Last week on Communication Networks

BGP sessions come in two flavors

- **eBGP** sessions are used to learn routes to external destinations
- **iBGP** sessions connect the routers in the same AS

**BGP** is the routing protocol “glueing” the Internet together

- **eBGP** sessions connect border routers in different ASes
- **iBGP** sessions connect the routers in the same AS
BGP needs to solve three key challenges: scalability, privacy and policy enforcement

There is a huge # of networks and prefixes
1M prefixes, >70,000 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

On the wire, BGP is a rather simple protocol composed of four basic messages

<table>
<thead>
<tr>
<th>type</th>
<th>used to</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPEN</td>
<td>establish TCP-based BGP sessions</td>
</tr>
<tr>
<td>NOTIFICATION</td>
<td>report unusual conditions</td>
</tr>
<tr>
<td>UPDATE</td>
<td>inform neighbor of a new best route</td>
</tr>
<tr>
<td></td>
<td>a change in the best route</td>
</tr>
<tr>
<td></td>
<td>the removal of the best route</td>
</tr>
<tr>
<td>KEEPALIVE</td>
<td>inform neighbor that the connection is alive</td>
</tr>
</tbody>
</table>

Attributes Usage

- **NEXT-HOP**: egress point identification
- **AS-PATH**: loop avoidance, outbound traffic control, inbound traffic control
- **LOCAL-PREF**: outbound traffic control
- **MED**: inbound traffic control

Border Gateway Protocol
policies and more

This week on Communication Networks

Follow the Money
BGP Policies
Protocol
How does it work?

Problems
- security, performance, ...

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>

Communication Networks | Mon 17 April 2023
Unlike 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 absent from the BGP specification. 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

BGP (lack of) security

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 into the router and reconfigures

- Getting other ASes to believe the 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

- 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

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

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

- **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 it 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?

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

Proposed Enhancements to BGP

- **Address attestations**
  - Claim the right to originate a prefix
    - Signed and distributed out-of-band
    - Checked through delegation chain from ICANN
- **Route attestations**
  - Distributed as an attribute in BGP update message
    - Signed by each AS as route traverses the network
- **S-BGP can validate**
  - AS path indicates the order ASes were traversed
    - No intermediate ASes were added or removed

Secure BGP

- **Origin Authentication + cryptographic signatures**
  - Anyone who knows v’s public key can verify that the message was sent by v.

S-BGP Secure Version of BGP

- **Address attestations**
  - Claim the right to originate a prefix
  - Signed and distributed out-of-band
  - Checked through delegation chain from ICANN
- **Route attestations**
  - Distributed as an attribute in BGP update message
  - Signed by each AS as route traverses the network
- **S-BGP can validate**
  - AS path indicates the order ASes were traversed
  - No intermediate ASes were added or removed

S-BGP Deployment Challenges

- **Complete, accurate registries of prefix “owner”**
- **Public Key Infrastructure**
  - To know the public key for any given AS
- **Cryptographic operations**
  - E.g., digital signatures on BGP messages
- **Need to perform operations quickly**
  - To avoid delaying response to routing changes
- **Difficulty of incremental deployment**
  - Hard to have a “flag day” to deploy S-BGP
BGP Security Today

- **Resource Public Key Infrastructure (RPKI)**
  - A framework to support improved BGP security:
    1. A secure way to map AS numbers to IP prefixes.
    2. A distributed repository system for storing and disseminating the mappings.

- **RPKI operations**
  - RPKI relies on cryptographic certificates (X.509)
  - The certificate infrastructure mimics the way IP prefixes are distributed: from IANA, to Regional Internet Registries (RIR), to end-customers.
  - A Route Origination Authorization (ROA) states which AS is authorised to originate certain IP prefixes.

With arbitrary policies, BGP may have multiple stable states

### Problems
- Reachability
- Security
- Convergence
- Performance
- Anomalies
- Relevance

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

The actual assignment depends on the ordering between the messages

With arbitrary policies, BGP may fail to converge

---

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

... this is not always possible

* https://www.nanog.org/meetings/nanog31/presentations/griffin.pdf
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?"

Logic gates

Easy, you build a computer using 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

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

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

Check our paper for more details
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 path selection is mostly economical, not based on accurate performance criteria

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

BGP configuration is hard to get right

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 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.

In August 2017
Someone in Google fat-thumbed a Border Gateway Protocol (BGP) advertisement and sent Japanese Internet traffic into a black hole.

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. 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.

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.

Ironically, this means that the Internet works better during the week-ends...

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

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