS1

Policy Dispute Oscillations



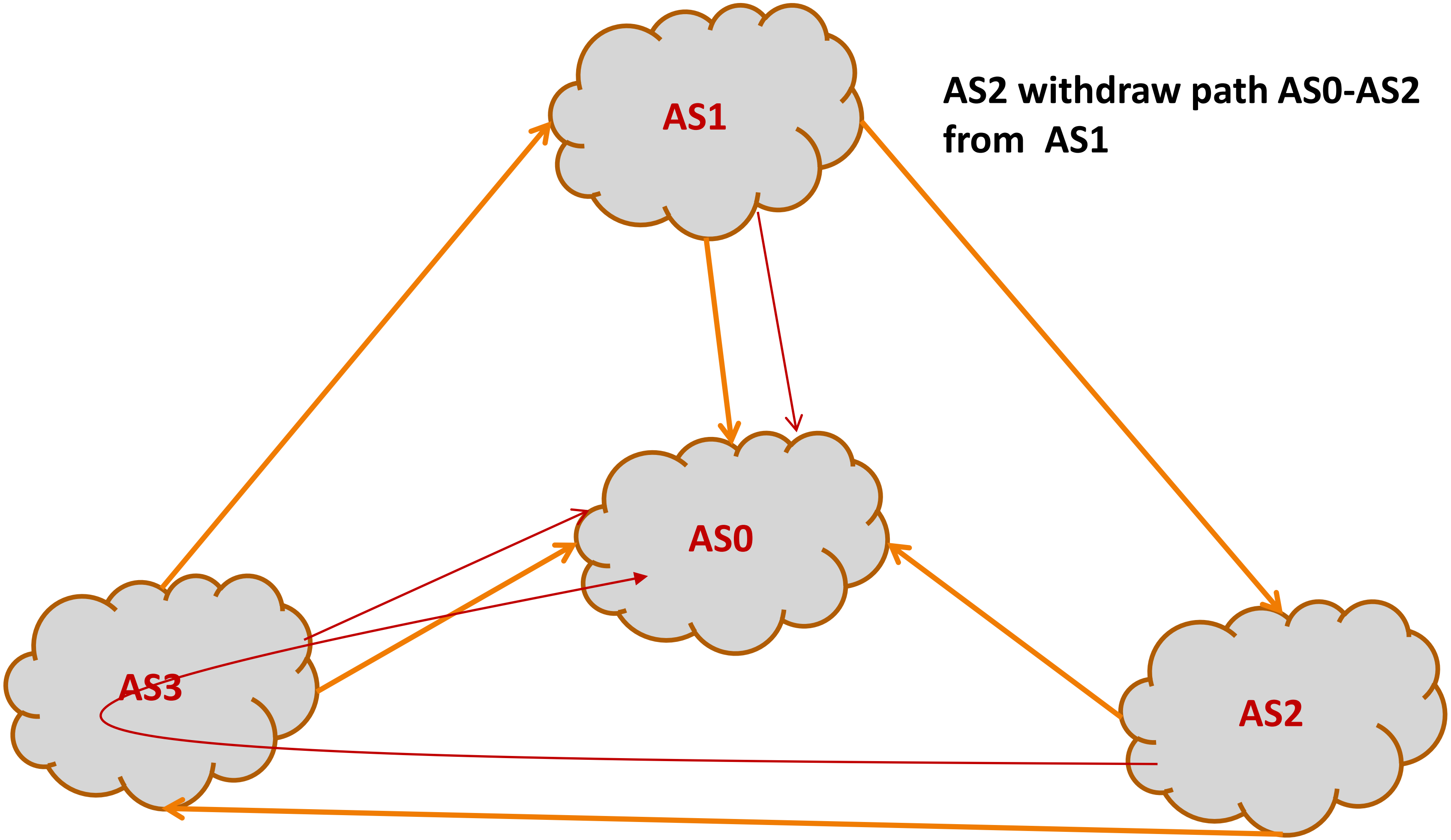
AS1 withdraw path AS0-AS1 from AS3

Policy Dispute Oscillations



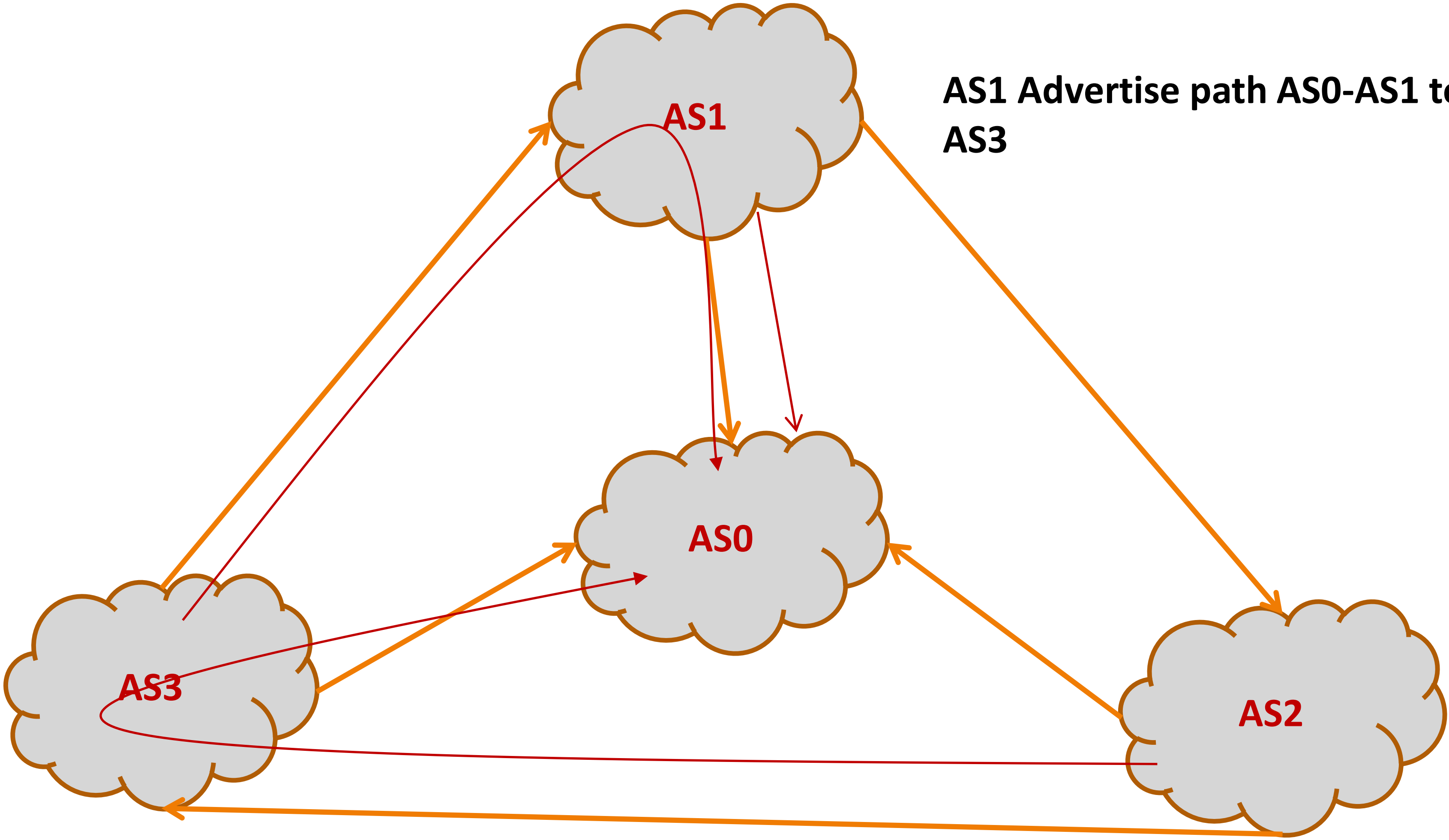
AS3 advertise path AS0-AS3 to AS2

Policy Dispute Oscillations



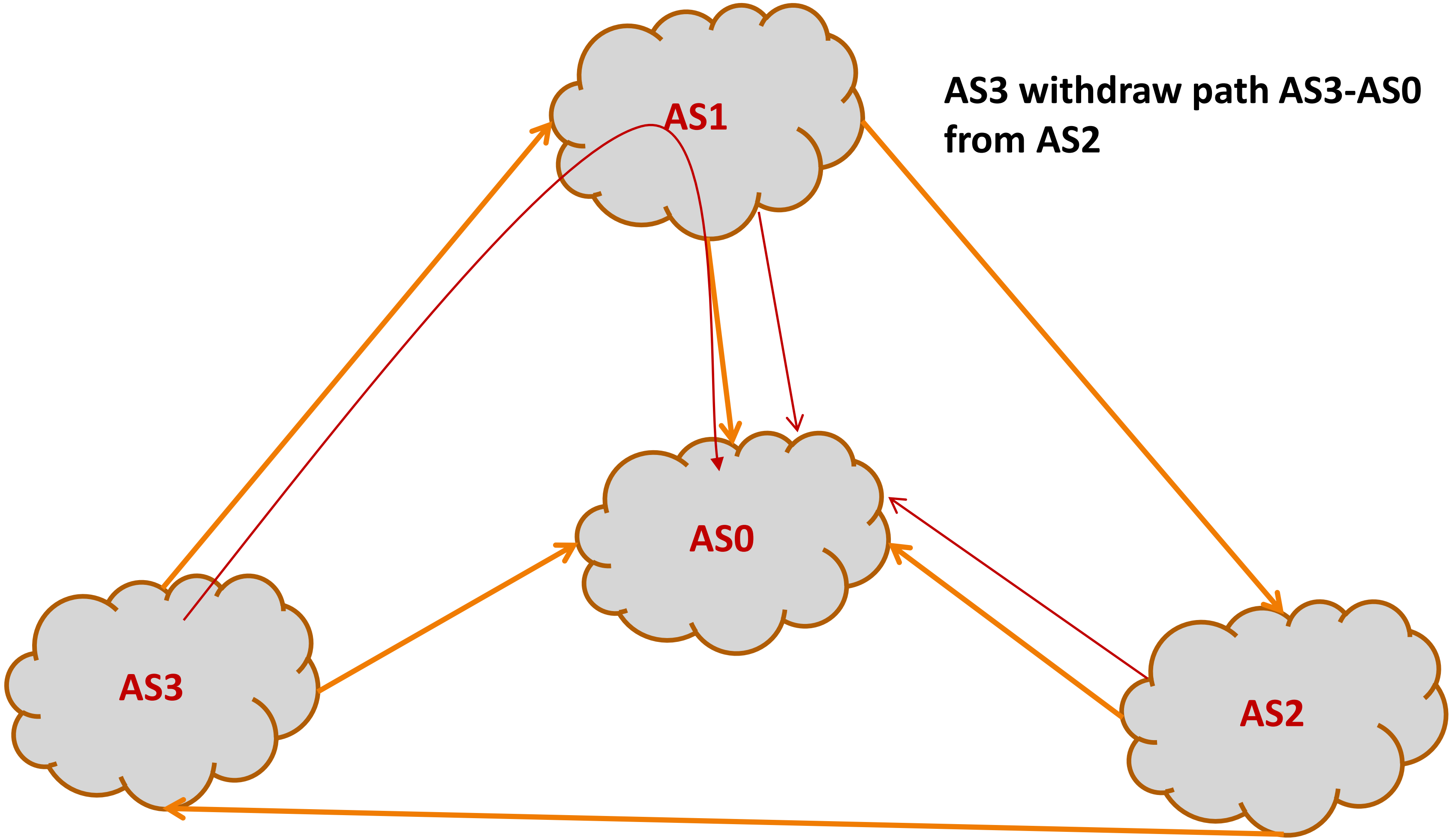
AS2 withdraw path AS0-AS2 from AS1

Policy Dispute Oscillations



AS1 Advertise path AS0-AS1 to AS3

Policy Dispute Oscillations



AS3 withdraw path AS3-AS0 from AS2

Why “Policy disputes” Occur…?

- There is no possible path assignment for which at-least one AS in the system does not have a better path available
  - Thus that AS would switch to the better route
  - This act of switching creates a different path assignment that is also not stable
  - It is analogous to a game where there is no pure strategy (i.e. Nash Equilibrium)

Can Inter-domain routing converge…?

- Argument made by Gao & Rexford
  - If every AS obeys certain local constraints on preference and export policies, then BGP is guaranteed to converge !
- Validity of the argument
  - There may be legitimate reasons to deviate from the guidelines
    - e.g. AS may decide to provide transit between two peer ASes as a part of special business relationship
- Question…?
  - Is it possible to design a policy-based protocol that always converges?
    - Without imposing policy restrictions

Non-Monotonic Ranking

- An AS can attach MED route attribute to express its preferences regarding which route the neighbor should use (ASes are connected at more than one place)
  - MED values are set by the AS that advertise the route
  - Receiving AS can not compare it with routes received from some other ASes for the same destination

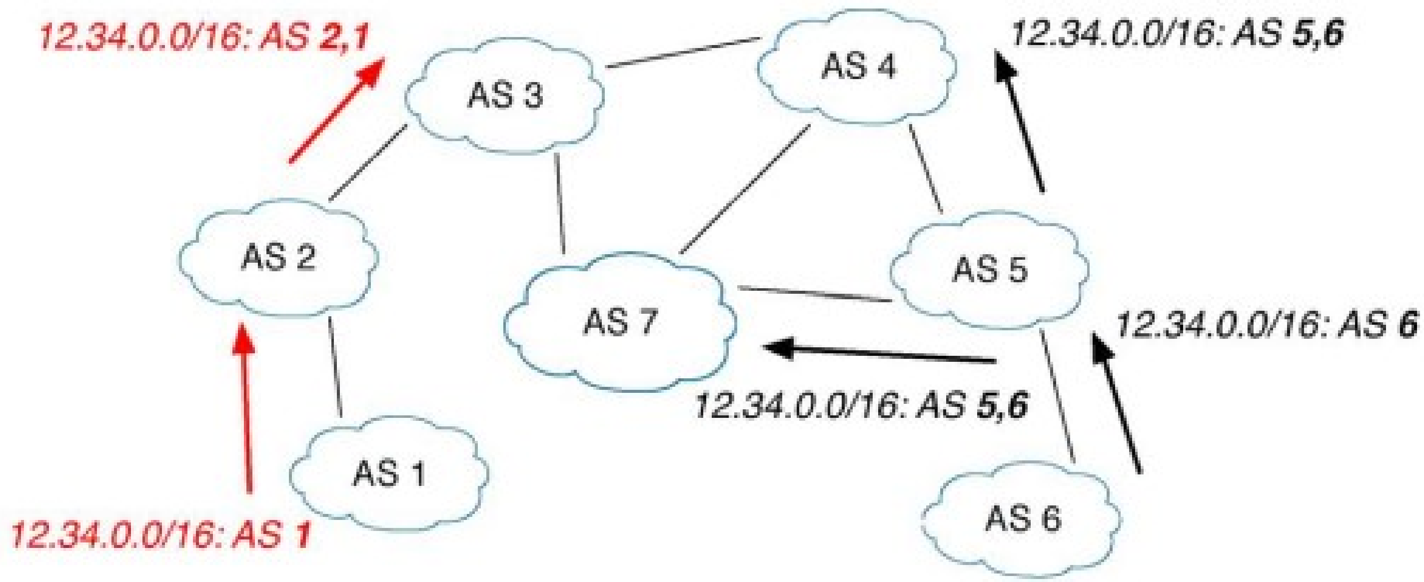
- Consequences
  - Routers may not have monotonic preferences between pairs of routes
    - Causes oscillations
  - Oscillations possible within a single AS also becoz routers within AS can not express monotonic ranking

Control Plane Security

- BGP does not allow an AS to verify that a route it learned is valid
- Also, it doesn't guarantees about where packets will actually go
- No support for controlling route announcements
  - An AS can advertise any arbitrary prefixes
  - No prefix authorization checking !!!
    - Vulnerable to prefix hijacking
- Difficult to check whether routes are policy complaint
  - ASes do not make their relationship public

- S-BGP Proposes IP Addresses – AS binding
  - Requires PKI (Public Key Infrastructure)
  - Costly as High message overhead

Example: Prefix Hijacking



Control Plane Security

- Open Questions
  1. Decentralized solution to verify prefix ownership….?
  2. Is it possible to design In-band verification scheme for ownership ?

Using TCP as the Transport Protocol

- Policy and routing information between two AS can be listen by an intruder
  - No security provision in TCP
- Possibility of the Man in the Middle Attack
  - BGP messages exchanged between two AS can be tempered (adding bogus info)
- Denial of Service Attacks
  - SYN Flooding attack
    - Some better routes can be neglected
    - Excessive messages can abort the session or crash the routers

Data Plane Security

- Assuming, the routes are authentic and policy complaint
  - **Does this verify** that route's AS path matches the actual forwarding path?
- A router should reject packets from sources that should not have a valid route through this router to the destination.
  - This requires to discover the routes from the source to that AS

Scalability Induced Problems

- BGP abstracts the routing details inside each AS and aggregates reach-ability information (i.e. prefix aggregation) about destinations
  - This provides scalability to BGP
  - But….makes difficult to determine cause of any routing updates
    - Convergence becomes slow
  - Hide fine-grained information about the reachability of destinations
    - Reduces AS control over incoming traffic

How to Secure BGP…?

- Secure message exchange between neighbors
  - No one should watch or tempering the exchanged messages

- Routing Information validity
  - Origin authentication
    - Is the prefix owned by the AS announcing it?
  - AS path authentication
    - Is AS path the sequence of ASes the update traversed?
  - AS path policy
    - Does AS path adhere to the routing policies of each AS?

Secure BGP Protocols

- S-BGP
  - Based on PKI
  - Validates path attributes between ASes using digital signatures and associated public key certificates

- Secure Origin BGP (SoBGP)
  - Use PKI for authorizing and authenticating entities and organizations

Cryptographic Techniques

- Shared key between two parties
  - Maintaining shared secret's Complexity becomes $O(n^2)$ for n peers
  - Needs frequent replacement
- Cryptographic Hash Functions
  - Message Digest Algorithm (MD 5)
  - Secure Hash Algorithm (SHA 1) Family
- Message Authentication Code
  - Unforgeable tag appended to message that provides security by guarantee the message integrity
- Public Key Cryptographic
  - Public key + Private key

BGP Security Today!!!

- Applying best common practices (BCPs)
  - Securing the session (authentication, encryption)
  - Filtering routes by prefix and AS path
  - Resetting attributes to default values
  - Packet filters to block unexpected control traffic

- This is not good enough
  - Depends on vigilant application of BCPs
    - … and not making configuration mistakes!
  - Doesn't address fundamental problems
    - Can't tell who owns the IP address block
    - Can't tell if the AS path is bogus or invalid
    - Can't be sure the data packets follow the chosen route

Next…

- BGP Mis-configurations

- Reference
  - Understanding BGP Mis-configuration [Ratul Mahajan, 2002]

BGP Mis-configurations Causes

- Accidental injection of routes into global BGP tables
  - e.g. due to address space hijacks

- Accidental export of routes in violation of an ISP's policy
  - e.g. due to human errors

BGP Mis-configurations Types

- Origin Misconfiguration: accidently injects a prefix into the global BGP tables
  - Failure to summaries an address space, leads to the injection of one or more specific prefixes
  - Prefix Hijack
- Export Misconfiguration: AS-path is in violation of the policies of one of the ASes in the path
  - Router exported a route, it should have filtered

Impact of Misconfigurations

- Adverse Impact of Misconfigurations
  - Routing load
  - Connectivity Disruption
  - Policy Violations

- Mis-configuration Identification
  - In general ***valid routes stay for a longer period***
  - Route changes last for ***less than a day*** are treated as misconfigured
    - Based on Real time observation
    - Vague conclusion! Why

Origin Mis-configuration Types

	Old route	New route
Self deagg-regation	a.b.0.0/16 X Y Z	a.b.c.0/24 X Y Z
Related origin	a.b.0.0/16 X Y Z	a.b.0.0/16 X Y a.b.0.0/16 X Y Z O a.b.c.0/24 X Y a.b.c.0/24 X Y Z O
Foreign origin	a.b.0.0/16 X Y Z	a.b.0.0/16 X Y O a.b.c.0/24 X Y O e.f.g.h/i X Y O

Related Origin: An existing prefix (or subset) is advertised by a new but related origin (one of the origins appears in the AS path of the other)

Possible Causes

Self De-aggregation: Forget to aggregate at a router

Related Origin: Likely connected to the network

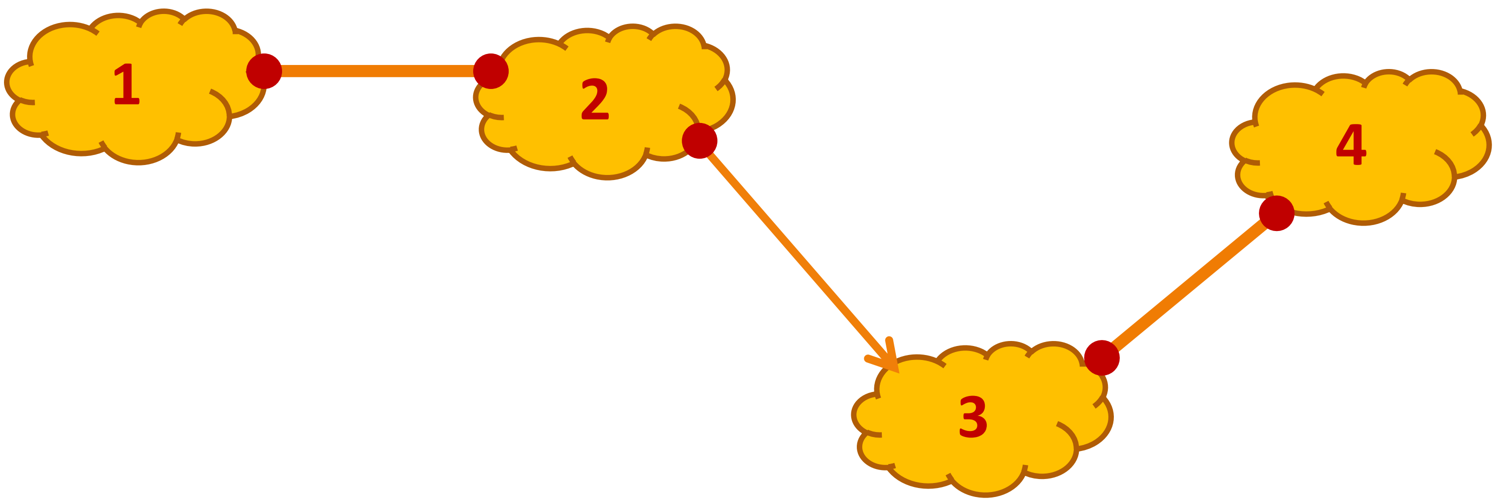
Foreign Origin: Due to Prefix Hijack

Export Policy

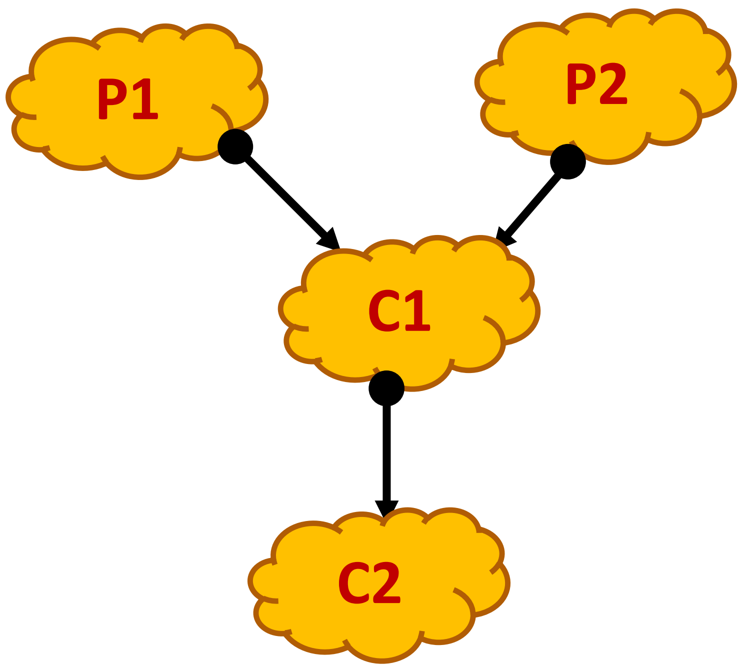
- Export policies arises from commercial relationship between ASes
- Knowing the relationship between ASes makes mis-configuration detection simple
  - But AS relationship is not available…!

- Observations from BGP Routing Tables
  - An AS path can have at most one peer to peer edge which occurs at highest point in the path
  - ASes with more neighbors are more likely to be providers
  - Valid AS paths are Valley free
    - Provider to customer is downward direction, sibling and peers same level
    - Routes that starts going downwards never goes up again

Example: Valley Free Violation



Example: Export Mis-configuration



- **<u>Intended policy at C1</u>**<u>:</u> Provide transit to **C2** through link **C1-C2**
- **<u>Configured policy:</u>** Export all routes originated by **C2** to **P1** and **P2**
- **<u>Correct policy</u>:** export only when AS path is "**C2**"

Questions!!!

Thank You!