Securing against route leaks

Amreesh Phokeer

AFRINIC
The Internet Numbers Registry for Africa
Routing Leak briefly takes down Google

Route leak cuts access to Amazon Web Services

BGP bungle downs big-name clients.

Australia's internet hit hard by massive Malaysian route leak

Telekom Malaysia apologises for BGP bungle.
Definition by RFC 7908

“A route leak is the propagation of routing announcement(s) beyond their intended scope. That is, an announcement from an Autonomous System (AS) of a learned BGP route to another AS is in violation of the intended policies of the receiver, the sender, and/or one of the ASes along the preceding AS path.”
Characteristics

- Mostly accidental (99%) but hard to detect (1% intentional)
- Last hours or sometimes days
- 6 different types (RFC 7908)
- Deceives RPKI OV
- Impact of route leaks
  - redirection of traffic through unintended path
  - Eavesdropping, traffic analysis
  - Black holing, MITM, DoS, Network overload
- Solution space being discussed at the IETF (IDR WG)
- Will probably need to modify the BGP protocol
- Other mitigation techniques and best practices
Type 1: Hairpin Turn with Full Prefix

- A multi-homed AS learns a route from one upstream ISP
- It propagates it to another upstream ISP
- Neither the prefix nor the AS_PATH is modified
- Similar to Kapela-Pilosov path-poisoning attack, with full prefix
- Usually accidental
- Poor egress filtering
- Why is it successful? Receiving ISP prefers customer announcement over peer announcement
- Result in traffic redirection or blocked at offending AS due to network overload
Type 1: Hairpin route leak

ISP 1
AS 1
Prefix P

ISP 2
AS 2
Route leak (P) propagates

Route leak (P) to upstream 2

Customer

Prefix P update

Route update P
Type 1: Example Telekom Malaysia (June 2015)

- Telekom Malaysia (AS4788) leaked 179,000 prefixes to Level3 (AS3549)
- Level3 accepted and propagated to its peers and customers
- Result: Traffic redirection through Telekom Malaysia
- Significant performance issue, slow Internet

BGP Update messages
BGP Update messages
Type 2: Lateral ISP-ISP-ISP Leak

- Peer to peer agreement
- No-transit
- ISP B will receive a route from ISP A and will leak to ISP C
- Usually reported by proper BGP monitoring by finding 3 large ISP’s ASNs in (1) AS_PATH
- Normally, large ISPs do not buy transit from each other
Type 3: Leak of Transit-Provider Prefixes to Peer

- Offending AS leaks route learned from transit provider to peer
- By principle Customer AS should only announce its own network to its peers
Type 4: Leak of Peer Prefixes to Transit Provider

- Offending AS leaks route learned from laterals (peer) to its own transit provider (upstream)
- Examples:
  - Axcelx-Hibernia route leak of Amazon AWS prefixes causing disruption of AWS platform
Type 4: Example Hathway-Airtel (March 2015)

- Google-Hathway (Peers)
- Airtel is transit provider for Hathway
- Customer networks preferred over peers
Type 5: Prefix Re-Origination with Data Path to Legitimate Origin

- Multi-homed AS learns route from upstream A and announces prefix to upstream B
- Offending AS strips the received AS_PATH and re-originate's the prefix
- Maintains (or not) original path to legitimate destination via offending AS
- Examples:
  - ChinaTelecom 2010
  - Indosat route leak 2014
Type 6: Accidental Leak of Internal Prefixes and More Specific Prefixes

ISP A

Prefix A

Route update A

ISP B

Leaks internal routes and more specifics to other ISPs

Customer

Leaks route and more specifics to other peers

Peer

Peer
# What RPKI and BGPSec can do?

<table>
<thead>
<tr>
<th>Type of Route leak</th>
<th>Current State of Detection Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 1: Hairpin Turn with Full Prefix</td>
<td>Neither OV nor BGPSec</td>
</tr>
<tr>
<td>Type 2: Lateral ISP-ISP-ISP</td>
<td>Neither OV nor BGPSec</td>
</tr>
<tr>
<td>Type 3: Leak of Transit-Provider Prefixes to Peer</td>
<td>Neither OV nor BGPSec</td>
</tr>
<tr>
<td>Type 4: Leak of Peer Prefixes to Transit Provider</td>
<td>Neither OV nor BGPSec</td>
</tr>
<tr>
<td>Type 5: Prefix re-origination with Data Path to legitimate Origin</td>
<td>OV detects</td>
</tr>
</tbody>
</table>
| Type 6: Accidental leak of Internal Prefixes and more specifics                    | • Internal prefixes not meant to be routed: OV can detect (no covering ROA or AS0 ROA)  
• ROA maxlen for more specifics                                                  |
New developments (RLP)

Solution Steps

Today: Current BGP (without route leak solution; assuming prefix filters aren’t doing job adequately) Vulnerable to accidental (99%) and malicious (1%) route leaks

Step 1: BGP with proposed route leak solution (with RPKI/OV but without BGPssec) Detects/mitigates accidental (99%) but not malicious (1%)

Step 2: BGP with proposed route leak solution (with RPKI/OV and BGPssec) Detects/mitigates accidental (99%) as well as malicious (1%)

Two drafts on route leaks mitigation techniques
For a majority of networks, the chance that your have one of those Tier 1 as transit customer is relatively low =>>> filter them out deny any announcement with AS_PATH containing big network ASes
[Method] Use of BGP communities

- Route leak type 4: Networks receiving routes from a peer should not leak those routes to other peers
- Tag routes with BGP communities on ingress, execute on egress (c.f. NANOG)

Outbound filters

\[
\text{IF} \\
\text{Prefixes contains proper BGP communities} \Rightarrow \text{ALLOW} \\
\text{ELSE} \\
\text{DENY} \\
\text{FI}
\]
[Method] Apply whitelist of prefix

- Allow only known customers to send routes to you
- Quite challenging: policies change frequently based on new business relations
- Difficult to implement programmatically, make use of filter generator to help you out (e.g. bgpq3, rpsltool, etc)
- But make sure not to leak out the whitelist of all your customers itself
[Method] Use maximum prefix settings

- Very easy!!!
- Use max_prefix filters on all your ingress and egress announcements (customers and peers)
- If unsure talk to your peers

neighbor 10.1.1.1 maximum-prefix 3000

!--- Drops the peering to 10.1.1.1 when !--- more than 3000 prefixes are received.
[Method] NTT Peerlocking mechanism

https://www.slideshare.net/InternetSociety/peering-and-transit-tutorials-practical-every-day-bgp-filtering

Relationship between providers (we know that):

- PCCW is not the upstream of AT&T
- AT&T is not the upstream of PCCW

So an AS_PATH: **2914_3491_7018** is a mistake

**X_PCCW_ATT**

Peerlocking extract and compile those relationship to create filters
Peerlock example

- Green is a bgp partner
- Blue is peerlocked
- Question: ask blue network
- So if you receive routes from PEER D or E, C, you would block those routes.
- APPLY on all eBGP sessions (customer and peering)

https://www.slideshare.net/InternetSociety/peering-and-transit-tutorials-practical-every-day-bgp-filtering
[Method] Bogon ASN

- AS 23456 => if your BGP speaker is 4-byte ASN compatible, DENY AS_PATH with AS23456
- DENY any private/reserved ASN
  - 0
  - 23456
  - 64496-131071
  - 4200000000 - 4294967295
- Buggy software implementation or misconfiguration
[Method] Bogon prefixes

10.0.0.0/8 is a bogon prefix (RFC1918).

Announced By

<table>
<thead>
<tr>
<th>Origin AS</th>
<th>Announcement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS38726</td>
<td>10.0.0.0/8</td>
<td>Internet Assigned Numbers Authority</td>
</tr>
</tbody>
</table>

Updated 15 Sep 2016 08:10 PST © 2016 Hurricane Electric
[Method] Effective Filtering (recap)

- Use maximum prefix settings
- Reject Fullbogon/Bogon prefixes (RFC1918, etc)
- Reject Bogon ASNs (AS0/AS23456, etc)
- Reject IXP LAN prefixes (i.e. the subnets)
- Build a whitelist of customers filter from IRR
- Use BGP communities to mark routes
- Reject AS_PATH based on “bignetworks” filter
- Prevent route leak using Peerlock filter
- Only accept VALIDATED RPKI routes
- Only accept Origin-AS registered in AS-SET in IRR
Protect yourself and your customers

- Always check the AS_PATH on routes coming from your peers
- Apply strict filter from your customers using prefix lists
- Use tools such as RIPE Stats, BGP looking glasses to see whether your AS is not misbehaving
- IRRToolset can be very handy to automate BGP filters from routing registries
- Create and maintain Bogon filters and reject RFC1918 announcements, bogon ASNs
- Use mailing lists and contact misbehaving ISPs
Mutually Agreed Norms for Routing Security (MANRS)

1. Prevent propagation of Incorrect routing information
2. Prevent traffic with spoofed source IP addresses
3. Facilitate global operational communication and coordination between network operators
4. Facilitate validation of routing information at global scale
References

MANRS - https://www.routingmanifesto.org/manrs/

Practical Everyday BGP Filtering by Job Snijders - http://www.slideshare.net/InternetSociety/peering-and-transit-tutorials-practical-every-day-bgp-filtering

Problem definition and classification of route leaks - RFC7908


BGPQ3 - https://github.com/snar/bgpq3
Questions?

amreesh@afrinic.net

irr@afrinic.net

rpki-help@afrinic.net
Resource Certification (RPKI)
@AFRINIC

Amreesh Phokeer
Agenda

- What is the rationale behind RPKI?
- What is resource certification?
- How to get your resources certified?
- How to sign your routing announcements?
- How to make your router talk RPKI?
- How to build filters based on validated routes?
- Demo
- Current hot topics
Rationale

- Routing mostly based on trust
- BGP offer amazing possibilities but poor security
- No systematic way to filter peers and customers
- Unreliable sources of policy information
- The Internet is full of stories of:
  - Route leaks you said?
  - BGP hijacks
  - Traffic redirection
  - Blackholing
Mitigation techniques

● Conservative filtering

● Use of Routing Registries but:
  ○ Do not have all routing information
  ○ Do not necessarily mirror each other
  ○ Routing policies not kept up-to-date
  ○ Error-prone
Youtube and Pakistan Telecom (2008)

208.65.152.0/22

The Internet

208.65.152.0/24

Pakistan Telecom
Timeline

- **AS 17557 (Pakistan Telecom)** starts announcing 208.65.153.0/24 on **Sunday, 24 February, 2008** at **18:47 UTC**

- **AS 36561 (YouTube)** starts announcing 208.65.153.128/25 and 208.65.153.0/25 on **20:07 UTC**

- **20:18 UTC**

- **AS 3491 (PCCW Global)** withdraws all prefixes originated by AS17557 (Pakistan Telecom) on **20:51 UTC**

- **21:01 UTC**

- **20:07 UTC**

- **AS 36561 (YouTube)** starts announcing 208.65.153.0/24 on **20:07 UTC**

- **20:18 UTC**

- **Hijacked /24 originated by AS 17557 (Pakistan Telecom) and AS3491 (PCCW Global)** on **21:01 UTC**
China Telecom - Traffic redirection

Fraudulently announced 203.190.56.0/21

AT&T

peers

China Telecom

peers

LEVEL3

Announces AS PATH LEVEL3, Verizon, Verizon-Wireless to 203.190.56.0/21

AS

Verizon

customer

203.190.56.0/21

AS

Verizon Wireless
How do we securing Internet Routing?

<table>
<thead>
<tr>
<th>Network</th>
<th>Next Hop</th>
<th>Metric</th>
<th>LocPrf</th>
<th>Weight</th>
<th>Path</th>
</tr>
</thead>
<tbody>
<tr>
<td>192.0.2.0/24</td>
<td>1.1.1.1</td>
<td>20</td>
<td></td>
<td>0</td>
<td>1010 1011 286 4040 i</td>
</tr>
<tr>
<td></td>
<td>2.2.2.2</td>
<td>20</td>
<td></td>
<td>0</td>
<td>2020 1011 4040 i</td>
</tr>
<tr>
<td></td>
<td>3.3.3.3</td>
<td>10</td>
<td>100</td>
<td>0</td>
<td>2020 702 4040 i</td>
</tr>
<tr>
<td></td>
<td>4.4.4.4</td>
<td>0</td>
<td>90</td>
<td>0</td>
<td>4040 i</td>
</tr>
<tr>
<td></td>
<td>5.5.5.5</td>
<td>2659</td>
<td></td>
<td>0</td>
<td>5050 1011 4040 i</td>
</tr>
<tr>
<td></td>
<td>6.6.6.6</td>
<td></td>
<td>80</td>
<td>0</td>
<td>6060 4040 i</td>
</tr>
<tr>
<td></td>
<td>7.7.7.7</td>
<td>10</td>
<td>100</td>
<td>0</td>
<td>7070 3356 3356 4040 i</td>
</tr>
</tbody>
</table>
How do we securing Internet Routing?

<table>
<thead>
<tr>
<th>Network</th>
<th>Next Hop</th>
<th>Metric</th>
<th>LocPrf</th>
<th>Weight</th>
<th>Path</th>
</tr>
</thead>
<tbody>
<tr>
<td>192.0.2.0/24</td>
<td>1.1.1.1</td>
<td>20</td>
<td></td>
<td>0</td>
<td>1010 1011 286 4040 i</td>
</tr>
<tr>
<td></td>
<td>2.2.2.2</td>
<td>20</td>
<td></td>
<td>0</td>
<td>2020 1011 4040 i</td>
</tr>
<tr>
<td></td>
<td>3.3.3.3</td>
<td>10</td>
<td>100</td>
<td>0</td>
<td>2020 702 4040 i</td>
</tr>
<tr>
<td></td>
<td>4.4.4.4</td>
<td>0</td>
<td>90</td>
<td>0</td>
<td>4040 i</td>
</tr>
<tr>
<td></td>
<td>5.5.5.5</td>
<td>2659</td>
<td></td>
<td>0</td>
<td>5050 1011 4040 i</td>
</tr>
<tr>
<td></td>
<td>6.6.6.6</td>
<td></td>
<td>80</td>
<td>0</td>
<td>6060 4040 i</td>
</tr>
<tr>
<td></td>
<td>7.7.7.7</td>
<td>10</td>
<td>100</td>
<td>0</td>
<td>7070 3356 3356 4040 i</td>
</tr>
</tbody>
</table>
How do we securing Internet Routing?

<table>
<thead>
<tr>
<th>Network</th>
<th>Next Hop</th>
<th>Metric</th>
<th>LocPrf</th>
<th>Weight</th>
<th>Path</th>
</tr>
</thead>
<tbody>
<tr>
<td>192.0.2.0/24</td>
<td>1.1.1.1</td>
<td>20</td>
<td></td>
<td>0</td>
<td>1010 1011 286 4040 i</td>
</tr>
<tr>
<td></td>
<td>2.2.2.2</td>
<td>20</td>
<td></td>
<td>0</td>
<td>2020 1011 4040 i</td>
</tr>
<tr>
<td>3.3.3.3</td>
<td></td>
<td>10</td>
<td>100</td>
<td>0</td>
<td>2020 702 4040 i</td>
</tr>
<tr>
<td>4.4.4.4</td>
<td></td>
<td>0</td>
<td>90</td>
<td>0</td>
<td>4040 i</td>
</tr>
<tr>
<td>5.5.5.5</td>
<td></td>
<td>2659</td>
<td></td>
<td>0</td>
<td>5050 1011 4040 i</td>
</tr>
<tr>
<td>6.6.6.6</td>
<td></td>
<td></td>
<td>80</td>
<td>0</td>
<td>6060 4040 i</td>
</tr>
<tr>
<td>7.7.7.7</td>
<td></td>
<td>10</td>
<td>100</td>
<td>0</td>
<td>7070 3356 3356 4040 i</td>
</tr>
</tbody>
</table>

Prefix origination
AS_PATH

You need to secure both!!!
Solution - RPKI

- SIDR group at the IETF
  - How to securely verify that an AS is authorised to announce a prefix? *(Origin Validation)*
  - How to make sure that the AS_PATH has not been modified? *(BGPSEC)*

- Origin validation
  - RFC 5280: X.509 Public Key Infrastructure
  - RFC 3779: Extensions for IP addresses and ASN

- BGPSEC (still on-going)
Role of the RIR

- Receives global allocations from IANA
- Distribute and manage resources at a regional level
- Make sure information are up-to-date and accurate
- Becomes the de-facto authority as sole registry regionally
Role of the RIR

- Receives global allocations from IANA
- Distribute and manage resources at a regional level
- Make sure information are up-to-date and accurate
- Becomes the de-facto authority as sole registry regionally
Resource Certificates

- RPKI defines two types of certificates:
  - CA - Certificate Authority (to issue CA or EE)
  - EE - End-entity (digital signature, etc)

- Certify resources - verifiable ownership!

- AFRINIC has a self-signed root certificate

- IANA one-day!

- Opt-in service, one year validity

- Exclude legacy space/members
Certificate hierarchy

Issuer: AFRINIC
Subject: ISP1
Resources: 192.2.0.0/16
Pub Key Info: <ISP1-key>
Signed by: <root-key-priv>

Issuer: ISP1
Subject: ISP2
Resources: 192.2.200.0/22
Pub Key Info: <isp2-key>
Signed by: <afринic-key-priv>

Issuer: ISP2
Subject: ISP2-EE
Resources: 192.2.200.0/24
Pub Key Info: <isp2-ee-key>
Signed by: <isp2-key-priv>

ROA
AS 3
192.2.200.0/24-24
Repositories

- Public
- Certificates and ROA
- CRL and MFT
- Hosted or delegated
- HOSTED MODE only
AFRINIC’s Repository

Index of /repository

<table>
<thead>
<tr>
<th>Name</th>
<th>Last modified</th>
<th>Size</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent Directory</td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>04E8B0D80F4D11E0B657..&gt;</td>
<td>2016-04-25 07:40</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>AfriNIC.cer</td>
<td>2015-05-25 11:32</td>
<td>4.1K</td>
<td></td>
</tr>
<tr>
<td>afrinic/</td>
<td>2016-09-18 20:00</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>apnic/</td>
<td>2016-09-18 20:05</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>arin/</td>
<td>2016-09-18 20:10</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>lacnic/</td>
<td>2016-09-18 20:15</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>member_repository/</td>
<td>2016-09-18 20:36</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>ripe/</td>
<td>2016-09-18 20:20</td>
<td></td>
<td>-</td>
</tr>
</tbody>
</table>

AFRINIC Root Certificate
AFRINIC member’s repository
Demo

https://my.afrinic.net
Validated caches

AFRINIC Validator Host: validator.afrinic.net Port: 8080
Configure your router

```
router bgp 12345
...
bgp rpki server tcp 192.168.179.3 port 43779 refresh 60
bgp rpki server tcp 147.28.0.84 port 93920 refresh 60
```
Route announcement status

- **Valid** – A matching/covering prefix was found with a matching AS number
- **Invalid** – A covering prefix was found, but the AS number did not match, and there was no other matching one
- **NotFound** – No matching or covering prefix was found, same as today
You define your own policy

**Fairly Secure**

route-map validity-0
  match rpki valid
  set local-preference 100
route-map validity-1
  match rpki not-found
  set local-preference 50
! invalid is dropped

**Paranoid**

route-map validity-0
  match rpki valid
  set local-preference 110
! everything else dropped

**Security Geek**

route-map validity-0
  match rpki invalid
  set local-preference 110
! everything else dropped
RPKI tools

● Validators:
  ○ RIPE Validator
  ○ Rcynic - [www.rpki.net](http://www.rpki.net) (CA+Validator)
  ○ RPSTIR

● Looking glasses:
  ○ bgp.he.net
  ○ Bgpmon
  ○ RIPEStat
RPKI Hot topics

- Global trust anchor
  - ICANN/IANA/NRO
  - Support from local RIR community
- RPKI Adverse actions
- RPKI Validation considered
  - Transfers of resources
  - ERX spaces
Questions

rpki-help@afrinic.net