Last week on Communication Networks

The Internet is a network of networks, referred to as Autonomous Systems (AS)

BGP is the routing protocol “glueing” the Internet together

Using BGP, ASes exchange information about the IP prefixes they can reach, directly or indirectly

BGP needs to solve three key challenges: scalability, privacy and policy enforcement

Inters-domain routing

Link-state protocols
Distance-vector protocols

Path-vector protocols

There is a huge # of networks and prefixes
700k prefixes, >50,000 networks, millions (!) of routers

Networks don’t want to divulge internal topologies or their business relationships

Networks need to control where to send and receive traffic without an Internet-wide notion of a link cost metric
BGP relies on path-vector routing to support flexible routing policies and avoid count-to-infinity.

**key idea** advertise the *entire path* instead of distances.

---

**Border Gateway Protocol**

policies and more

---

**BGP suffers from many rampant problems**

<table>
<thead>
<tr>
<th>Problems</th>
<th>Reachability</th>
<th>Security</th>
<th>Convergence</th>
<th>Performance</th>
<th>Anomalies</th>
<th>Relevance</th>
</tr>
</thead>
</table>

---

Like normal routing, policy routing does not guarantee reachability even if the graph is connected.

Because of policies, Swisscom cannot reach DT even if the graph is connected.

---

Many security considerations are simply **absent** from BGP specifications.

- ASes can advertise any prefixes even if they don’t own them!
- ASes can arbitrarily modify route content e.g., change the content of the AS-PATH.
- ASes can forward traffic along different paths than the advertised one.
BGP (lack of) security

#1 BGP does not validate the origin of advertisements
#2 BGP does not validate the content of advertisements

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

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

Hijacking is Hard to Debug

- The victim AS doesn’t see the problem
  - Picks its own route, might not learn the bogus route
- May not cause loss of connectivity
  - Snooping, with minor 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 BGP 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 do not discard the bogus route
  - E.g., not doing protective filtering

YouTube Outage on Feb 24, 2008

- YouTube (AS 36561)
  - Web site www.youtube.com (208.65.152.0/22)
- Pakistan Telecom (AS 17557)
  - Government order to block access to YouTube
  - Announces 208.65.153.0/24 to PCCW (AS 3491)
    - All packets to YouTube get dropped on the floor
- Mistakes were made
  - AS 17557: announce to everyone, not just customers
  - AS 3491: not filtering routes announced by AS 17557
- Lasted 100 minutes for some, 2 hours for others
Timeline (UTC Time)

- 18:47:45
  - First evidence of hijacked /24 route in Asia
- 18:48:00
  - Several big trans-Pacific providers carrying the route
- 18:49:30
  - Bogus route fully propagated
- 20:07:25
  - YouTube starts advertising /24 to attract traffic back
- 20:08:30
  - Many (but not all) providers are using valid route

Timeline (UTC Time)

- 20:18:43
  - YouTube announces two more-specific /25 routes
- 20:19:37
  - Some more providers start using the /25 routes
- 20:50:59
  - AS 17557 starts prepending (“3491 17557 17557”)
- 20:59:39
  - AS 3491 disconnects AS 17557
- 21:00:00
  - Videos of cats flushing toilets are available again!

Another Example: Spammers

- Spammers sending spam
  - Form a (bidirectional) TCP connection to mail server
  - Send a bunch of spam e-mail, then disconnect
- But, best not to use your real IP address
  - Relatively easy to trace back to you
- Could hijack someone’s address space
  - But you might not receive all the (TCP) return traffic
- How to evade detection
  - Hijack unused (i.e., unallocated) address block
  - Temporarily use the IP addresses to send your spam

BGP (lack of) security

#1 BGP does not validate the origin of advertisements
#2 BGP does not validate the content of advertisements

Bogus AS Paths

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

Bogus AS Paths

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

Bogus AS Paths

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

Invalid Paths

- AS exports a route it shouldn’t
  - AS path is a valid sequence, but violated policy
- Example: customer misconfiguration
  - Exports routes from one provider to another
- Interacts with provider policy
  - Provider prefers customer routes
  - Directing all traffic through customer
- Main defense
  - Filtering routes based on prefixes and AS path
**Missing/Inconsistent Routes**

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

**BGP Security Today**

- **Applying best common practices (BCPs)**
  - Securing the session (authentication, encryption)
  - Filtering routes by prefix and AS path
  - Packet filters to block unexpected control traffic
- **This is not good enough**
  - Depends on vigilant application of BCPs
  - 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

Routing attacks can be used to de-anonymize Tor users

Routing attacks can be used to partition the Bitcoin network

With arbitrary policies, BGP may have multiple stable states

If AS2 is the first to advertise 2 0, the system stabilizes in a state where AS 1 is happy

If AS1 is the first one to advertise 1 0, the system stabilizes in a state where AS 2 is happy

Routing attacks can be used to partition the Bitcoin network
The actual assignment depends on the ordering between the messages

Note that AS1/AS2 could change the outcome by manual intervention


With arbitrary policies, BGP may fail to converge

Initially, all ASes only know the direct route to 0

AS 1 advertises its path to AS 2

Upon reception, AS 2 switches to 2 1 0 (preferred)

AS 3 advertises its path to AS 1
Upon reception, AS 1 switches to 1 3 0 (preferred)

AS 1 advertises its new path 1 3 0 to AS 2

Upon reception, AS 2 reverts back to its initial path 2 0

AS 2 advertises its path 2 0 to AS 3

Upon reception, AS 3 switches to 3 2 0 (preferred)

AS 3 advertises its new path 3 2 0 to AS 1

Upon reception, AS 1 reverts back to 1 0 (initial path)

AS 1 advertises its new path 1 0 to AS 2
Upon reception, AS 2 switches to 2 1 0 (preferred)

AS 2 advertises its new path 2 1 0 to AS 3

Upon reception, AS 3 switches to its initial path 3 0

We are back where we started, from there on, the oscillation will continue forever

Policy oscillations are a direct consequence of policy autonomy

Guaranteeing the absence of oscillations is hard even when you know all the policies!

How come?

Computationally, a BGP network is as "powerful" as

see "Using Routers to Build Logic Circuits: How Powerful is BGP?"
How do you prove such a thing?

Easy, you build a computer using BGP...

Logic gates

$$\text{Logic gates}$$

$$\text{Input 1}$$

$$\text{Input 2}$$

$$\text{Output}$$

$$\text{Logic gates}$$

$$\text{Memory}$$

$$\text{Clock}$$

BGP has it all!

$$\text{BGP has it all!}$$

$$\text{Input 1}$$

$$\text{Input 2}$$

$$\text{Output}$$

$$\text{BGP config}$$

$$\text{Memory}$$

$$\text{Clock}$$

famous incorrect BGP configurations (Griffin et al.)

Together, BGP routers form the largest computer in the world!

Router-level view of the Internet, OPTE project

Instead of using Minecraft for building a computer... use BGP!

Hack III, Minecraft's largest computer to date

Instead of using Minecraft for building a computer... use BGP!
Theorem 1
Determining whether a finite BGP network converges is PSPACE-hard.

Theorem 2
Determining whether an infinite BGP network converges is Turing-complete.

Checking BGP correctness is as hard as checking the termination of a general program.

Reachability
Security
Convergence
Performance
Anomalies
Relevance

Problems

BGP path selection is mostly economical, not based on accurate performance criteria.

BGP configuration is hard to get right, you’ll understand that very soon.

BGP is both “bloated” and underspecified
lots of knobs and (sometimes, conflicting) interpretations

BGP is often manually configured
humans make mistakes, often

BGP abstraction is fundamentally flawed
disjoint, router-based configuration to effect AS-wide policy

In practice though, BGP does not oscillate “that” often

known as “Gao-Rexford” rules

Theorem
If all AS policies follow the cust/peer/provider rules, BGP is guaranteed to converge.

Intuition
Oscillations require “preferences cycles” which make no economical sense.

BGP says that path 4 -> 1 is better than path 3 -> 2 -> 1.

BGP configuration is hard to get right, you’ll understand that very soon.

https://www.theregister.co.uk/2017/08/27/google_routing_blunder_sent_japans_internet_dark/
In August 2017
Someone in Google fat-thumbed a Border Gateway Protocol (BGP) advertisement and sent Japanese Internet traffic into a black hole.

In August 2017
Someone in Google fat-thumbed a Border Gateway Protocol (BGP) advertisement and sent Japanese Internet traffic into a black hole. Traffic from Japanese giants like NTT and KDDI was sent to Google on the expectation it would be treated as transit.

The outage in Japan only lasted a couple of hours but was so severe that the country’s Internal Affairs and Communications ministries want carriers to report on what went wrong.

An example, this time from November 2017

For a little more than 90 minutes, Internet service for millions of users in the U.S. and around the world slowed to a crawl. The cause was yet another BGP routing leak, a router misconfiguration directing Internet traffic from its intended path to somewhere else.

"Human factors are responsible for 50% to 80% of network outages"

Juniper Networks, What’s Behind Network Downtime?, 2008

Ironically, this means that the Internet works better during the week-ends...
The world of BGP policies is rapidly changing

ISPs are now eyeballs talking to content networks
e.g., Swisscom and Netflix/Spotify/YouTube

Transit becomes less important and less profitable
traffic move more and more to interconnection points

No systematic practices, yet
details of peering arrangements are private anyway