

CSCI-1680
Network Layer:
Inter-domain Routing – Policy and Security

Rodrigo Fonseca



Administrivia

- **Homework 2 is out**
- **Rodrigo's office hours:**
 - Monday 1-3 (or by appointment)



Today

- **BGP Continued**
 - Policy routing, instability, vulnerabilities



Route Selection

- **More specific prefix**
- **Next-hop reachable?**
- **Prefer highest weight**
 - Computed using some AS-specific local policy
- **Prefer highest local-pref**
- **Prefer locally originated routes**
- **Prefer routes with shortest AS path length**
- **Prefer eBGP over iBGP**
- **Prefer routes with lowest cost to egress point**
 - Hot-potato routing
- **Tie-breaking rules**
 - E.g., oldest route, lowest router-id



Customer/Provider AS relationships

- **Customer pays for connectivity**
 - E.g. Brown contracts with OSHEAN
 - Customer is stub, provider is a transit
- **Many customers are multi-homed**
 - E.g., OSHEAN connects to Level3, Cogent
- **Typical policies:**
 - Provider tells all neighbors how to reach customer
 - Provider prefers routes from customers (\$\$)
 - Customer does not provide transit service



Peer Relationships

- **ASs agree to exchange traffic for free**
 - Penalties/Renegotiate if imbalance
- **Tier 1 ISPs have no default route: all peer with each other**
- **You are Tier $i + 1$ if you have a default route to a Tier I**
- **Typical policies**
 - AS only exports customer routes to peer
 - AS exports a peer's routes only to its customers
 - Goal: avoid being transit when no gain



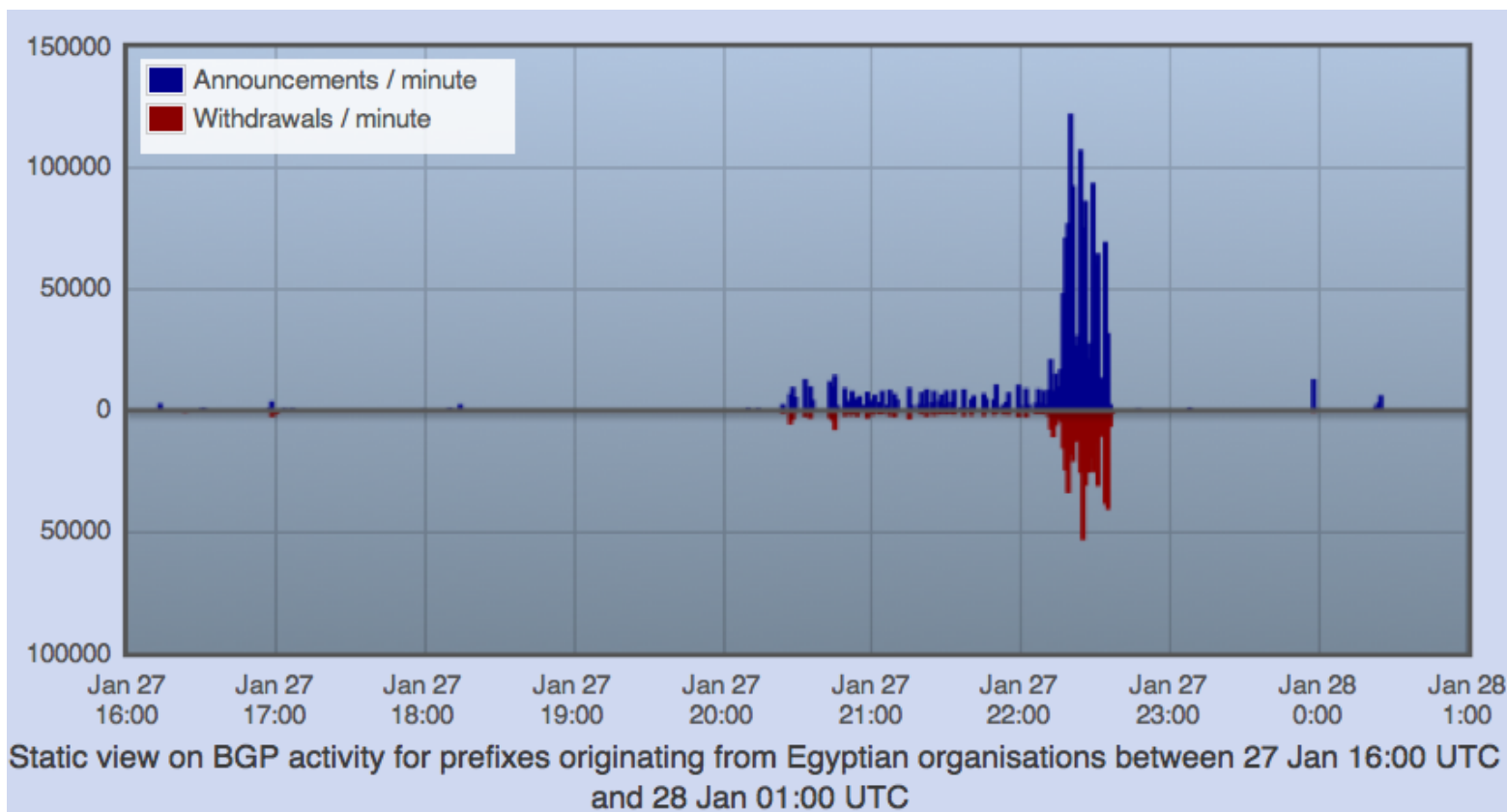
Peering Drama

- Cogent vs. Level3 were peers
- In 2003, Level3 decided to start charging Cogent
- Cogent said no
- **Internet partition:** Cogent's customers couldn't get to Level3's customers and vice-versa
 - Other ISPs were affected as well
- Took 3 weeks to reach an undisclosed agreement



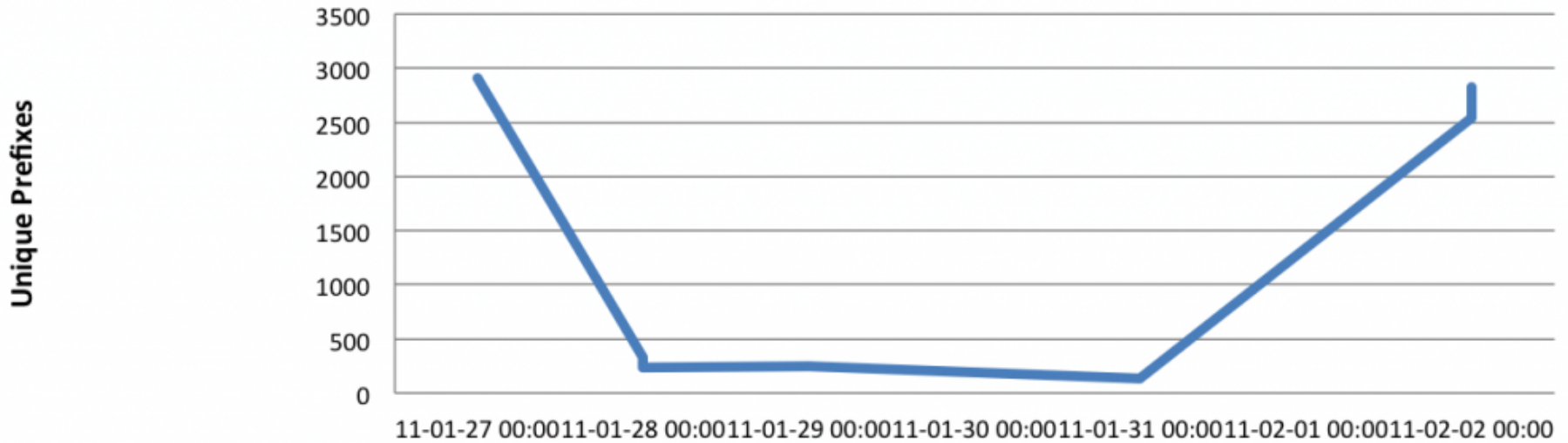
“Shutting off” the Internet

- Starting from Jan 27th, 2011, Egypt was disconnected from the Internet
 - 2769/2903 networks withdrawn from BGP (95%)!



Egypt Incident

Number of Egyptian networks



	11-01-27 00:00	11-01-28 02:00	11-01-28 16:00	11-01-28 20:00	11-01-29 00:00	11-01-29 18:00	11-01-31 22:00	11-02-02 10:00	11-02-02 12:00
Number of Egyptian networks	2903	327	239	241	242	243	134	2539	2825



Some BGP Challenges

- **Convergence**
- **Traffic engineering**
 - How to assure certain routes are selected
- **Scaling (route reflectors)**
- **Security**

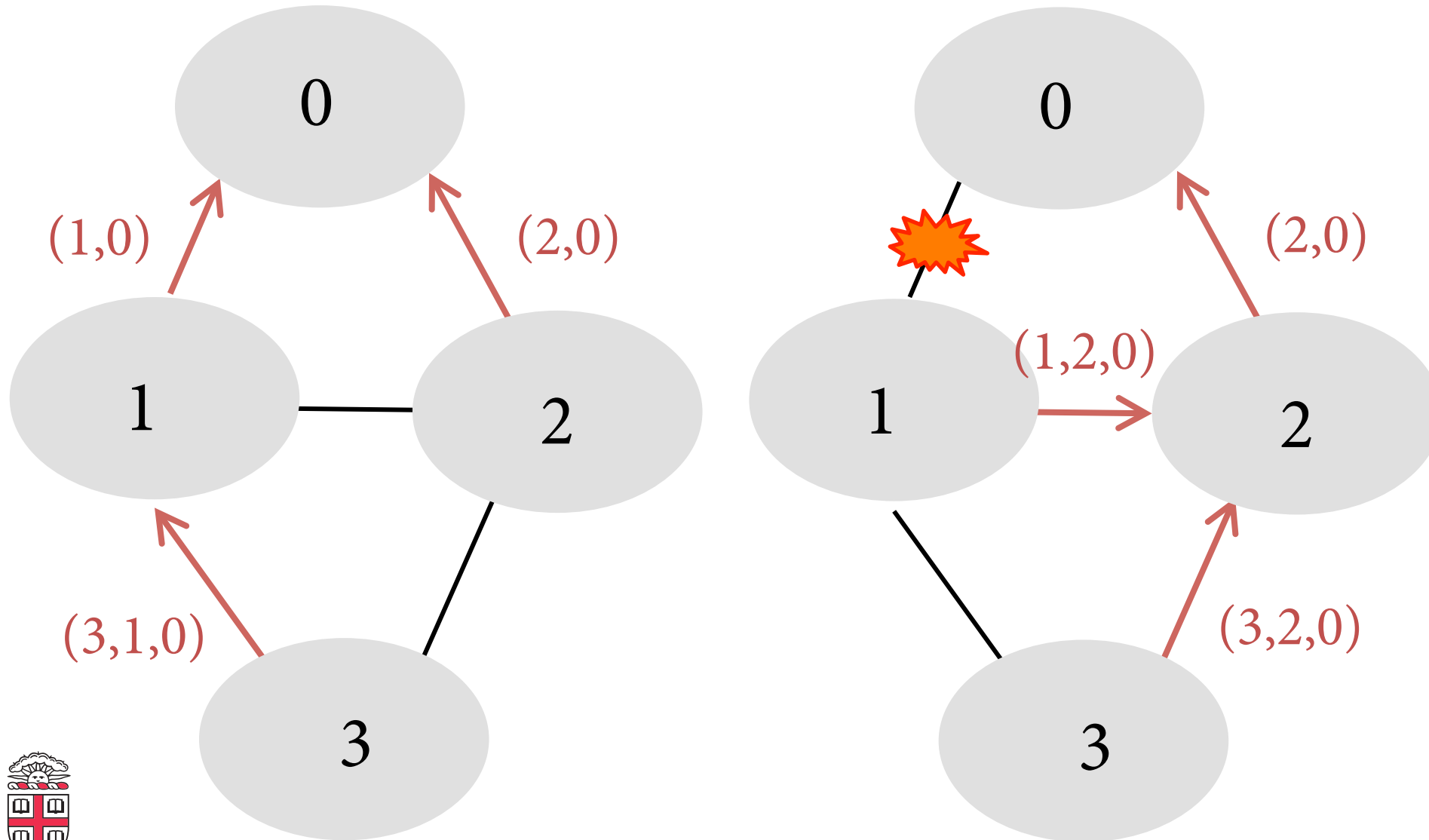


Convergence

- **Given a change, how long until the network re-stabilizes?**
 - Depends on change: sometimes never
 - Open research problem: “tweak and pray”
 - Distributed setting is challenging
- **Some reasons for change**
 - Topology changes
 - BGP session failures
 - Changes in policy
 - Conflicts between policies can cause oscillation

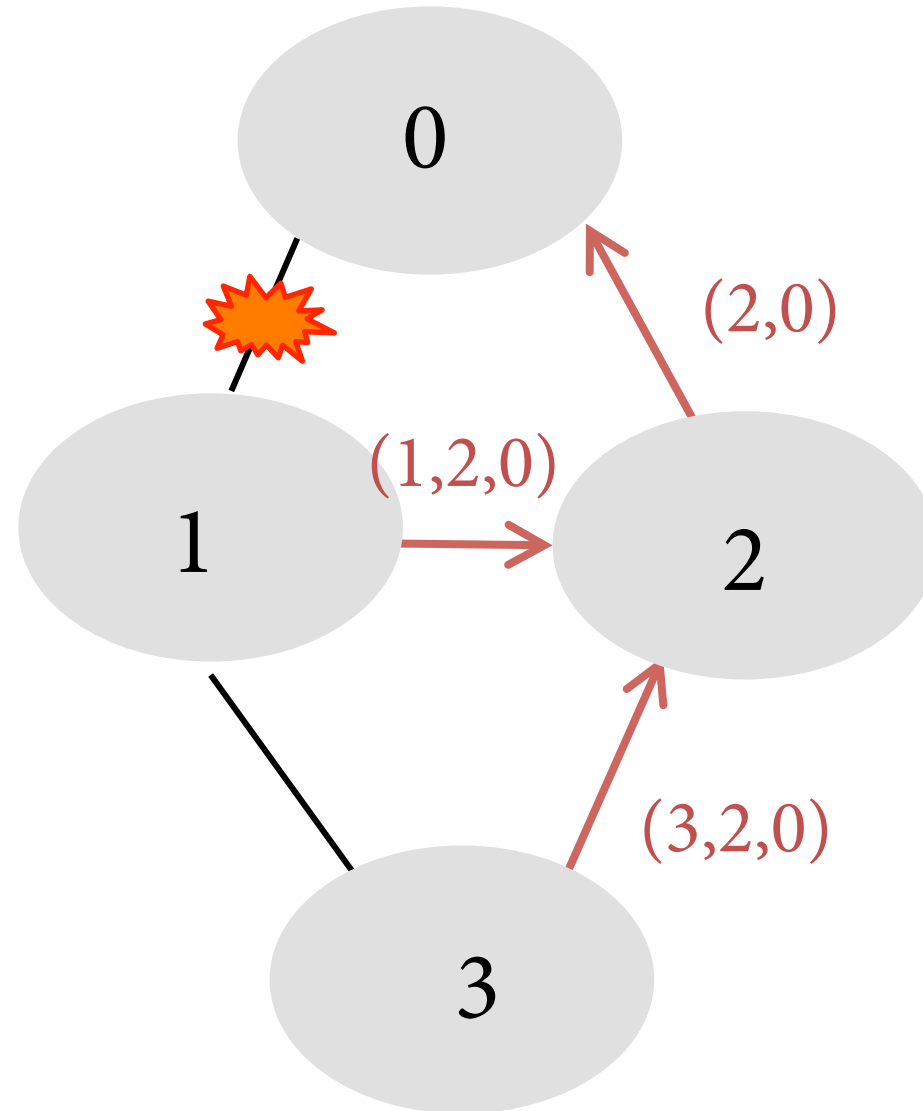


Routing Change: Before and After



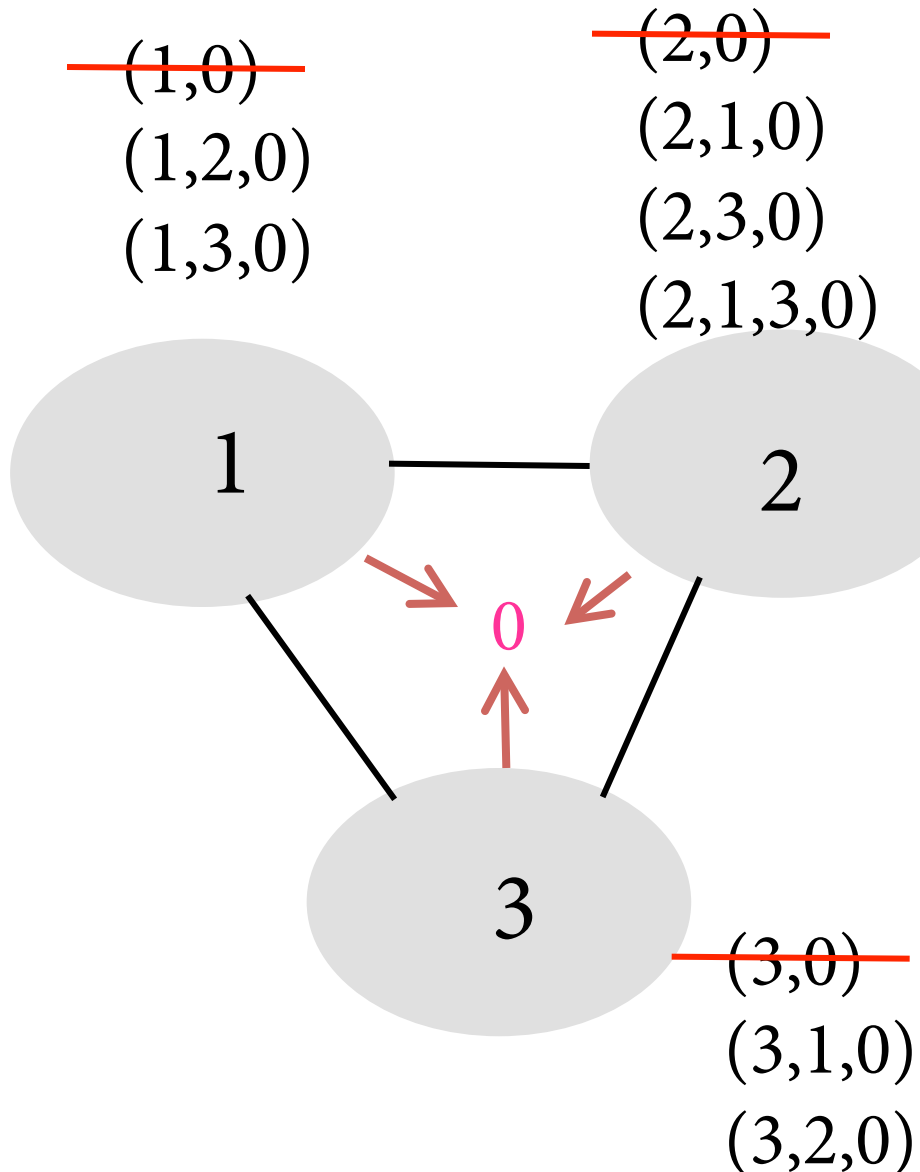
Routing Change: Path Exploration

- **AS 1**
 - Delete the route (1,0)
 - Switch to next route (1,2,0)
 - Send route (1,2,0) to AS 3
- **AS 3**
 - Sees (1,2,0) replace (1,0)
 - Compares to route (2,0)
 - Switches to using AS 2



Routing Change: Path Exploration

- **Initial situation**
 - Destination 0 is alive
 - All ASes use direct path
- **When destination dies**
 - All ASes lose direct path
 - All switch to longer paths
 - Eventually withdrawn
- **E.g., AS 2**
 - $(2,0) \rightarrow (2,1,0)$
 - $(2,1,0) \rightarrow (2,3,0)$
 - $(2,3,0) \rightarrow (2,1,3,0)$
 - $(2,1,3,0) \rightarrow \text{null}$



• **Convergence may be slow!**



Route Engineering

- **Route filtering**
- **Setting weights**
- **More specific routes: longest prefix**
- **AS prepending: “477 477 477 477”**
- **More of an art than science**



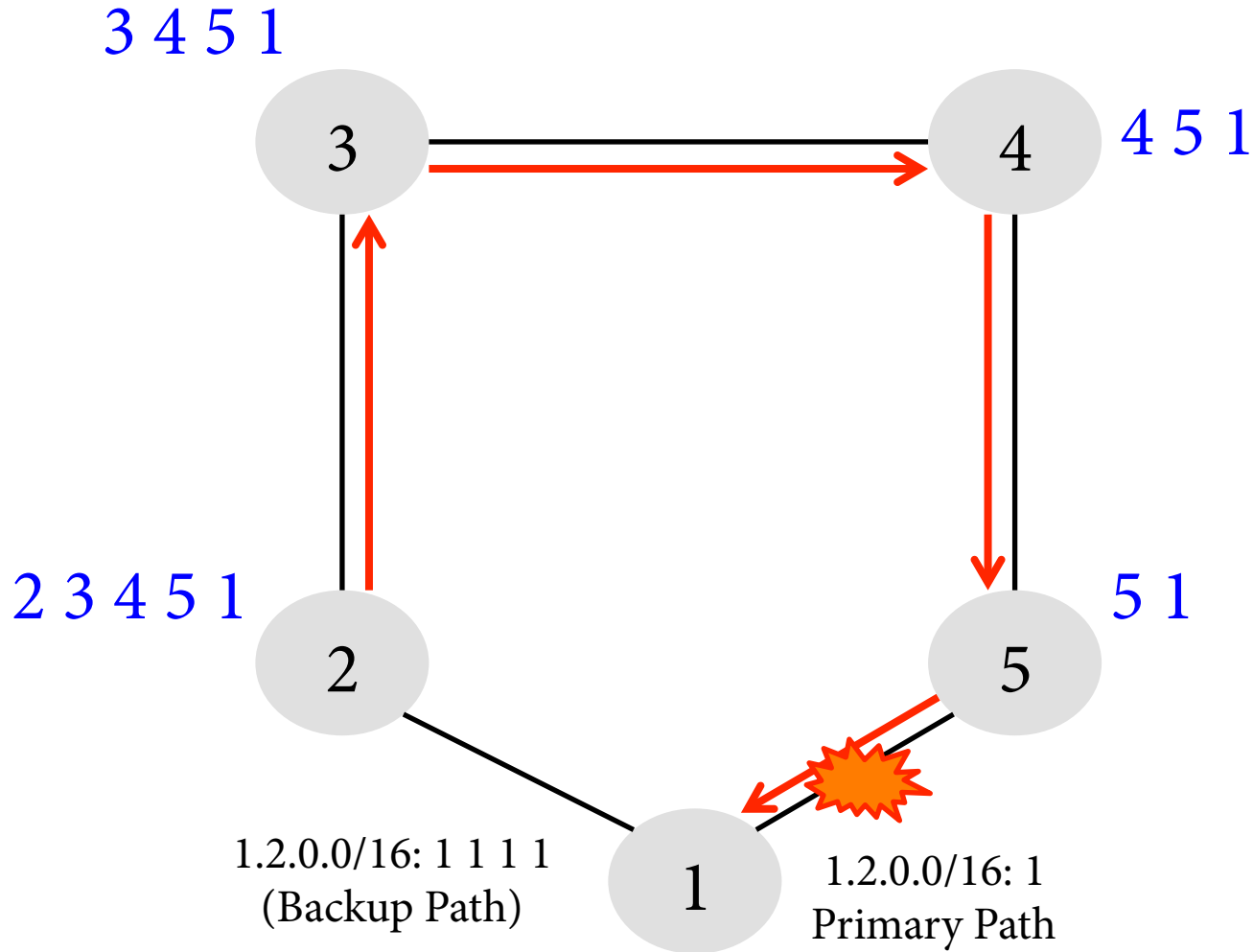
Multiple Stable Configurations

BGP Wedgies [RFC 4264]

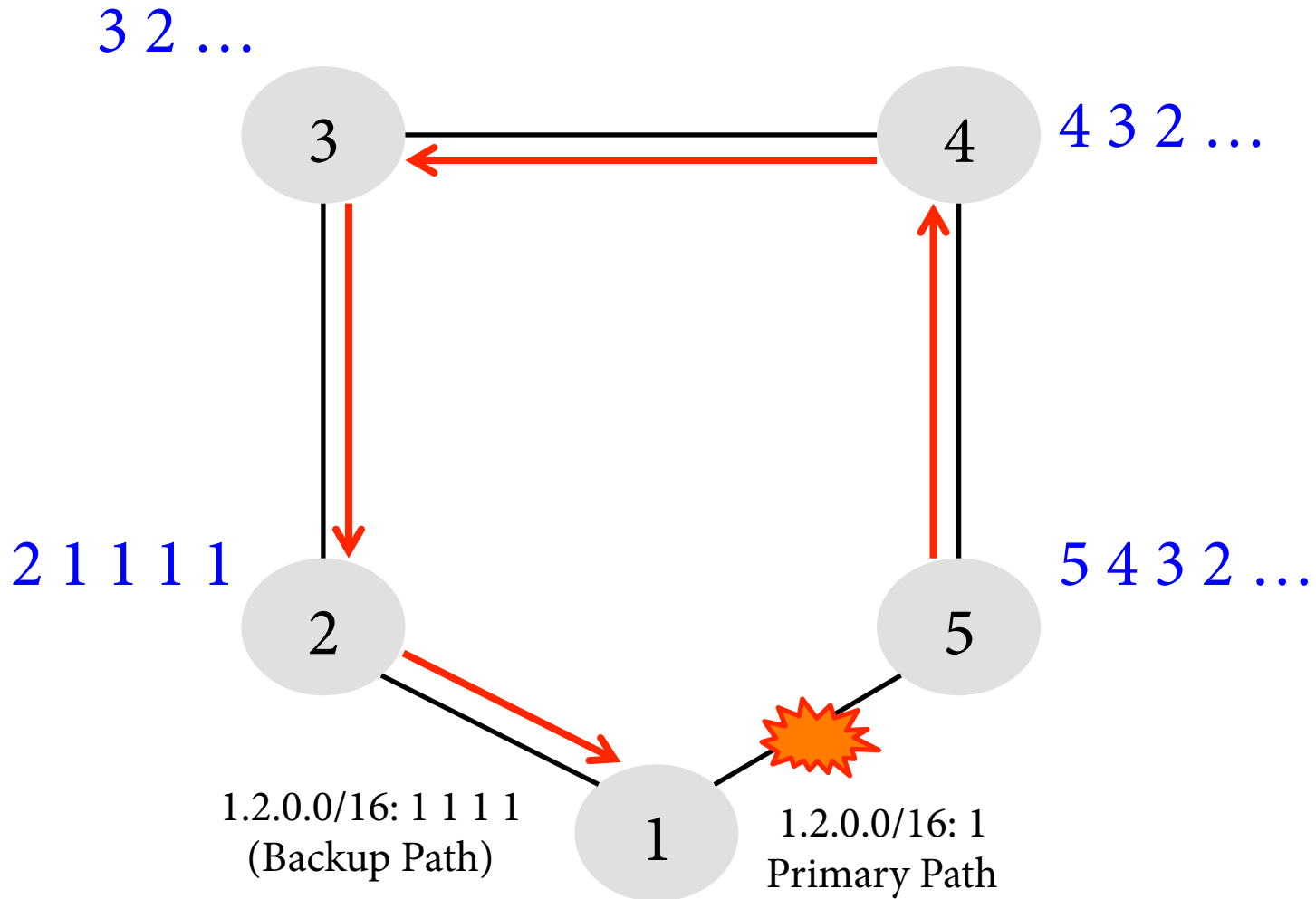
- **Typical policy:**
 - Prefer routes from customers
 - Then prefer shortest paths



BGP Wedgies



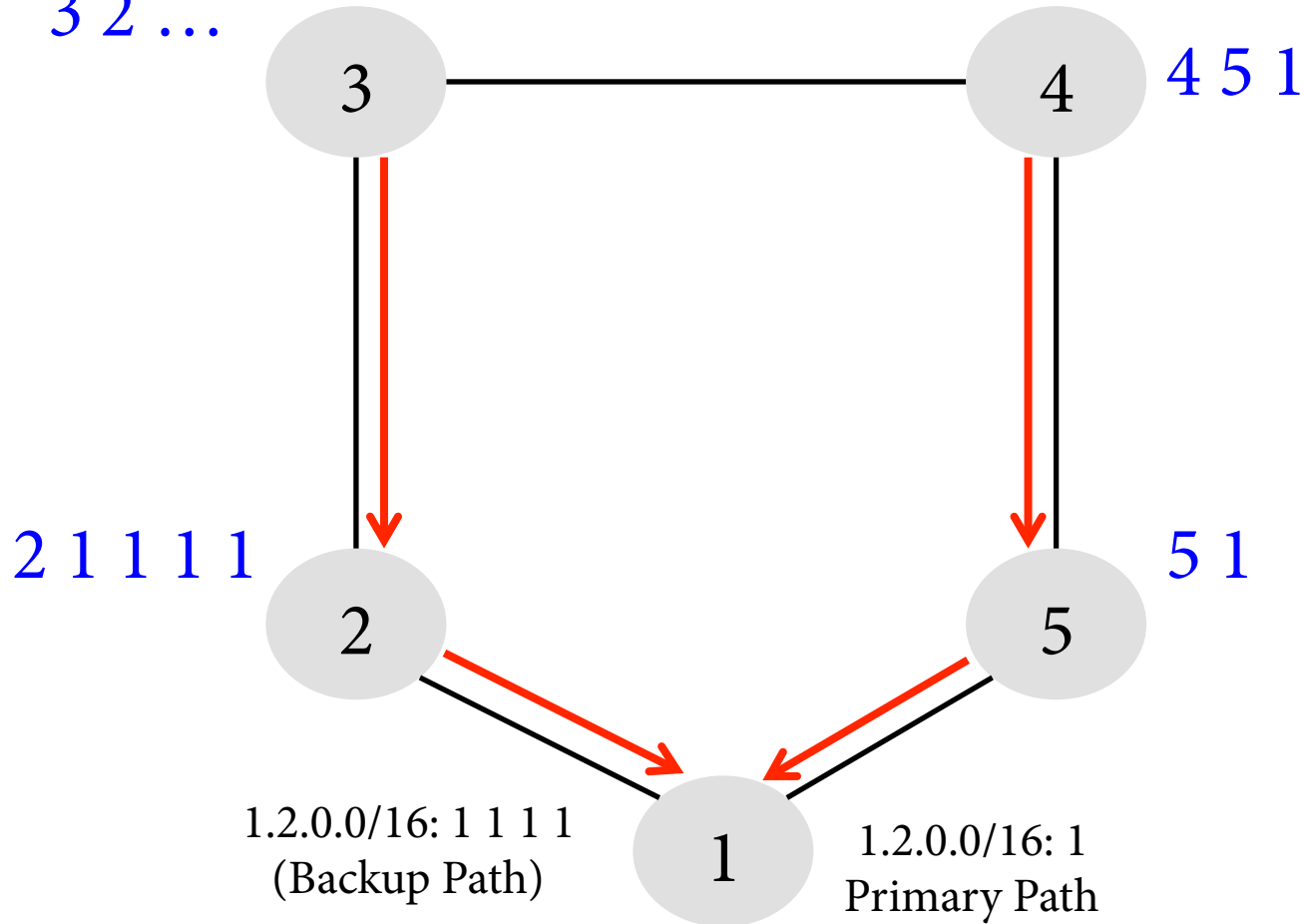
BGP Wedgies



BGP Wedgies

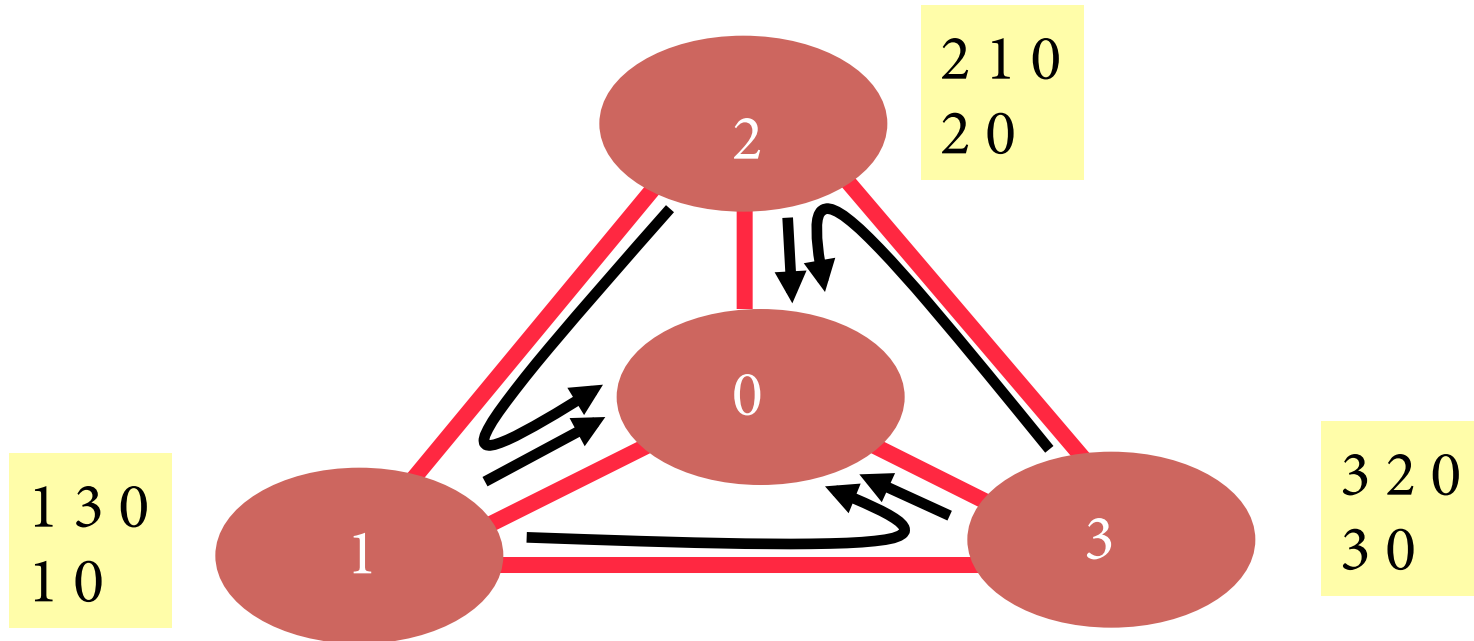
3 prefers customer route: stable configuration!

3 2 ...



Unstable Configurations

- Due to policy conflicts



Avoiding BGP Instabilities

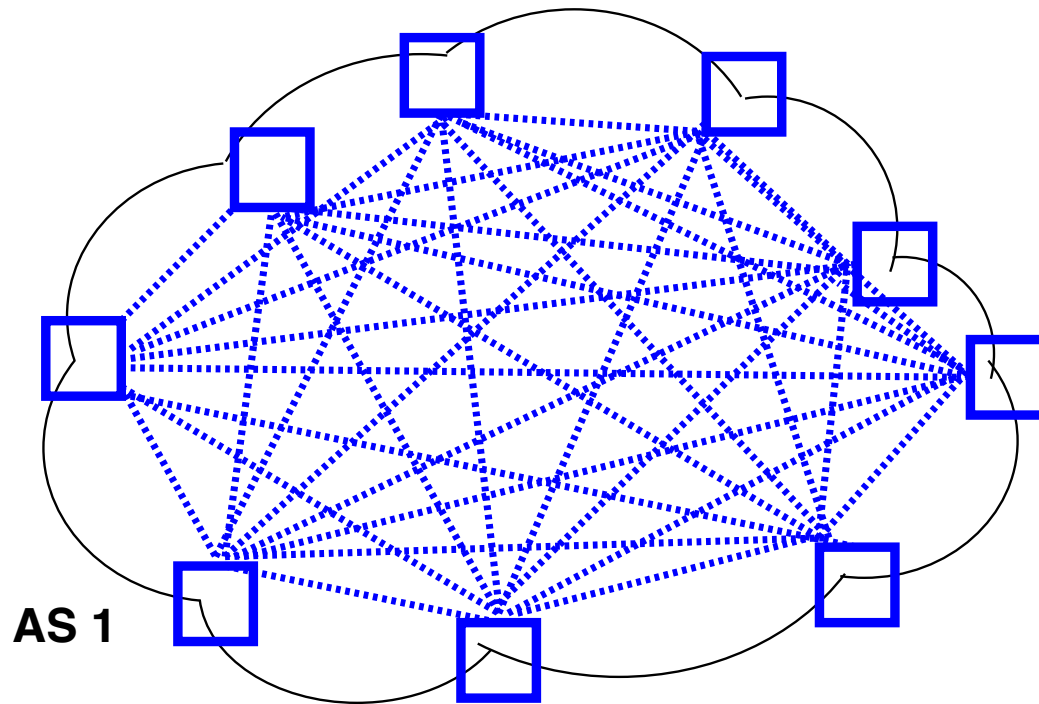
- **Detecting conflicting policies**
 - Centralized: NP-Complete problem!
 - Distributed: open research problem
 - Requires too much cooperation
- **Detecting oscillations**
 - Monitoring for repetitive BGP messages
- **Restricted routing policies and topologies**
 - Some topologies / policies proven to be safe*

* Gao & Rexford, “Stable Internet Routing without Global Coordination”, IEEE/ACM ToN, 2001



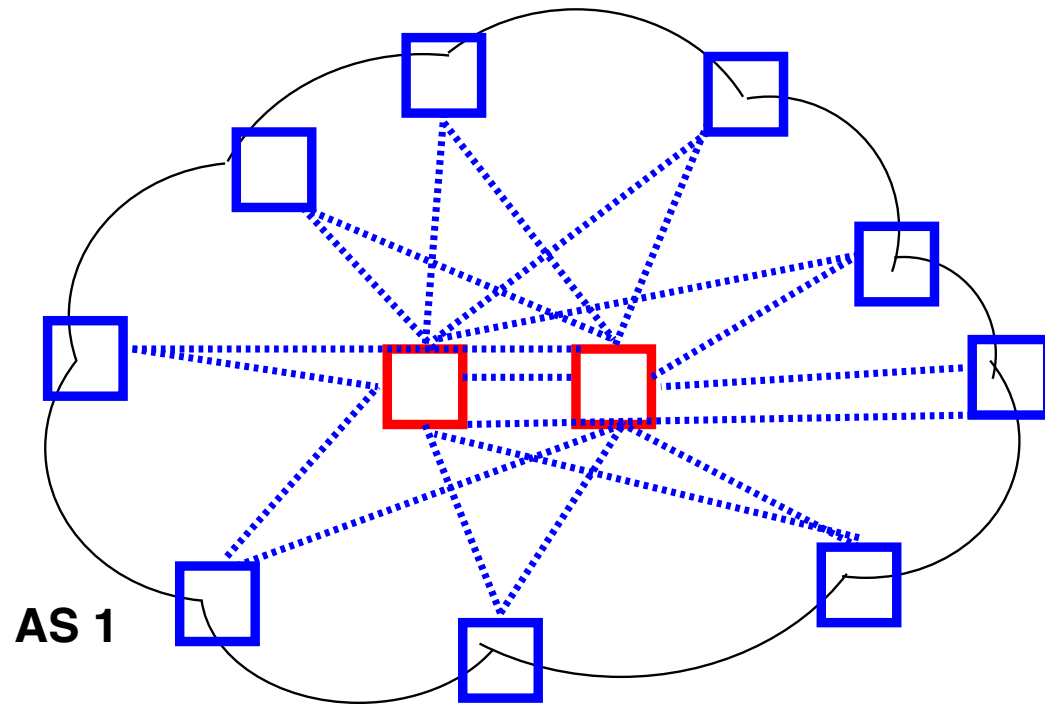
Scaling iBGP: route reflectors

iBGP Mesh == $O(n^2)$ mess



Scaling iBGP: route reflectors

Solution: Route Reflectors
 $O(n*k)$



BGP Security Goals

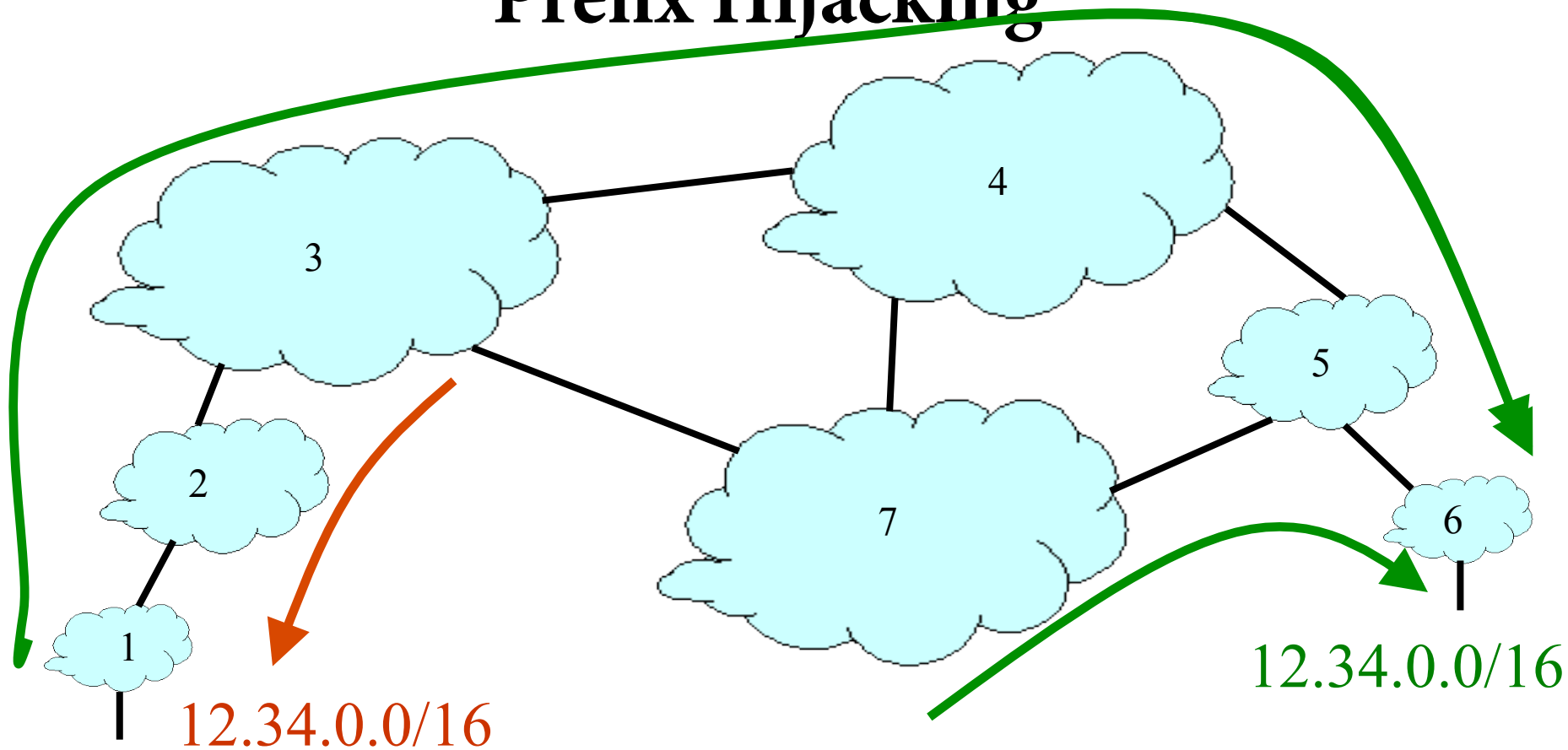
- Confidential message exchange between neighbors
- **Validity of routing information**
 - Origin, Path, Policy
- Correspondence to the data path



Origin: IP Address Ownership and Hijacking

- **IP address block assignment**
 - Regional Internet Registries (ARIN, RIPE, APNIC)
 - Internet Service Providers
- **Proper origination of a prefix into BGP**
 - By the AS who owns the prefix
 - ... or, by its upstream provider(s) in its behalf
- **However, what's to stop someone else?**
 - Prefix hijacking: another AS originates the prefix
 - BGP does not verify that the AS is authorized
 - Registries of prefix ownership are inaccurate

Prefix Hijacking



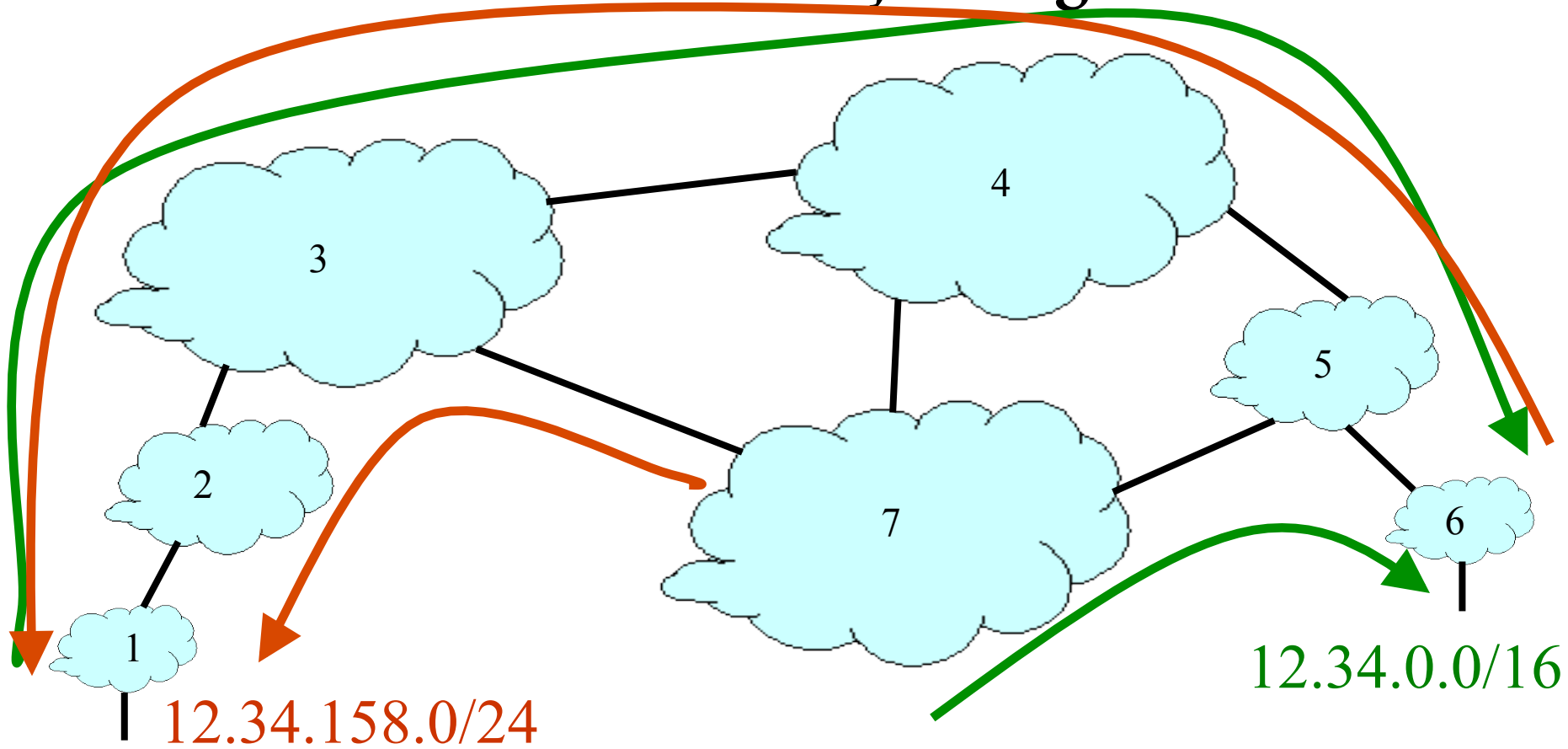
- **Consequences for the affected ASes**

- Blackhole: data traffic is discarded
- Snooping: data traffic is inspected, and then redirected
- Impersonation: data traffic is sent to bogus destinations

Hijacking is Hard to Debug

- **Real origin AS doesn't see the problem**
 - Picks its own route
 - Might not even learn the bogus route
- **May not cause loss of connectivity**
 - E.g., if the bogus AS snoops and redirects
 - ... may only cause performance degradation
- **Or, loss of connectivity is isolated**
 - E.g., only for sources in parts of the Internet
- **Diagnosing prefix hijacking**
 - Analyzing updates from many vantage points
 - Launching traceroute from many vantage points

Sub-Prefix Hijacking



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

How to Hijack a Prefix

- **The hijacking AS has**
 - Router with eBGP session(s)
 - Configured to originate the prefix
- **Getting access to the router**
 - Network operator makes configuration mistake
 - Disgruntled operator launches an attack
 - Outsider breaks in to the router and reconfigures
- **Getting other ASes to believe bogus route**
 - Neighbor ASes not filtering the routes
 - ... e.g., by allowing only expected prefixes
 - But, specifying filters on *peering* links is hard

Pakistan Youtube incident

- Youtube's has prefix 208.65.152.0/22
- Pakistan's government order Youtube blocked
- Pakistan Telecom (AS 17557) announces 208.65.153.0/24 in the wrong direction (outwards!)
- Longest prefix match caused worldwide outage
- <http://www.youtube.com/watch?v=IzLPKuAOe50>



Many other incidents

- **Spammers steal unused IP space to hide**
 - Announce very short prefixes (e.g., /8). Why?
 - For a short amount of time
- **China incident, April 8th 2010**
 - China Telecom's AS23724 generally announces 40 prefixes
 - On April 8th, announced ~37,000 prefixes
 - About 10% leaked outside of China
 - Suddenly, going to www.dell.com might have you routing through AS23724!



Attacks on BGP Paths

- **Remove an AS from the path**
 - E.g., 701 3715 88 -> 701 88
- **Why?**
 - Attract sources that would normally avoid AS 3715
 - Make AS 88 look like it is closer to the core
 - Can fool loop detection!
- **May be hard to tell whether this is a lie**
 - 88 could indeed connect directly to 701!



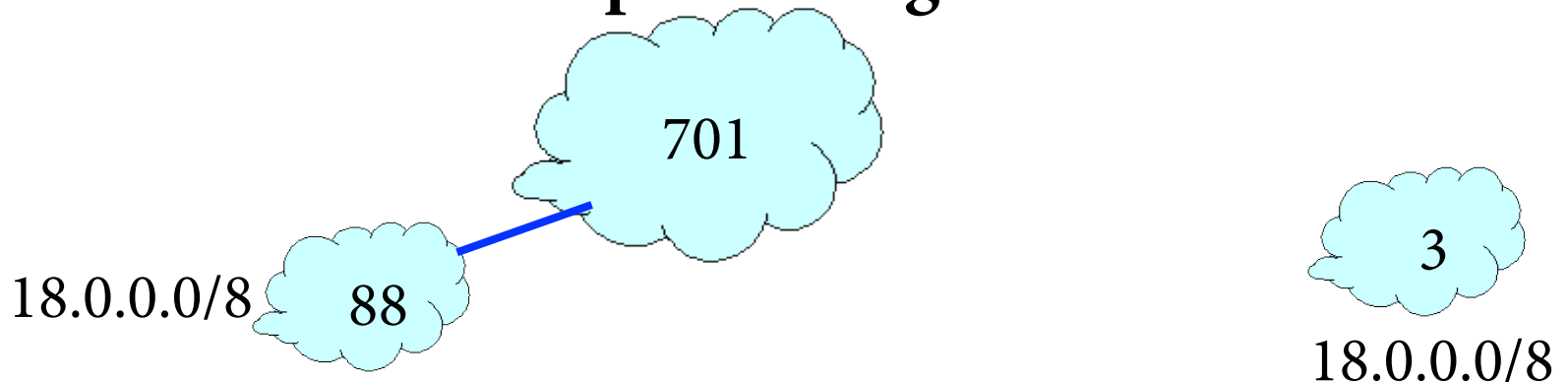
Attacks on BGP Paths

- **Adding ASes to the path**
 - E.g., 701 88 -> 701 3715 88
- **Why?**
 - Trigger loop detection in AS 3715
 - This would block unwanted traffic from AS 3715!
 - Make your AS look more connected
- **Who can tell this is a lie?**
 - AS 3715 could, if it could see the route
 - AS 88 could, but would it really care?



Attacks on BGP Paths

- **Adding ASes at the end of the path**
 - E.g., 701 88 into 701 88 3
- **Why?**
 - Evade detection for a bogus route (if added AS is legitimate owner of a prefix)
- **Hard to tell that the path is bogus!**



Proposed Solution: S-BGP

- **Based on a public key infrastructure**
- **Address attestations**
 - Claims the right to originate a prefix
 - Signed and distributed out of band
 - Checked through delegation chain from ICANN
- **Route attestations**
 - Attribute in BGP update message
 - Signed by each AS as route along path
- **S-BGP can avoid**
 - Prefix hijacking
 - Addition, removal, or reordering of intermediate ASes



S-BGP Deployment

- **Very challenging**
 - PKI
 - Accurate address registries
 - Need to perform cryptographic operations on all path operations
 - Flag day almost impossible
 - Incremental deployment offers little incentive
- **But there is hope! [Goldberg et al, 2011]**
 - Road to incremental deployment
 - Change rules to break ties for secure paths
 - If a few top Tier-1 ISPs
 - Plus their respective stub clients deploy simplified version (just sign, not validate)
 - Gains in traffic => \$ => adoption!



Data Plane Attacks

- **Routers/ASes can advertise one route, but not necessarily follow it!**
- **May drop packets**
 - Or a fraction of packets
 - What if you just slow down some traffic?
- **Can send packets in a different direction**
 - Impersonation attack
 - Snooping attack
- **How to detect?**
 - Congestion or an attack?
 - Can let ping/traceroute packets go through
 - End-to-end checks?
- **Harder to pull off, as you need control of a router**



BGP Recap

- **Key protocol that holds Internet routing together**
- **Path Vector Protocol among Autonomous Systems**
- **Policy, feasibility first; non-optimal routes**
- **Important security problems**



Next Class

- **Network layer wrap up**
- **Intro to TCP**

