Securing Internet Routing

RPKI & Route Origin Validation

Tashi Phuntsho (tashi@apnic.net)
Senior Network Janitor/Technical Trainer
Recent - Fat-finger/Hijacks/Leaks

- BGP Optimizers impact Internet – June 2019

  - Most CF (AS13335) hosted sites were not reachable during the leak
    - About 15% of their global traffic!!
    - ~ 120mins

On Mon, Jun 24, 2019 at 3:57 AM [Name Redacted] wrote:
Hello are there any issues with CloudFlare services now?

https://twitter.com/atoonk/status/1143143943531454464/photo/1
Recent - Fat-finger/Hijacks/Leaks

- BGP Optimizers impact Internet (contd...)
  - How and What happened?

BGP Optimizers (Was: Validating possible BGP MITM attack)

From: Job Snijders <job () mtt.net>
Date: Thu, 31 Aug 2017 22:06:49 +0200

Dear all,

disclaimer:

[ The following is targeted at the context where a BGP optimizer generates BGP announcement that are ordinarily not seen in the Default-Free Zone. The OP indicated they announce a /23, and were unpleasantly surprised to see two unauthorized announcements for /24 more-specifics pop up in their alerting system. No permission was granted to create and announce these more-specifics. The AS_PATH for those /24 announcements was entirely fabricated. Original thread https://mailman.nanog.org/pipermail/nanog/2017-August/092124.html ]

On Thu, Aug 31, 2017 at 11:13:18AM -0700, Andy Litzinger wrote:

Presuming it was a route optimizer and the issue was ongoing, what would be the suggested course of action?

I strongly recommend to turn off those BGP optimizers, glue the ports shut, burn the hardware, and salt the grounds on which the BGP optimizer sales people walked.

Recent - Fat-finger/Hijacks/Leaks

- **Google prefix leaks – Nov 2018**
  - Google services (G-Suite, Google search and Google analytics) affected by the leak
    - Traffic dropped at AS4809 (China Telecom)
    - ~ 74mins
Recent - Fat-finger/Hijacks/Leaks

- Google prefix leaks (contd...)
  - How did it happen?
    - AS37282 (MainOne) leaked Google prefixes to AS4809 (CT) at IXPN, who leaked it to other transit providers like AS20485 (TransTelecom)

https://blog.thousandeyes.com/internet-vulnerability-takes-down-google/
Recent - Fat-finger/Hijacks/Leaks

• Amazon (AS16509) Route53 hijack – April 2018
  - AS10279 (eNET) originated more specifics (/24s) of Amazon Route53’s prefix (205.251.192.0/21)
    205.251.192.0/24 ...... 205.251.199.0/24
    https://ip-ranges.amazonaws.com/ip-ranges.json
  - Its peers, like AS6939 (HE), shared these routes with 100s of their own peers...
  - The motive?
    - During the period, DNS servers in the hijacked range only responded to queries for myetherwallet.com
    - Responded with addresses associated with AS41995/AS48693
Recent - Fat-finger/Hijacks/Leaks

- Route53 hijack (contd...)
  - Resolvers querying any Route53 managed names, would ask the authoritative servers controlled through the BGP hijack
    - Possibly, used an automated cert issuer to get a cert for myetherwallet.com
  - use _THEIR_ crypto to end-users to see everything (including passwords)

https://blog.cloudflare.com/bgp-leaks-and-crypto-currencies
Recent - Fat-finger/Hijacks/Leaks

• Bharti (AS9498) originates 103.0.0.0/10 - **Dec 2017**
  - ~2 days
  - No damage done—more than 8K specific routes!

• Google brings down Internet in Japan - **Aug 2017**
  - ~ 24 hours)
  - Google (AS15169) leaked **>130K** prefixes to Verizon (AS701) – in Chicago
    - Normally ~ 50 prefixes
    - ~25K of those were NTT OCN’s (AS4713) more specifics
    - which was leaked onwards to KDDI and IIJ (and accepted)
  - Everyone who received the leaked more specifics, preferred the Verizon-Google path to reach NTT OCN!
Recent - Fat-finger/Hijacks/Leaks

- Google leak (contd...)

![Before leak (JP->JP)](image1)

![After leak (JP->JP)](image2)

![After leak (EU->EU)](image3)

Fat-finger/Hijacks/Leaks

- YouTube (AS36561) Incident - **Feb 2008**
  - ~ 2 hours
  - AS17557 (PT) announced 208.65.153.0/24 (208.65.152.0/22)
    - Propagated by AS3491 (PCCW)
Why do we keep seeing these?

• Because NO ONE is in charge?
  - No single authority model for the Internet
  - No reference point for what’s right in routing
Why do we keep seeing these?

- Routing works by RUMOUR
  - Tell what you know to your neighbors, and Learn what your neighbors know
  - Assume everyone is correct (and *honest*)
    - Is the originating network the rightful owner?
Why do we keep seeing these?

• Routing is VARIABLE
  □ The view of the network depends on where you are
    ▪ Different routing outcomes at different locations

□ ~ no reference view to compare the local view 😞
Why do we keep seeing these?

- Routing works in REVERSE
  - Outbound advertisement affects inbound traffic
  - Inbound *(Accepted)* advertisement influence outbound traffic
Why do we keep seeing these?

• As always, there is no E-bit (evil!)
  - A bad routing update does not identify itself as BAD
  - All we can do is identify GOOD updates
  - But how do we identify what is GOOD???
Why should we worry?

- Because it’s just so easy to do bad in routing!
How do we address these?

- **Filtering!**
  - Filters with your peers, upstream(s) and customers
    - Prefix filters
    - Prefix limit
    - AS-PATH filters
    - AS-PATH limit
    - RFC 8212 – BGP default reject or something similar
Current practice

1. Peering/Transit Request
2. LOA Check
3. Filters (in/out)
Tools & Techniques

LOA Check

- Whois (manual)
- Letter of Authority
- IRR (RPSL)
Tools & Techniques

- Look up **whois**
  - verify holder of a resource

```
tashi@tashi -> whois -h whois.apnic.net 202.125.96.0
% [whois.apnic.net]
% Whois data copyright terms http://www.apnic.net/db/dbcopyright.html
% Information related to '202.125.96.0 - 202.125.96.255'
% Abuse contact for '202.125.96.0 - 202.125.96.255' is 'training@apnic.net'

inetnum: 202.125.96.0 - 202.125.96.255
netname: APNICTRAINING-AP
descr: Prefix for APNICTRAINING LAB DC
country: AU
admin-c: AT480-AP
tech-c: AT480-AP
status: ALLOCATED NON-PORTABLE
mnt-by: MAINT-AU-APNICTRAINING
mnt-irt: IRT-APNICTRAINING-AU
last-modified: 2016-06-17T00:17:28Z
source: APNIC

irt: IRT-APNICTRAINING-AU
address: 6 Cordelia Street
address: South Brisbane
address: QLD 4101
e-mail: training@apnic.net
abuse-mailbox: training@apnic.net
admin-c: AT480-AP
tech-c: AT480-AP
auth: # Filtered
mnt-by: MAINT-AU-APNICTRAINING
last-modified: 2013-10-31T11:01:10Z
source: APNIC

```

role: APNIC Training
address: 6 Cordelia Street
address: South Brisbane
address: QLD 4101
country: AU
phone: +61 7 3858 3100
fax-no: +61 7 3858 3199
e-mail: training@apnic.net
admin-c: JW3997-AP
tech-c: JW3997-AP
nic-hdl: AT480-AP
mnt-by: MAINT-AU-APNICTRAINING
last-modified: 2017-08-22T04:59:14Z
source: APNIC

% Information related to '202.125.96.0/24AS131107'

role: APNIC Training
address: 6 Cordelia Street
address: South Brisbane
address: QLD 4101
country: AU
phone: +61 7 3858 3100
fax-no: +61 7 3858 3199
e-mail: training@apnic.net
admin-c: JW3997-AP
tech-c: JW3997-AP
nic-hdl: AT480-AP
mnt-by: MAINT-AU-APNICTRAINING
last-modified: 2016-06-16T23:23:00Z
source: APNIC
Tools & Techniques

• Ask for a **Letter of Authority**
  - Absolve from any liabilities
Tools & Techniques

- Look up (or ask to enter) details in **internet routing registries** (IRR)
  - describes route origination and inter-AS routing policies

```
toshi@toshi:~$ whois -h whois.road.net AS17660
out-num:  AS17660
as-name:  BT-Bhutan
descr:    Divinetworks for BT
admin-c:  DUMMY-RIPE
tech-c:   DUMMY-RIPE
status:   OTHER
mnt-by:   YP67641-MNT
mnt-by:   ES6436-RIPE
created:  2012-11-29T10:31:33Z
last-modified:  2018-09-04T15:26:24Z
source:   RIPE-NONAUTH
remarks:  **************************************************************************
 remarks:  * THIS OBJECT IS MODIFIED
 remarks:  * Please note that all data that is generally regarded as personal
 remarks:  * data has been removed from this object.
 remarks:  * To view the original object, please query the RIPE Database at:
 remarks:  * http://www.ripe.net/whois
 remarks:  **************************************************************************

aut-num:  AS17660
as-name:  DRUKNET-AS
descr:    DrukNet ISP
descr:    Bhutan Telecom
descr:    Thimphu
country:  BT
org:      ORG-BTL2-AP
import:   action pref=100; accept ANY
export:   to AS6461 announce AS-DRUKNET-TRANSIT
import:   action pref=150; accept ANY
export:   to AS2914 announce AS-DRUKNET-TRANSIT
import:   action pref=100; accept ANY
export:   to AS9453 announce AS-DRUKNET-TRANSIT

route:    61.45.248.0/24
origin:   AS17660
mnt-by:   MAINT-A54826
changed:  noc@vocus.com.au  20160702
source:   RADB
```
Tools & Techniques

- IRR
  - Helps auto generate network (prefix/as-path) filters using RPSL tools
    - Filter out route advertisements not described in the registry
Tools & Techniques

• Problem(s) with IRR
  - No single authority model
    - How do I know if a RR entry is genuine and correct?
    - How do I differentiate between a current and a lapsed entry?
  - Many RRs
    - If two RRs contain conflicting data, which one do I trust and use?
  - Incomplete data - Not all resources are registered in an IRR
    - If a route is not in a RR, is the route invalid or is the RR just missing data?
  - Scaling
    - How do I apply IRR filters to upstream(s)?
Tools & Techniques

- Automating network filters (IRR filters) - **Caution**
  
  - IRR filters only as good as the correctness of the IRR entries
    - Might require manual overrides and offline verification of resource holders
    - Good idea to use specific sources (`-S` in `bgpq3`, `-s` in `rtconfig`) when generating filters, assuming mirrors are up to date
Back to basics – identify GOOD

- Could we use a digital signature to convey the “authority to use”?
  - Using a private key to sign the authority, and
  - the public key to validate the authority

- The idea being:
  - If the holder of the resource has the private key, it can sign/authorize the use of the resource
How about trust?

- How do we build a chain of trust in this framework?
  - Follow the resource allocation/delegation hierarchy
    - IANA → RIRs → NIRs/LIRs → End Holders
    - |  
    - V  
    - End Holders
  - To describe the address allocation using digital certificates
RPKI Chain of Trust

Allocation Hierarchy

ARIN
AFRINIC
APNIC
LACNIC
RIPE-NCC

Trust Anchor Certificate

IANA
Cert (CA)

NIR
Cert (CA)

ISP
Cert (EE)

Certificate chain mirrors the allocation hierarchy
RPKI Chain of Trust

- RIRs hold a self-signed root certificate for all the resources they have in the registry
  - they are the Trust Anchor for the system

- The root certificate signs the resource certificates for end-holder allocations
  - binds the resources to the end-holders public key

- Any attestations signed by the end-holder’s private key, can now be validated up the chain of trust
**X.509 Certificates recap (RFC5280)**

- Associates a public key with an individual or an organization

<table>
<thead>
<tr>
<th>FIELD</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>VERSION</td>
<td>Version of X.509</td>
</tr>
<tr>
<td>SERIAL NUMBER</td>
<td>Uniquely identifies the certificate</td>
</tr>
<tr>
<td>SIGNATURE ALGORITHM</td>
<td>Algorithms used by the CA to sign the cert</td>
</tr>
<tr>
<td>ISSUER NAME</td>
<td>Id of the CA (that issued the cert)</td>
</tr>
<tr>
<td>VALIDITY PERIOD</td>
<td>Cert validity</td>
</tr>
<tr>
<td>SUBJECT NAME</td>
<td>Entity associated with the public key</td>
</tr>
<tr>
<td>SUBJECT PUBLIC KEY</td>
<td>Owner’s public key</td>
</tr>
<tr>
<td>EXTENSIONS (ISSUER KEY ID)</td>
<td>Identify the pub key of issuer of the cert</td>
</tr>
<tr>
<td>EXTENSIONS (SUBJECT KEY ID)</td>
<td>Extra info (owner of the cert)</td>
</tr>
<tr>
<td>EXTENSIONS (CRL)</td>
<td>Extensions (CRL)</td>
</tr>
<tr>
<td>CA DIGITAL SIGNATURE</td>
<td>Certifies the binding between the pub key &amp; subject of the cert</td>
</tr>
</tbody>
</table>
RF 3779 extensions – binds a list of resources (IPv4/IPv6, ASN) to the subject of the certificate (private key holder)

SIA (subject information access) contains a URI that identifies the publication point of the objects signed by the subject of the cert.
Resource Certificates

- When an address holder A (*IRs) allocates resources (IP address/ASN) to B (end holders)

  - A issues a public-key/resource certificate that binds the allocated address with B’s public key, all signed by A’s (CA) private key

  - The resource certificate proves the holder of the private key (B) is the legitimate holder of the number resource!
Route Origin Authorization (ROA)

- The resource holder (B) can now sign *authorities* using its private key, which can be validated by any third party against the TA.

- For routing, the address holder can *authorize* a network (ASN) to *originate* a route into the BGP routing system, and *sign* this permission with its private key (ROA).
Route Origin Authorization (ROA)

- Digitally signed object
  - list of prefixes and the nominated ASN
  - can be verified cryptographically

<table>
<thead>
<tr>
<th>Prefix</th>
<th>203.176.32.0/19</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max-length</td>
<td>/24</td>
</tr>
<tr>
<td>Origin ASN</td>
<td>AS17821</td>
</tr>
</tbody>
</table>

- **Multiple ROAs can exist for the same prefix**
What can RPKI do?

• Authoritatively proof:
  - Who is the legitimate owner of an address, and
  - Identify which ASNs have the permission from the holder to originate the address

• Hence, can help:
  - prevent route hijacks/mis-origination/misconfiguration
RPKI Components

- **Issuing Party** – Internet Registries (*IRs)
  - Certificate Authority (CA) that issues resource certificates to end-holders
  - Publishes the objects (ROAs) signed by the resource certificate holders
RPKI Components

- **Relying Party (RP)**
  - RPKI Validator tool that gathers data (ROA) from the distributed RPKI repositories
  - Validates each entry’s signature against the TA to build a "Validated cache"
RPKI Service Models

- **Hosted model:**
  - The RIR (APNIC) runs the CA functions on members’ behalf
    - Manage keys, repo, etc.
    - Generate certificates for resource delegations

- **Delegated model:**
  - Member becomes the CA (delegated by the parent CA) and operates the full RPKI system
    - JPNIC, TWNIC, CNNIC (IDNIC in progress)
Route Origin Validation (ROV)

Global (RPKI) Repository

TA

TA

TA

RSYNC/RRDP

2406:6400::/32-48
17821

ROA

2406:6400::/32-48
17821

2406:6400::/48

2406:6400::/48

17821

TA

AS17821

ASXXXX

RPKI-to-Router (RtR)

RPKI Validator/
RPKI Cache server

2406:6400::/32-48
17821
Route Origin Validation

- Router fetches ROA information from the validated RPKI cache
  - Crypto stripped by the validator

- BGP checks each received BGP update against the ROA information and labels them
Validation States

- **Valid**
  - the prefix and AS pair found in the database.

- **Invalid**
  - prefix is found, but origin AS is wrong, OR
  - the prefix length is longer than the maximum length

- **Not Found/Unknown**
  - No valid ROA found
    - Neither valid nor invalid (perhaps not created)
Validation States

### ROA

<table>
<thead>
<tr>
<th>ASN</th>
<th>Prefix</th>
<th>Max Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>65420</td>
<td>10.0.0.0/16</td>
<td>18</td>
</tr>
</tbody>
</table>

### BGP Routes

<table>
<thead>
<tr>
<th>ASN</th>
<th>Prefix</th>
<th>RPKI State</th>
</tr>
</thead>
<tbody>
<tr>
<td>65420</td>
<td>10.0.0.0/16</td>
<td>VALID</td>
</tr>
<tr>
<td>65420</td>
<td>10.0.128.0/17</td>
<td>VALID</td>
</tr>
<tr>
<td>65421</td>
<td>10.0.0.0/16</td>
<td>INVALID</td>
</tr>
<tr>
<td>65420</td>
<td>10.0.10.0/24</td>
<td>INVALID</td>
</tr>
<tr>
<td>65430</td>
<td>10.0.0.0/8</td>
<td>NOT FOUND</td>
</tr>
</tbody>
</table>
Possible actions - RPKI states

- **Do Nothing** (observe & learn)
- **Tag with BGP communities**
  - If you have downstream customers or run a route server (IXP)
    - Let them decide
  - Ex:
    - **Valid** (ASN:65XX1)
    - **Not Found** (ASN:65XX2)
    - **Invalid** (ASN:65XX3)
- **Modify preference values**
  - **RFC7115** (High, Low, Lowest)
- **Drop Invalids**
  - ~6K IPv4 routes (might want to check your top flows)

ROV – Industry trends

- **AT&T (AS7018) drops Invalids!**
  
  11 Feb 2019

**AT&T/as7018 now drops invalid prefixes from peers**

Jay Borkenhagen jayb at bracburn.org

Mon Feb 11 14:53:45 UTC 2019

- Previous message (by thread): BGP topological vs centralized route reflector
- Next message (by thread): AT&T/as7018 now drops invalid prefixes from peers
- Messages sorted by: [date] [thread] [subject] [author]

FYI:

The AT&T/as7018 network is now dropping all RPKI-invalid route announcements that we receive from our peers.

We continue to accept invalid route announcements from our customers, at least for now. We are communicating with our customers whose invalid announcements we are propagating, informing them that these routes will be accepted by fewer and fewer networks over time.

Thanks to those of you who are publishing ROAs in the RPKI. We would also like to encourage other networks to join us in taking this step to improve the quality of routing information in the internet.

Thanks!

Jay B.
• **Workonline Comms (AS37271) & SEACOM (AS37100) drops Invalids!**
  - 1 and 5 April 2019 (does not use ARIN’s TAL)

---

**[aops] RPKI ROV & Dropping of Invalids - Africa**

- **To:** aops@aops.net
- **Subject:** [aops] RPKI ROV & Dropping of Invalids - Africa
- **From:** Mark Tinka <mark.tinka@seacom.mu>
- **Date:** Tue, 9 Apr 2019 14:05:03 +0200

Hello all,

In November 2018 during the ZAPF (South Africa Peering Forum) meeting in Cape Town, 3 major ISPs in Africa announced that they would enable RPKI’s ROV (Route Origin Validation) and the dropping of Invalid routes as part of an effort to clean up the BGP Internet, on the 1st April, 2019.

On the 1st of April, Workonline Communications (AS37271) enabled ROV and the dropping of Invalid routes. This applies to all eBGP sessions for IPv4 and IPv6.

On the 5th of April, SEACOM (AS37100) enabled ROV and the dropping of Invalid routes. This applies to all eBGP sessions with public peers, private peers and transit providers, both for IPv4 and IPv6. eBGP sessions toward downstream customers will follow in 3 months from now.

We are still standing by for the 3rd ISP to complete their implementation, and we are certain they will communicate with the community accordingly.

Please note that for the legal reasons previously discussed on various fora, neither Workonline Communications nor SEACOM are utilising the ARIN TAL. As a result, any routes covered only by a ROA issued under the ARIN TAL will fall back to a status of Not Found. Unfortunately, this means that ARIN members will not see any improved routing security for their prefixes on our networks until this is resolved. We will each re-evaluate this decision if and when ARIN’s policy changes. We are hopeful that this will happen sooner rather than later.

If you interconnect with either of us and may be experiencing any routing issues potentially related to this new policy, please feel free to reach out to:

- noc@workonline.africa
- peering@seacom.mu

Workonline Communications and SEACOM hope that this move encourages the rest of the ISP community around the world to ramp up their deployment of RPKI ROV and dropping of Invalid routes, as we appreciate the work that AT&T have carried out in the same vein.

In the mean time, we are happy to answer any questions you may have about our deployments. Thanks.

Mark Tinka (SEACOM) & Ben Maddison (Workonline Communications).
Are ROAs enough?

- What if I forge the origin AS in the AS path?
  - Would be accepted as “good” – pass origin validation!

- Which means, we need to secure the AS path as well
  - Need AS path validation (per-prefix)

- We can use RPKI certificates for this
AS keys (per-router keys)

IANA
Cert (CA)

APNIC
Cert (CA)

APNIC Training

APNIC Training (CA)
202.125.96.0/24
AS45192
Public Key

Prefix EE
202.125.96.0/24
Public Key

ROA
202.125.96.0/24
AS45192

AS Cert (CA)
AS45192
Public Key

Router EE
AS45192
rtr-00
Public Key

Encodes ASN and router ID
AS path validation - BGPsec

- AS1 router crypto signs the message to AS2
- AS2 router signs the message to AS3 and AS4, encapsulating AS1’s message

- A BGPsec speaker validates the received update by checking:
  - If there is a ROA that describes the prefix and origin AS
  - If the received AS path can be validated as a chain of signatures (for each AS in the AS path) using the AS keys
So why is AS path validation NOT happening?

- Cannot have partial adoption
  - Cannot jump across non-participating networks

- More HW resources
  - CPU - high crypto overhead to validate signatures, and
  - Memory
    - Updates in BGPsec would be per prefix
    - New attributes carrying signatures and certs/key IDs for every AS in the AS path

- No clarity on how to distribute the collection of certificates required to validate the signatures

- Given so much overhead, can it prevent more than route hijacks?
  - Route leaks?
RPKI Further Reading

- RFC 5280: X.509 PKI Certificates
- RFC 3779: Extensions for IP Addresses and ASNs
- RFC 6481-6493: Resource Public Key Infrastructure
Acknowledgement

- Geoff Huston, APNIC
- Randy Bush, IIJ Labs/Arrcus
Implementation
Create & publish your ROA

- Login to MyANIC
  - Go to Resources → Route Management (see image below)

Create (publish) your ROA

• Select **Create route** (as shown below)
Create (publish) your ROA

- Example for IPv6 below
Create (publish) your ROA

Confirm route creation

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ROA</strong></td>
<td>Enabled</td>
</tr>
<tr>
<td><strong>Whois</strong></td>
<td>Disabled</td>
</tr>
<tr>
<td><strong>Prefix</strong></td>
<td>2406:6400::/32</td>
</tr>
<tr>
<td><strong>Origin AS</strong></td>
<td>45192</td>
</tr>
<tr>
<td><strong>Most specific announcement</strong></td>
<td>/48 (distance from prefix length: 16)</td>
</tr>
</tbody>
</table>

*Sub-route management is only available when the distance from the most specific announcement to the prefix length is less than 16*
Create (publish) your ROA

- Example for **IPv4**
Create (publish) your ROA

• Your ROAs are ready!
Check your ROA

http://nong.rand.apnic.net:8080/roas

Validated ROAs

Validated ROAs from **APNIC RPKI Root, ARIN, AfriNIC RPKI Root, LACNIC RPKI Root, RIPE NCC RPKI Root.**

<table>
<thead>
<tr>
<th>ASN</th>
<th>Prefix</th>
<th>Maximum Length</th>
<th>Trust Anchor</th>
</tr>
</thead>
<tbody>
<tr>
<td>135533</td>
<td>61.45.248.0/24</td>
<td>24</td>
<td>APNIC RPKI Root</td>
</tr>
</tbody>
</table>

Showing 1 to 1 of 1 entries (filtered from 83,128 total entries)
Check your ROA

# whois -h rr.ntt.net 2001:df2:ee00::/48

route6: 2001:df2:ee00::/48
descr: RPKI ROA for 2001:df2:ee00::/48
remarks: This route object represents routing data retrieved from the RPKI
remarks: The original data can be found here: https://rpki.gin.ntt.net/r/AS131107/2001:df2:ee00::/48
remarks: This route object is the result of an automated RPKI-to-IRR conversion process.
remarks: maxLength 48
origin: AS131107
mnt-by: MAINT-JOB
changed: job@ntt.net 20180802
source: RPKI # Trust Anchor: APNIC RPKI Root
Check your ROA

# whois -h whois.bgpmon.net 2001:df2:ee00::/48

Prefix: 2001:df2:ee00::/48
Prefix description: APNICTRAINING-DC
Country code: AU
Origin AS: 131107
Origin AS Name: APNICTRAINING LAB DC
RPKI status: ROA validation successful
First seen: 2016-06-30
Last seen: 2018-01-21
Seen by #peers: 97

# whois -h whois.bgpmon.net "--roa 131107 2001:df2:ee00::/48"

------------------------
ROA Details
------------------------
Origin ASN: AS131107
Not valid Before: 2016-09-07 02:10:04
Not valid After: 2020-07-30 00:00:00 Expires in 2y190d9h34m23.2000000029802s
Trust Anchor: rpki.apnic.net
Prefixes: 2001:df2:ee00::/48 (max length /48) 202.125.96.0/24 (max length /24)
Check your ROA

https://bgp.he.net/

<table>
<thead>
<tr>
<th>Announced By</th>
</tr>
</thead>
<tbody>
<tr>
<td>Origin AS</td>
</tr>
<tr>
<td>AS131107</td>
</tr>
</tbody>
</table>
Deploy RPKI Validator

- Many options:
  - RIPE RPKI Validator
  - Dragon Research Labs RPKI Toolkit
    - [https://github.com/dragonresearch/rpki.net](https://github.com/dragonresearch/rpki.net)
  - Routinator
    - [https://github.com/NLnetLabs/routinator](https://github.com/NLnetLabs/routinator)
  - OctoRPKI & GoRTR (Cloudflare’s RPKI toolkit)
    - [https://github.com/cloudflare/cfrpki](https://github.com/cloudflare/cfrpki)
• Policies based on validation:

```
route-map ROUTE-VALIDATION permit 10
  match rpki valid
  set local-preference 110
!
route-map ROUTE-VALIDATION permit 20
  match rpki not-found
  set local-preference 100
!
route-map ROUTE-VALIDATION permit 10
  match rpki invalid
  set local-preference 90
!
```
Configuration (IOS)

• Apply the route-map to inbound updates

```
router bgp 131107
!-----output omitted-----!
address-family ipv4
   bgp bestpath prefix-validate allow-invalid
   neighbor X.X.X.169 activate
   neighbor X.X.X.169 route-map ROUTE-VALIDATION in
exit-address-family
!
address-family ipv6
   bgp bestpath prefix-validate allow-invalid
   neighbor X6:X6:X6::151 activate
   neighbor X6:X6:X6::151 route-map ROUTE-VALIDATION in
exit-address-family
!
```
Establishing session with the validator

```
routing-options {
    autonomous-system 131107;
    validation {
        group rpki-validator {
            session <validator-IP> {
                refresh-time 120;
                port <323/3323/8282>;
                local-address X.X.X.253;
            }
        }
    }
}
```
### Configuration (JunOS)

- Define policies based on the validation states

```plaintext
policy-options {
    policy-statement ROUTE-VALIDATION {
        term valid {
            from {
                protocol bgp;
                validation-database valid;
            }
            then {
                local-preference 110;
                validation-state valid;
                accept;
            }
        }
        term invalid {
            from {
                protocol bgp;
                validation-database invalid;
            }
            then {
                local-preference 90;
                validation-state invalid;
                accept;
            }
        }
        term unknown {
            from {
                protocol bgp;
                validation-database unknown;
            }
            then {
                local-preference 100;
                validation-state unknown;
                accept;
            }
        }
    }
}
```
Router Configuration (JunOS)

- Apply the policy to inbound updates

```
protocols {
    bgp {
        group external-peers {
            neighbor X.X.X.1 {
                import ROUTE-VALIDATION;
                family inet {
                    unicast;
                }
            }
        }
        group external-peers-v6 {
            neighbor X6:X6:X6:X6::1 {
                import ROUTE-VALIDATION;
                family inet6 {
                    unicast;
                }
            }
        }
    }
}
```
RPKI Verification (IOS)

- IOS has only

```bash
#sh bgp ipv6 unicast rpki ?
  servers  Display RPKI cache server information
  table    Display RPKI table entries

#sh bgp ipv4 unicast rpki ?
  servers  Display RPKI cache server information
  table    Display RPKI table entries
```
RPKI Verification (IOS)

- Check the RTR session

```
#sh bgp ipv4 unicast rpki servers

BGP SOVC neighbor is X.X.X.47/323 connected to port 323
Flags 64, Refresh time is 120, Serial number is 1516477445, Session ID is 8871
InQ has 0 messages, OutQ has 0 messages, formatted msg 7826
Session IO flags 3, Session flags 4008
  Neighbor Statistics:
  Prefixes 45661
  Connection attempts: 1
  Connection failures: 0
  Errors sent: 0
  Errors received: 0

Connection state is ESTAB, I/O status: 1, unread input bytes: 0
Connection is ECN Disabled, Minimum incoming TTL 0, Outgoing TTL 255
Local host: X.X.X.225, Local port: 29831
Foreign host: X.X.X.47, Foreign port: 323
```
## RPKI Verification (IOS)

- Check the RPKI cache

```
#sh bgp ipv4 unicast rpki table
37868 BGP sovc network entries using 6058880 bytes of memory
39655 BGP sovc record entries using 1268960 bytes of memory

<table>
<thead>
<tr>
<th>Network</th>
<th>Maxlen</th>
<th>Origin-AS</th>
<th>Source</th>
<th>Neighbor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.9.0.0/16</td>
<td>24</td>
<td>4788</td>
<td>0</td>
<td>202.125.96.47/323</td>
</tr>
<tr>
<td>1.9.12.0/24</td>
<td>24</td>
<td>65037</td>
<td>0</td>
<td>202.125.96.47/323</td>
</tr>
<tr>
<td>1.9.21.0/24</td>
<td>24</td>
<td>24514</td>
<td>0</td>
<td>202.125.96.47/323</td>
</tr>
<tr>
<td>1.9.23.0/24</td>
<td>24</td>
<td>65120</td>
<td>0</td>
<td>202.125.96.47/323</td>
</tr>
</tbody>
</table>

#sh bgp ipv6 unicast rpki table
5309 BGP sovc network entries using 976856 bytes of memory
6006 BGP sovc record entries using 192192 bytes of memory

<table>
<thead>
<tr>
<th>Network</th>
<th>Maxlen</th>
<th>Origin-AS</th>
<th>Source</th>
<th>Neighbor</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001:200::/32</td>
<td>32</td>
<td>2500</td>
<td>0</td>
<td>202.125.96.47/323</td>
</tr>
<tr>
<td>2001:200:900::/40</td>
<td>40</td>
<td>7660</td>
<td>0</td>
<td>202.125.96.47/323</td>
</tr>
<tr>
<td>2001:200:8000::/35</td>
<td>35</td>
<td>4690</td>
<td>0</td>
<td>202.125.96.47/323</td>
</tr>
</tbody>
</table>
```
Check routes (IOS)

```
#sh bgp ipv4 unicast 202.144.128.0/19
BGP routing table entry for 202.144.128.0/19, version 3814371
Paths: (1 available, best #1, table default)
    Advertised to update-groups: 2
    Refresh Epoch 15
    4826 17660
    49.255.232.169 from 49.255.232.169 (114.31.194.12)
    Origin IGP, metric 0, localpref 110, valid, external, best
    Community: 4826:5101 4826:6570 4826:51011 24115:17660
    path 7F50C7CD98C8 RPKI State valid
    rx pathid: 0, tx pathid: 0x0

#sh bgp ipv6 unicast 2402:7800::/32
BGP routing table entry for 2402:7800::/32, version 1157916
Paths: (1 available, best #1, table default)
    Advertised to update-groups: 2
    Refresh Epoch 15
    4826
    2402:7800:10:2::151 from 2402:7800:10:2::151 (114.31.194.12)
    Origin IGP, metric 0, localpref 100, valid, external, best
    path 7F50B266CBD8 RPKI State not found
    rx pathid: 0, tx pathid: 0x0
```
RPKI Verification (JunOS)

- Check the RPKI cache

```
>show validation session
Session  State  Flaps  Uptime  #IPv4/IPv6 records
X.X.X.46  Up     75     09:20:59     40894/6747

>show validation session 202.125.96.46
Session  State  Flaps  Uptime  #IPv4/IPv6 records
X.X.X.46  Up     75     09:21:18     40894/6747
```
**RPKI Verification (JunOS)**

- **Check the RPKI cache**

```
>show validation database
RV database for instance master

<table>
<thead>
<tr>
<th>Prefix</th>
<th>Origin-AS</th>
<th>Session</th>
<th>State</th>
<th>Mismatch</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.9.0.0/16–24</td>
<td>4788</td>
<td>202.125.96.46</td>
<td>valid</td>
<td></td>
</tr>
<tr>
<td>1.9.12.0/24–24</td>
<td>65037</td>
<td>202.125.96.46</td>
<td>valid</td>
<td></td>
</tr>
<tr>
<td>1.9.21.0/24–24</td>
<td>24514</td>
<td>202.125.96.46</td>
<td>valid</td>
<td></td>
</tr>
<tr>
<td>1.9.23.0/24–24</td>
<td>65120</td>
<td>202.125.96.46</td>
<td>valid</td>
<td></td>
</tr>
<tr>
<td>2001:200::/32–32</td>
<td>2500</td>
<td>202.125.96.46</td>
<td>valid</td>
<td></td>
</tr>
<tr>
<td>2001:200:900::/40–40</td>
<td>16567</td>
<td>202.125.96.46</td>
<td>valid</td>
<td></td>
</tr>
<tr>
<td>2001:200:c000::/35–35</td>
<td>23634</td>
<td>202.125.96.46</td>
<td>valid</td>
<td></td>
</tr>
<tr>
<td>2001:200:e000::/35–35</td>
<td>7660</td>
<td>202.125.96.46</td>
<td>valid</td>
<td></td>
</tr>
</tbody>
</table>
```

*Would have been nice if they had per AF!*
RPKI Verification (JunOS)

- Can filter per origin ASN

```
>show validation database origin-autonomous-system 45192
RV database for instance master

<table>
<thead>
<tr>
<th>Prefix</th>
<th>Origin-AS</th>
<th>Session</th>
<th>State</th>
<th>Mismatch</th>
</tr>
</thead>
<tbody>
<tr>
<td>202.125.97.0/24-24</td>
<td>45192</td>
<td>202.125.96.46</td>
<td>valid</td>
<td></td>
</tr>
<tr>
<td>203.176.189.0/24-24</td>
<td>45192</td>
<td>202.125.96.46</td>
<td>valid</td>
<td></td>
</tr>
<tr>
<td>2001:df2:ee01::/48-48</td>
<td>45192</td>
<td>202.125.96.46</td>
<td>valid</td>
<td></td>
</tr>
</tbody>
</table>
```

IPv4 records: 2
IPv6 records: 1

*_IOS should have something similar!*
Check routes (JunOS)

>show route protocol bgp 202.144.128.0

inet.0: 693024 destinations, 693024 routes (693022 active, 0 holddown, 2 hidden)
+ = Active Route, - = Last Active, * = Both

202.144.128.0/20 *[BGP/170] 1w4d 21:03:04, MED 0, localpref 110, from 202.125.96.254
  AS path: 4826 17660 I, validation-state: valid
  >to 202.125.96.225 via ge-1/1/0.0

>show route protocol bgp 2001:201::/32

inet6.0: 93909 destinations, 93910 routes (93909 active, 0 holddown, 0 hidden)
+ = Active Route, - = Last Active, * = Both

2001:201::/32 *[BGP/170] 21:18:14, MED 0, localpref 100, from 2001:df2:ee00::1
  AS path: 65332 I, validation-state: unknown
  >to fe80::dab1:90ff:fedc:fd07 via ge-1/1/0.0
Propagating RPKI states to iBGP peers

- To avoid every BGP speaker having an RTR session, and

- All BGP speakers have consistent information

- Relies on extended BGP communities (RFC8097)

<table>
<thead>
<tr>
<th>Value</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Lookup result = &quot;valid&quot;</td>
</tr>
<tr>
<td>1</td>
<td>Lookup result = &quot;not found&quot;</td>
</tr>
<tr>
<td>2</td>
<td>Lookup result = &quot;invalid&quot;</td>
</tr>
</tbody>
</table>

- Sender (one that has RTR session) attaches the extended community to Updates, and receiver derives the validation states from it
- Must be enabled on both sender and receiver!
Propagating RPKI states (IOS)

- Sender (one with RTR session)

```plaintext
router bgp 131107
  bgp rpki server tcp <validator-IP> port <323/8282/3323> refresh 120
!---output omitted-----!
address-family ipv4
  neighbor X.X.X.X activate
  neighbor X.X.X.X send--community both
  neighbor X.X.X.X announce rpki state
exit-address-family
!
address-family ipv6
  neighbor X6:X6:X6::X6 activate
  neighbor X6:X6:X6:X6::X6 send--community both
  neighbor X6:X6:X6::X6 announce rpki state
exit-address-family
!
```
Propagating RPKI states (IOS)

- Receiver (iBGP peer)

```conf
router bgp 131107
!---output omitted-----!
address-family ipv4
  neighbor Y.Y.Y.Y activate
  neighbor Y.Y.Y.Y send-community both
  neighbor Y.Y.Y.Y announce rpki state
exit-address-family
!
address-family ipv6
  neighbor Y6:Y6:Y6:Y6::Y6 activate
  neighbor Y6:Y6:Y6:Y6::Y6 send-community both
  neighbor Y6:Y6:Y6:Y6::Y6 announce rpki state
exit-address-family
!
```

- If `announce rpki state` is not configured for the neighbor, all prefixes received from the iBGP neighbor will be marked VALID!
Propagating RPKI states (JunOS)

- Sender (one with RTR session)

```plaintext
policy-statement ROUTE-VALIDATION {
    term valid {
        from {
            protocol bgp;
            validation-database valid;
        }
        then {
            local-preference 110;
            validation-state valid;
            community add origin-validation-state-valid;
            accept;
        }
    }
    term invalid {
        from {
            protocol bgp;
            validation-database invalid;
        }
        then {
            local-preference 90;
            validation-state invalid;
            community add origin-validation-state-invalid;
            accept;
        }
    }
    term unknown {
        from {
            protocol bgp;
            validation-database unknown;
        }
        then {
            local-preference 100;
            validation-state unknown;
            community add origin-validation-state-unknown;
            accept;
        }
    }
}
```
Propagating RPKI states (JunOS)

- Receiver (iBGP peer)

```plaintext
policy-statement ROUTE-VALIDATION-1 {
  term valid {
    from community origin-validation-state-valid;
    then validation-state valid;
  }
  term invalid {
    from community origin-validation-state-invalid;
    then validation-state invalid;
  }
  term unknown {
    from community origin-validation-state-unknown;
    then validation-state unknown;
  }
}
```
Propagating RPKI states – potential issues

- IOS as BR, propagating states to JunOS iBGP peers
  unknown iana 4300

  Hack:
  - Either act on the states at the border, or
  - Match and tag them with custom communities before propagating
Configuration - Reference Link

- **Cisco**

- **Juniper**

- **RIPE**
Operational Caveats

- When RTR session goes down, the validation state changes to **Not Found** for all routes after a while
  - *Invalid* => *Not Found*
  - *we need at least two RTR sessions* and/or need careful filtering policies

- During a router reload, do we receive ROAs first or BGP updates first?
  - If BGP update is faster than ROA, will propagate even invalid routes to its iBGP peers
**Useful tools**

- **RIPEstat – prefix/ASN**
  - [https://stat.ripe.net/data/rpki-validation/data.json?resource=45192&prefix=202.125.96.0/24](https://stat.ripe.net/data/rpki-validation/data.json?resource=45192&prefix=202.125.96.0/24)

```
<table>
<thead>
<tr>
<th>JSON</th>
<th>Raw Data</th>
<th>Headers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Save</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Copy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Collapse All</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Expand All</td>
</tr>
</tbody>
</table>

```json

```
status: "ok"
success: "app004"
status_code: 200
version: "0.2"
cached: false
see_also: []
time: "2019-04-09T08:44:30.950267"
messages: []
data_call_status: "supported"
process_time: 34
build_version: "2019.4.8.02"
query_id: "20190409094430-515c3dbb-4a99-4036-9e66-2112d5d7d36"
data:
  validating_roas:
    - origin: "AS131187"
    - source: "APNIC RPKI Root"
    - prefix: "202.125.96.0/24"
    - max_length: 24
    - validity: "invalid_asn"
status: "invalid_asn"
prefix: "202.125.96.0/24"
resource: "45192"
```
https://www.apnic.net/community/security/resource-certification/#routing
Any questions?