APNIC

# Securing Internet Routing (with RPKI)

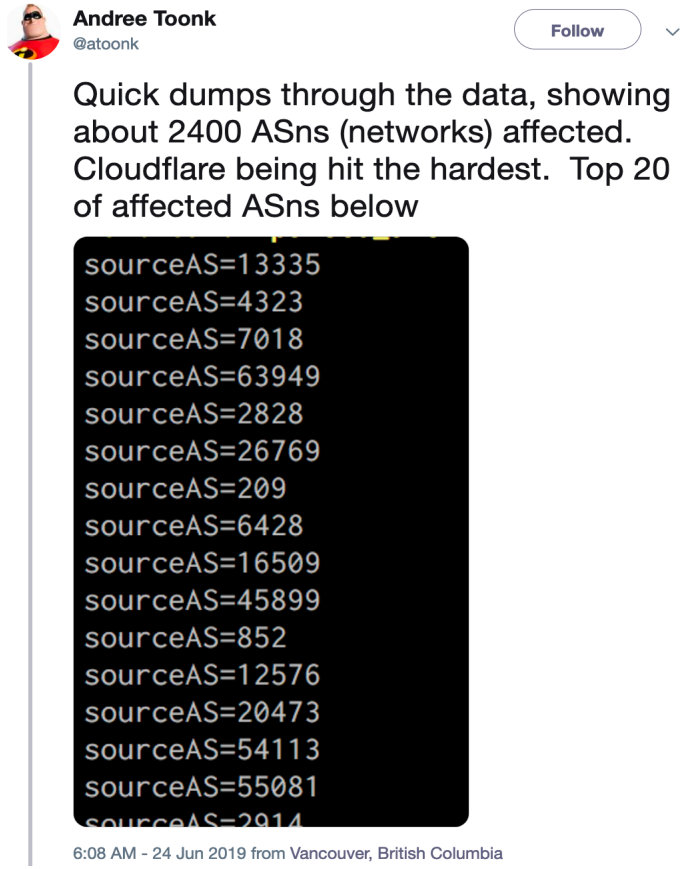

v1.0

# Headlines

- BGP Optimizers impact Internet – **June 2019**
  - 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 [redacted] wrote:

> Hello are there any issues with CloudFlare services now?

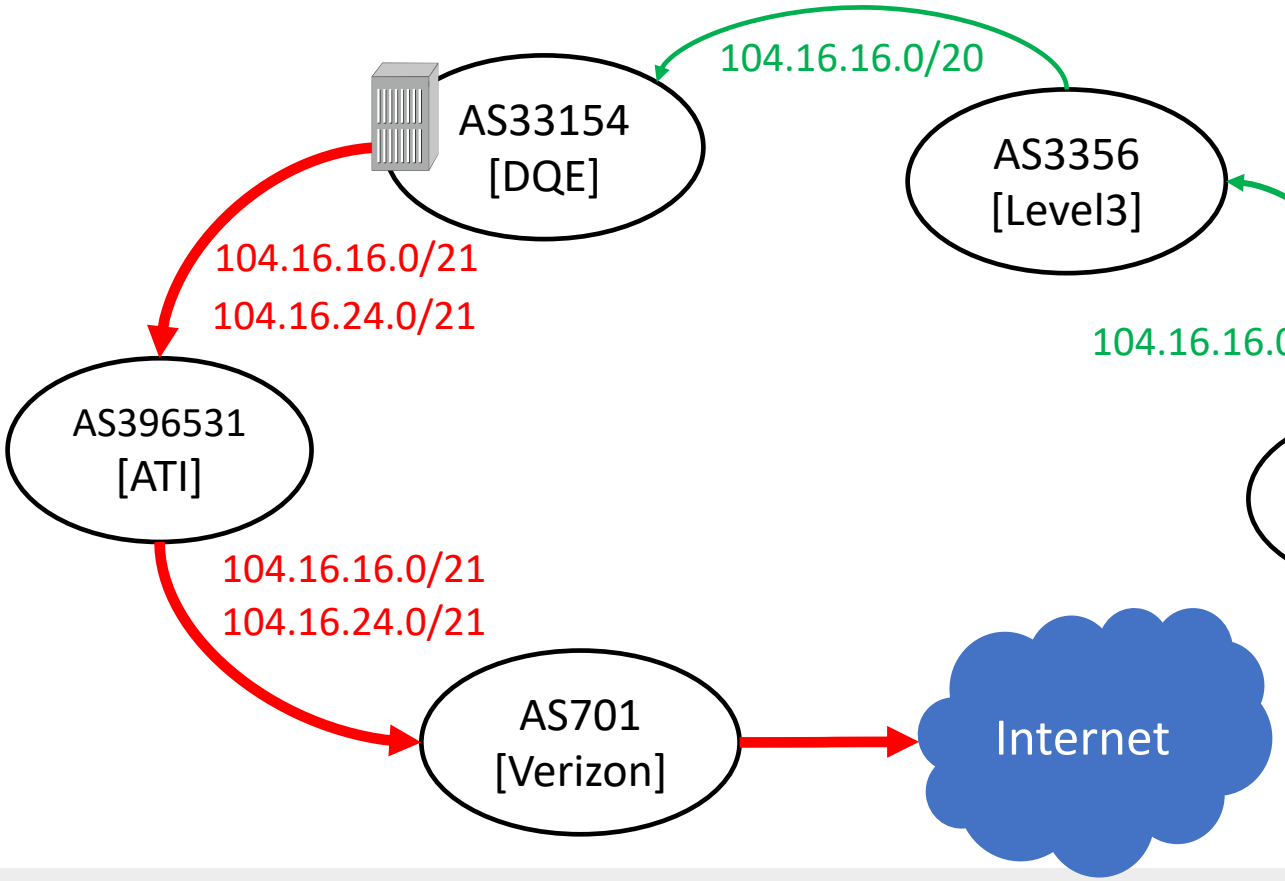



https://twitter.com/atoonk/status/1143143943531454464/photo/1


v1.0

# Headlines

- BGP Optimizers impact Internet (contd…)
  - What & How?




**BGP Optimizers (Was: Validating possible BGP MITM attack)**

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

Dear all,

disclaimer:

[ The following is targetted 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.

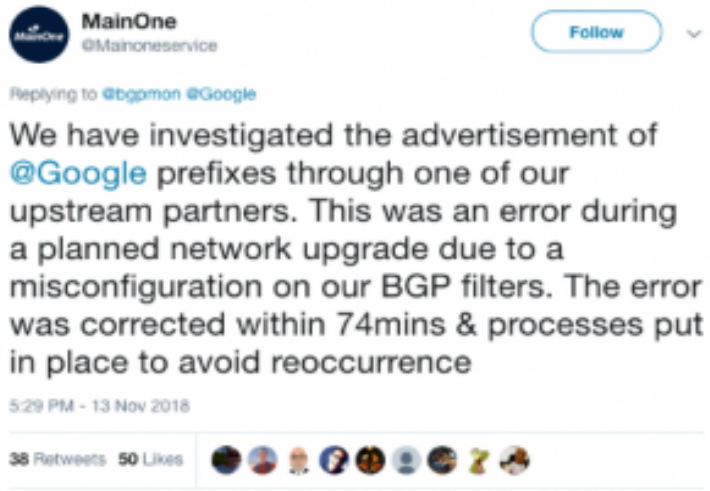

v1.0

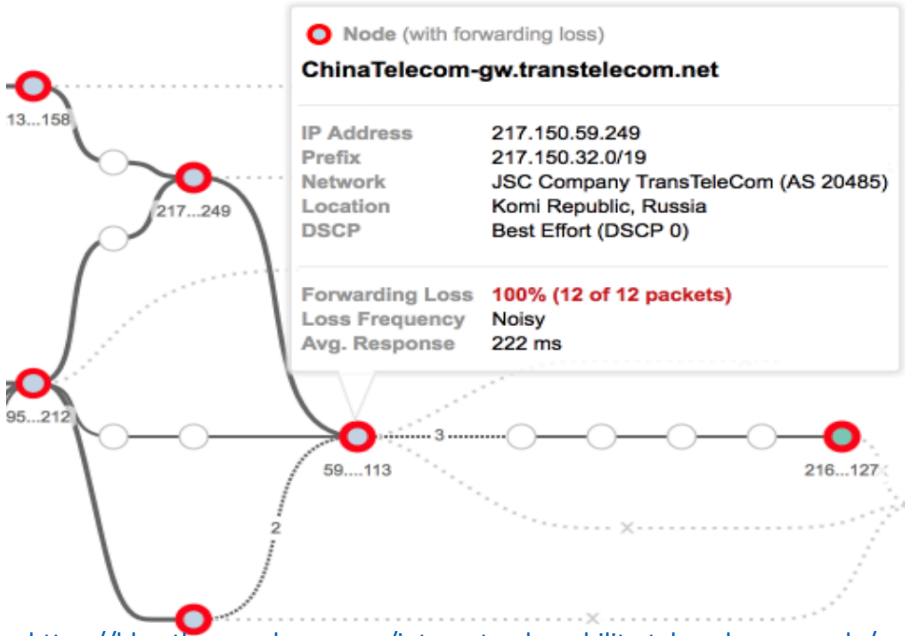
# Headlines

- 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

# Headlines

- Google prefix leaks (contd…)
  - How?
    - AS37282 (MainOne) leaked to AS4809 (CT) at IXPN, who leaked it to others like AS20485 (TransTelecom)




https://blog.thousandeyes.com/internet-vulnerability-takes-down-google/

# Headlines

- Amazon (AS16509) Route53 hijack – **April2018**
  - AS10279 (eNET) originated more specifics (/24s) of Amazon Route53's prefix (205.251.192.0/21)

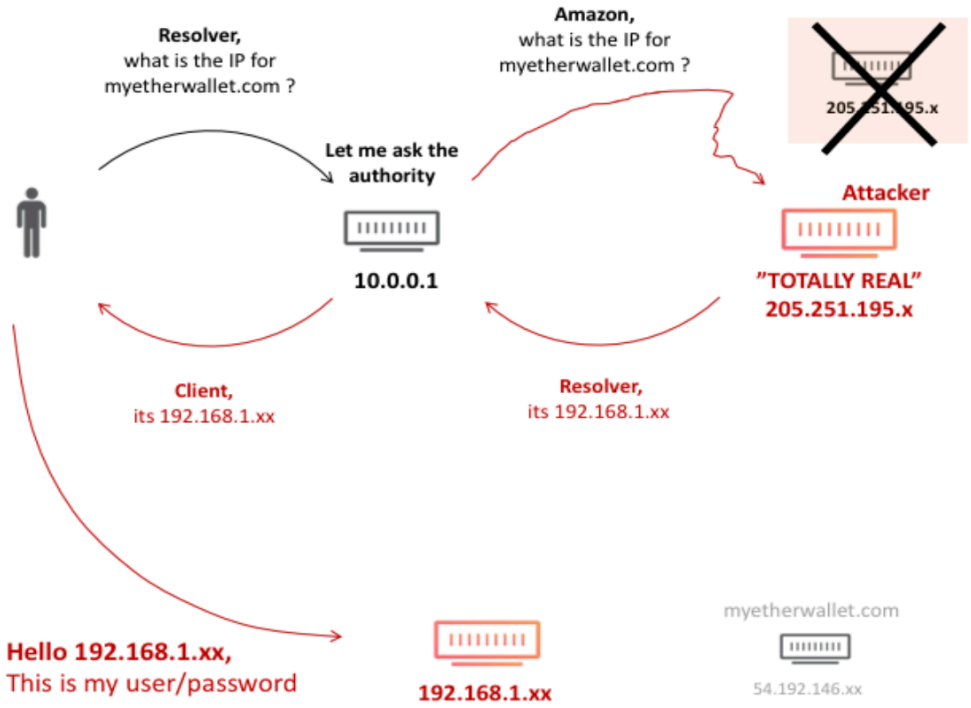
    **205.251.192.0/24 ……. 205.251.199.0/24**

    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


v1.0

# Headlines

- 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](myetherwallet.com)
  - use *_THEIR_* crypto to end-users to see everything (including passwords)




https://blog.cloudflare.com/bgp-leaks-and-crypto-currencies

# Headlines

- ~~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!


v1.0

# Headlines

- Google leak (contd…)

```
trace from Tokyo, Japan to Inuyama, Japan at 04:44 Aug 24, 2017
1  *
2  202.177.203.50    xe-0-0-0.gw401.ty2.ap.equinix.com  Tokyo       Japan    0.717
3  183.177.32.143    xe-1-1-1.gw402.ty1.ap.equinix.com  Tokyo       Japan    0.755
4  143.90.232.25     25.143090232.odn.ne.jp             Tokyo       Japan    1.411
5  143.90.161.73                                        Tokyo       Japan    2.757
6  143.90.47.14      STOrs-01Te0-1-0-1.nw.odn.ad.jp     Tokyo       Japan    3.552
7  210.252.167.230   230.210252167.odn.ne.jp            Tokyo       Japan    4.094
8  *
9  60.37.54.105      OCN (AS4713) CIDR BLOCK 70         Tokyo       Japan    4.088
10 125.170.97.85     OCN (AS4713) CIDR BLOCK 77                     Japan    4.017
11 125.170.97.74     OCN (AS4713) CIDR BLOCK 77         Ōsaka-shi   Japan   12.263
12 153.149.219.22    OCN (AS4713) CIDR BLOCK 93         Ōsaka-shi   Japan   12.362
13 153.146.148.18    OCN (AS4713) CIDR BLOCK 93         Tokyo       Japan    14.45
14 60.37.32.250      OCN (AS4713) CIDR BLOCK 70                     Japan   13.116
15 118.23.141.202    OCN (AS4713) CIDR BLOCK 86                     Japan   13.332
16 118.23.142.99     OCN (AS4713) CIDR BLOCK 86                     Japan   22.307
17 211.11.83.160     OCN (AS4713) CIDR BLOCK 23         Inuyama     Japan   15.672
```

**Before leak (JP->JP)**

**After leak (JP->JP)**

```
trace from Tokyo, Japan to Inuyama, Japan at 03:28 Aug 25, 2017
1  *
2  183.177.32.145    Equinix Asia Pacific               Tokyo       Japan    0.249
3  210.130.154.37    IIJ IPv4 BLOCK ( AS2497 )          Tokyo       Japan    0.618
4  58.138.102.109    tky001bb11.IIJ.Net                 Tokyo       Japan    0.877
5  58.138.88.86      sjc002bb12.IIJ.Net                 San Jose    United States   97.797
6  152.179.48.117    TenGigE0-3-0-8.GW6.SJC7.ALTER.NET  San Jose    United States   97.869
7  *
8  152.179.105.110   google-gw.customer.alter.net       Chicago     United States   337.19
9  108.170.243.197   Google Inc.                        Chicago     United States   246.325
10 *
11 209.85.241.43     Google Inc.                                    United States   256.188
12 72.14.238.38      Google Inc.                        Vancouver   Canada          247.849
13 209.85.245.110    Google Inc.                        Vancouver   Canada          249.291
14 *
15 108.170.242.138   Google Inc.                        Tokyo       Japan           246.267
16 211.0.193.21      OCN (AS4713) CIDR BLOCK 21         Tokyo       Japan           246.351
17 122.1.245.65      OCN (AS4713) CIDR BLOCK 81         Tokyo       Japan           246.426
18 *
19 153.149.218.10    OCN (AS4713) CIDR BLOCK 93         Ōsaka-shi   Japan           256.027
20 125.170.96.38     OCN (AS4713) CIDR BLOCK 77                     Japan           255.683
21 *
22 60.37.32.250      OCN (AS4713) CIDR BLOCK 70                     Japan           254.989
23 118.23.141.202    OCN (AS4713) CIDR BLOCK 86                     Japan           254.526
24 *
25 211.11.83.160     OCN (AS4713) CIDR BLOCK 23         Inuyama     Japan           256.212
```

**After leak (EU->EU)**

```
trace from London, England to Nürnberg, Germany at 03:30 Aug 25, 2017
1  *
2  195.66.248.190    fe0-2.tr2.linx.net                 London          United Kingdom   0.327
3  195.66.249.10     ge0-2-502.tr5.linx.net             London          United Kingdom   0.441
4  195.66.249.13     ge0-2-501.tr4.linx.net             London          United Kingdom   0.477
5  195.66.248.10     uunet-uk-transit.thn.linx.net      London          United Kingdom   0.507
6  158.43.193.245    POS0-0.CR2.LND6.ALTER.NET          London          United Kingdom   0.497
7  140.222.239.41    0.xe-0-0-0.IL1.NYC50.ALTER.NET     New York        United States   108.146
8  146.188.4.197     xe-0-0-1.IL1.NYC41.ALTER.NET       New York        United States    75.719
9  140.222.234.221   0.et-10-1-0.GW7.CHI13.ALTER.NET    Chicago         United States    94.793
10 152.179.105.110   google-gw.customer.alter.net       Chicago         United States   224.352
11 *
12 216.239.40.189    Google Inc.                        Northlake       United States   202.193
13 216.239.58.255    Google Inc.                                                        203.995
14 216.239.58.12     Google Inc.                                                        207.026
15 209.85.253.184    Google Inc.                        Luxembourg      Luxembourg      212.944
16 209.85.252.215    Google Inc.                                                        213.112
17 108.170.252.71    Google Inc.                                                        213.265
18 72.14.222.53      Google Inc.                                        Germany         212.061
19 188.111.165.169   Vodafone GmbH                                      Germany         227.077
20 178.7.138.112     Vodafone D2 GmbH                   Nürnberg        Germany         234.226
```

https://dyn.com/blog/large-bgp-leak-by-google-disrupts-internet-in-japan/

# Headlines

- YouTube (AS36561) Incident - **Feb 2008**
  - ~ 2 hours
  - AS17557 (PT) announced 208.65.153.0/24 (208.65.152.0/22)
    - Propagated by AS3491 (PCCW)


v1.0

# 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


v1.0

# 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?


v1.0

# 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 ☹


v1.0

# Why do we keep seeing these?

- Routing works in REVERSE
  - Outbound advertisement affects inbound traffic
  - Inbound (*Accepted*) advertisement influence outbound traffic


v1.0

# Why do we keep seeing these?

- As always, there is no E-bit (evil!)
  - RFC3514 ☺

  - 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???


v1.0

# Why should we worry?

- Because it’s just so easy to do bad in routing!

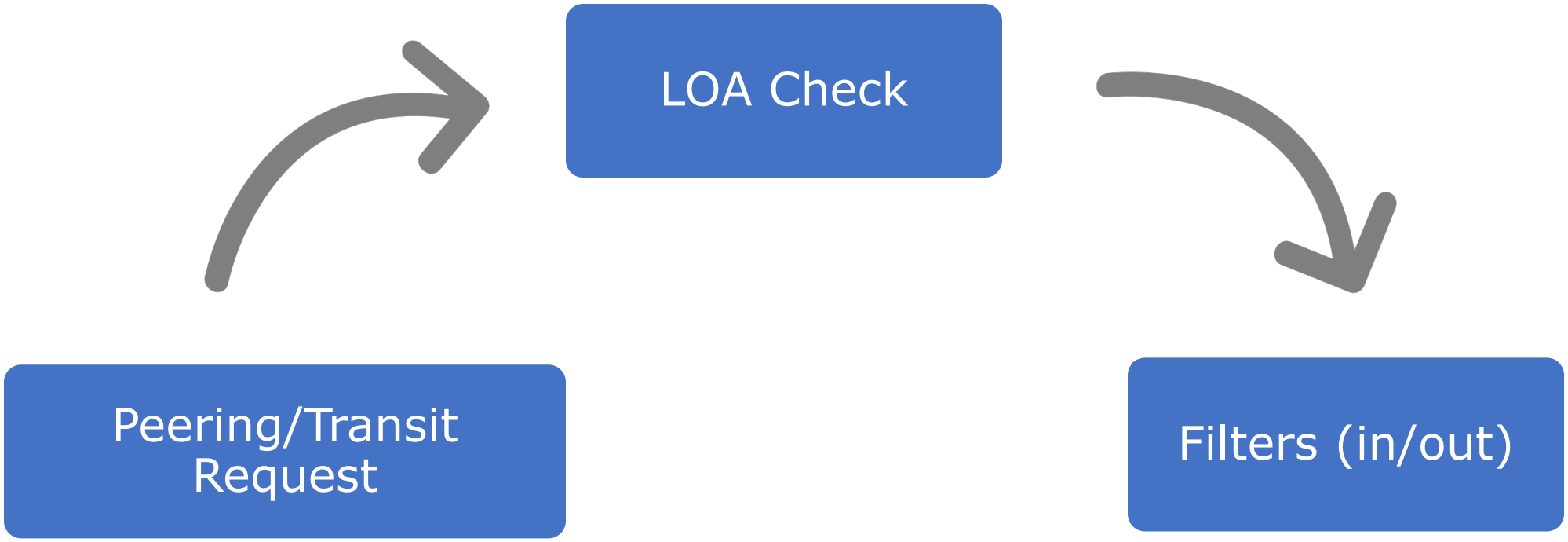
By Source (WP:NFCC#4), Fair use,
https://en.wikipedia.org/w/index.php?curid=42515224

# How do we address these?

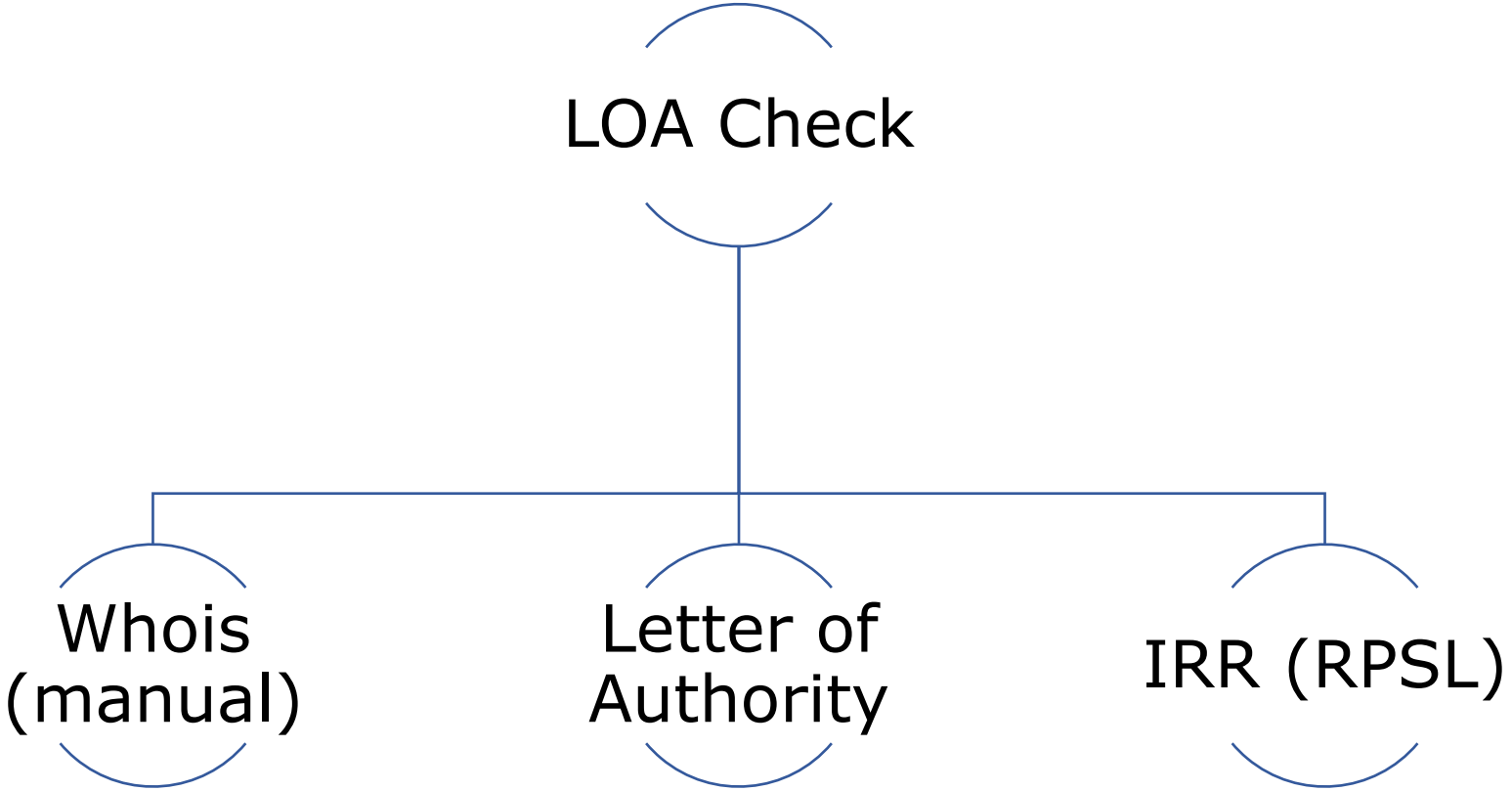
- **Good Hygiene ~ Filter Filter Filter!**
  - your peers, upstream(s), and customers
    - Prefix filters/Prefix limit
    - AS-PATH filters/AS-PATH limit
    - RFC 8212 – BGP default reject or something similar


v1.0

# Current practice



# Tools & Techniques



# 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

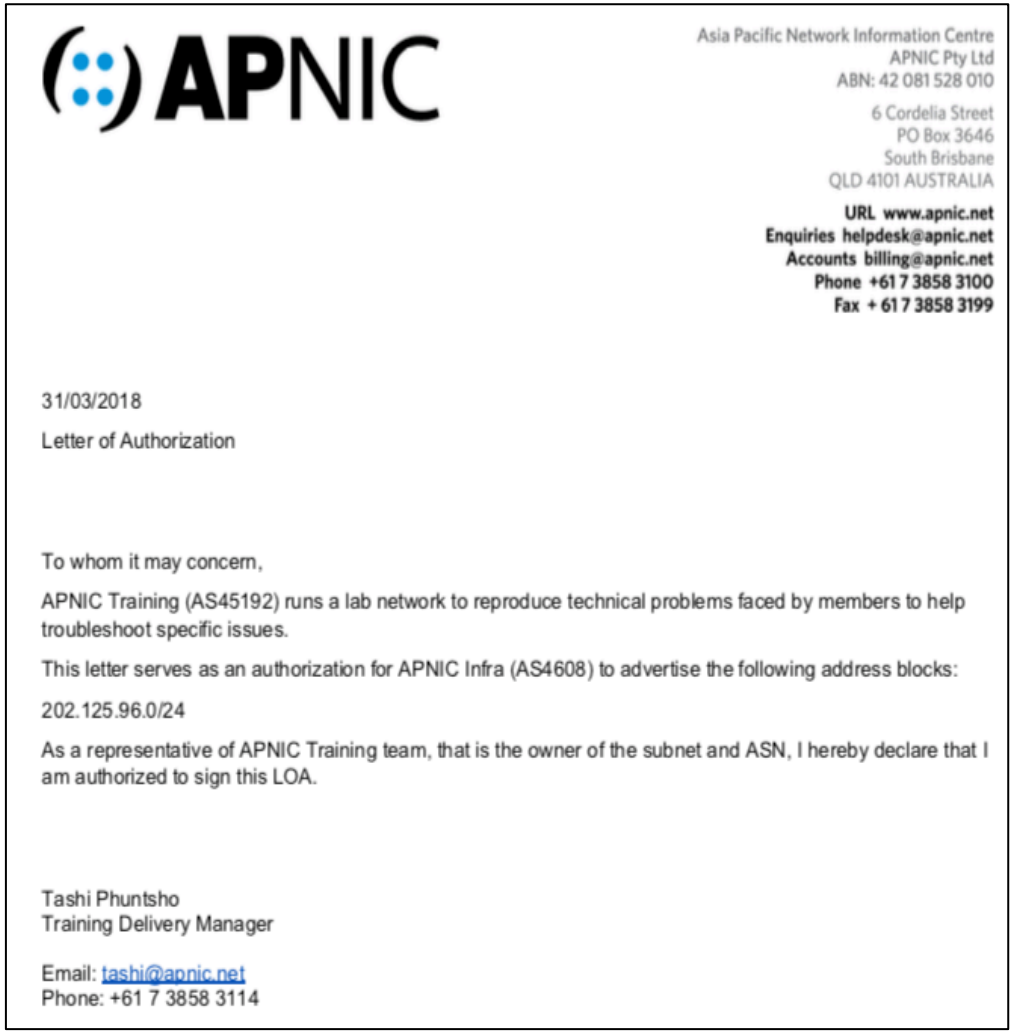
role:           APNIC Training
address:        6 Cordelia Street
address:        South Brisbane
address:        QLD 4101
country:        AU
phone:          +61 7 3858 3100
fax-no:         +61 7 3858 3199
e-mail:         training@apnic.net
admin-c:        JW3997-AP
tech-c:         JW3997-AP
nic-hdl: 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

% Information related to '202.125.96.0/24AS131107'

route:          202.125.96.0/24
descr:          Prefix for APNICTRAINING LAB DC
origin:         AS131107
mnt-by:         MAINT-AU-APNICTRAINING
country:        AU
last-modified:  2016-06-16T23:23:00Z
source:         APNIC
```

# Tools & Techniques

- Ask for a **Letter of Authority**
  - Absolve from any liabilities



APNIC

Asia Pacific Network Information Centre
APNIC Pty Ltd
ABN: 42 081 528 010

6 Cordelia Street
PO Box 3646
South Brisbane
QLD 4101 AUSTRALIA

URL www.apnic.net
Enquiries helpdesk@apnic.net
Accounts billing@apnic.net
Phone +61 7 3858 3100
Fax + 61 7 3858 3199

31/03/2018

Letter of Authorization

To whom it may concern,

APNIC Training (AS45192) runs a lab network to reproduce technical problems faced by members to help troubleshoot specific issues.

This letter serves as an authorization for APNIC Infra (AS4608) to advertise the following address blocks:

202.125.96.0/24

As a representative of APNIC Training team, that is the owner of the subnet and ASN, I hereby declare that I am authorized to sign this LOA.

Tashi Phuntsho
Training Delivery Manager

Email: tashi@apnic.net
Phone: +61 7 3858 3114

# Tools & Techniques

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

```
tashi@tashi ~> whois -h whois.radb.net 61.45.248.0/24
route:          61.45.248.0/24
descr:          APNICTRAINING-DC
origin:         AS135533
mnt-by:         MAINT-AS4826
changed:        noc@vocus.com.au 20160702
source:         RADB

route:          61.45.248.0/24
descr:          Prefix for APNICTRAINING LAB - AS135533
origin:         AS135533
mnt-by:         MAINT-AU-APNICTRAININGLAB
country:        AU
last-modified:  2017-10-19T01:36:37Z
source:         APNIC
```

```
tashi@tashi ~> whois -h whois.radb.net AS17660
aut-num:        AS17660
as-name:        BT-Bhutan
descr:          Divinetworks for BT
admin-c:        DUMY-RIPE
tech-c:         DUMY-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:         from AS6461   action pref=100;     accept ANY
export:         to AS6461     announce AS-DRUKNET-TRANSIT
import:         from AS2914   action pref=150;     accept ANY
export:         to AS2914     announce AS-DRUKNET-TRANSIT
import:         from AS6453   action pref=100;     accept ANY
export:         to AS6453     announce AS-DRUKNET-TRANSIT
```

# Tools & Techniques

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

```
tashi@tashi ~> bgpq3 -Al PEER-v4IN AS17660
no ip prefix-list PEER-v4IN
ip prefix-list PEER-v4IN permit 45.64.248.0/22
ip prefix-list PEER-v4IN permit 103.7.252.0/22
ip prefix-list PEER-v4IN permit 103.7.254.0/23
ip prefix-list PEER-v4IN permit 103.245.240.0/22
ip prefix-list PEER-v4IN permit 103.245.242.0/23
ip prefix-list PEER-v4IN permit 119.2.96.0/19
ip prefix-list PEER-v4IN permit 119.2.96.0/20
ip prefix-list PEER-v4IN permit 202.89.24.0/21
ip prefix-list PEER-v4IN permit 202.144.128.0/19
ip prefix-list PEER-v4IN permit 202.144.128.0/23
ip prefix-list PEER-v4IN permit 202.144.144.0/20
ip prefix-list PEER-v4IN permit 202.144.148.0/22
tashi@tashi ~> bgpq3 -6Al PEER-v6IN AS17660
no ipv6 prefix-list PEER-v6IN
ipv6 prefix-list PEER-v6IN permit 2405:d000::/32
ipv6 prefix-list PEER-v6IN permit 2405:d000:7000::/36
```

```
tashi@tashi ~> bgpq3 -Abl PEER-v4IN AS17660
PEER-v4IN = [
    45.64.248.0/22,
    103.7.252.0/22,
    103.7.254.0/23,
    103.245.240.0/22,
    103.245.242.0/23,
    119.2.96.0/19,
    119.2.96.0/20,
    202.89.24.0/21,
    202.144.128.0/19,
    202.144.128.0/23,
    202.144.144.0/20,
    202.144.148.0/22
];
tashi@tashi ~> bgpq3 -6Abl PEER-v6IN AS17660
PEER-v6IN = [
    2405:d000::/32,
    2405:d000:7000::/36
];
```

```
tashi@tashi ~> bgpq3 -f 38195 -lSUPERLOOP-IN AS-SUPERLOOP
no ip as-path access-list SUPERLOOP-IN
ip as-path access-list SUPERLOOP-IN permit ^38195(_38195)*$
ip as-path access-list SUPERLOOP-IN permit ^38195(_[0-9]+)*_(681|4647|4749|4785)$
ip as-path access-list SUPERLOOP-IN permit ^38195(_[0-9]+)*_(4846|4858|7477|7578)$
ip as-path access-list SUPERLOOP-IN permit ^38195(_[0-9]+)*_(7585|7604|7628|7631)$
ip as-path access-list SUPERLOOP-IN permit ^38195(_[0-9]+)*_(7699|9290|9297|9336)$
ip as-path access-list SUPERLOOP-IN permit ^38195(_[0-9]+)*_(9499|9544|9549|10143)$
ip as-path access-list SUPERLOOP-IN permit ^38195(_[0-9]+)*_(10145|11031|12041|15133)$
ip as-path access-list SUPERLOOP-IN permit ^38195(_[0-9]+)*_(15967|17462|17498|17766)$
ip as-path access-list SUPERLOOP-IN permit ^38195(_[0-9]+)*_(17829|17907|17991|18000)$
ip as-path access-list SUPERLOOP-IN permit ^38195(_[0-9]+)*_(18110|18201|18292|23156)$
ip as-path access-list SUPERLOOP-IN permit ^38195(_[0-9]+)*_(23456|23677|23858|23935)$
ip as-path access-list SUPERLOOP-IN permit ^38195(_[0-9]+)*_(24007|24065|24093|24129)$
ip as-path access-list SUPERLOOP-IN permit ^38195(_[0-9]+)*_(24231|24233|24238|24341)$
ip as-path access-list SUPERLOOP-IN permit ^38195(_[0-9]+)*_(24459|27232|30215|30762)$
ip as-path access-list SUPERLOOP-IN permit ^38195(_[0-9]+)*_(36351|37993|38263|38269)$
ip as-path access-list SUPERLOOP-IN permit ^38195(_[0-9]+)*_(38451|38534|38549|38570)$
ip as-path access-list SUPERLOOP-IN permit ^38195(_[0-9]+)*_(38595|38716|38719|38790)$
ip as-path access-list SUPERLOOP-IN permit ^38195(_[0-9]+)*_(38809|38830|38858|42909)$
ip as-path access-list SUPERLOOP-IN permit ^38195(_[0-9]+)*_(44239|45158|45267|45278)$
ip as-path access-list SUPERLOOP-IN permit ^38195(_[0-9]+)*_(45570|45577|45638|45671)$
ip as-path access-list SUPERLOOP-IN permit ^38195(_[0-9]+)*_(45844|46571|55411|55419)$
ip as-path access-list SUPERLOOP-IN permit ^38195(_[0-9]+)*_(55455|55506|55575|55707)$
ip as-path access-list SUPERLOOP-IN permit ^38195(_[0-9]+)*_(55752|55766|55803|55845)$
ip as-path access-list SUPERLOOP-IN permit ^38195(_[0-9]+)*_(55884|55931|55954|56037)$
ip as-path access-list SUPERLOOP-IN permit ^38195(_[0-9]+)*_(56098|56135|56178|56225)$
ip as-path access-list SUPERLOOP-IN permit ^38195(_[0-9]+)*_(56271|56287|58422|58443)$
ip as-path access-list SUPERLOOP-IN permit ^38195(_[0-9]+)*_(58511|58606|58634|58676)$
ip as-path access-list SUPERLOOP-IN permit ^38195(_[0-9]+)*_(58712|58739|58750|58868)$
ip as-path access-list SUPERLOOP-IN permit ^38195(_[0-9]+)*_(58914|59256|59330|59339)$
ip as-path access-list SUPERLOOP-IN permit ^38195(_[0-9]+)*_(59356|60592|60758|63926)$
ip as-path access-list SUPERLOOP-IN permit ^38195(_[0-9]+)*_(63937|63956)$
```

# 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)?


v1.0

# 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


v1.0

# Back to basics – identify GOOD

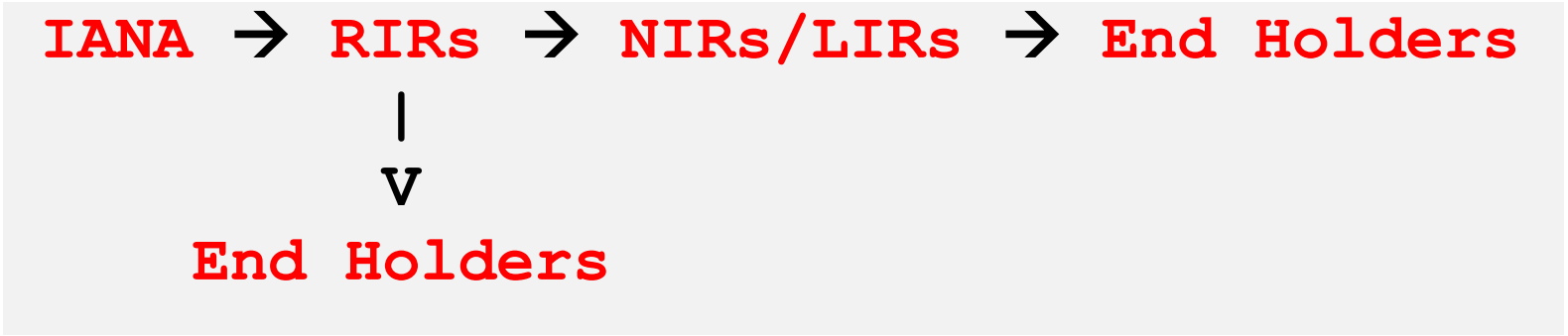
- Could we use a digital signature to convey the *authority to use*?
  - Private key to *sign* the *authority*, and
  - Public key to *validate* the *authority*

- ~ If the holder of the resource has the private key, it can sign/authorize the use of the resource


v1.0

# How about trust?

- How do we build a chain of trust in this framework??
  - Follow the resource allocation/delegation hierarchy




    - To describe the address allocation using digital certificates

# RPKI Chain of Trust

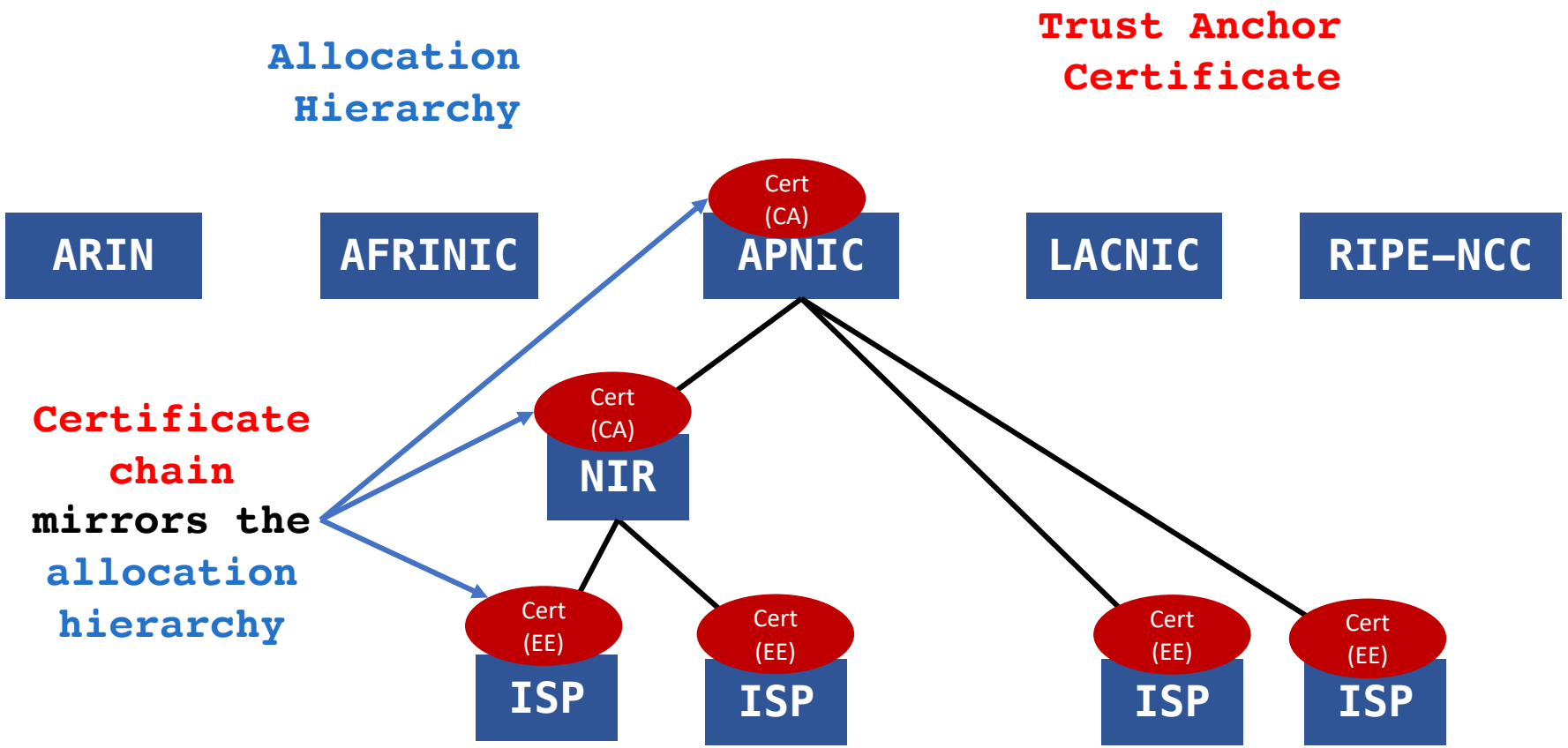


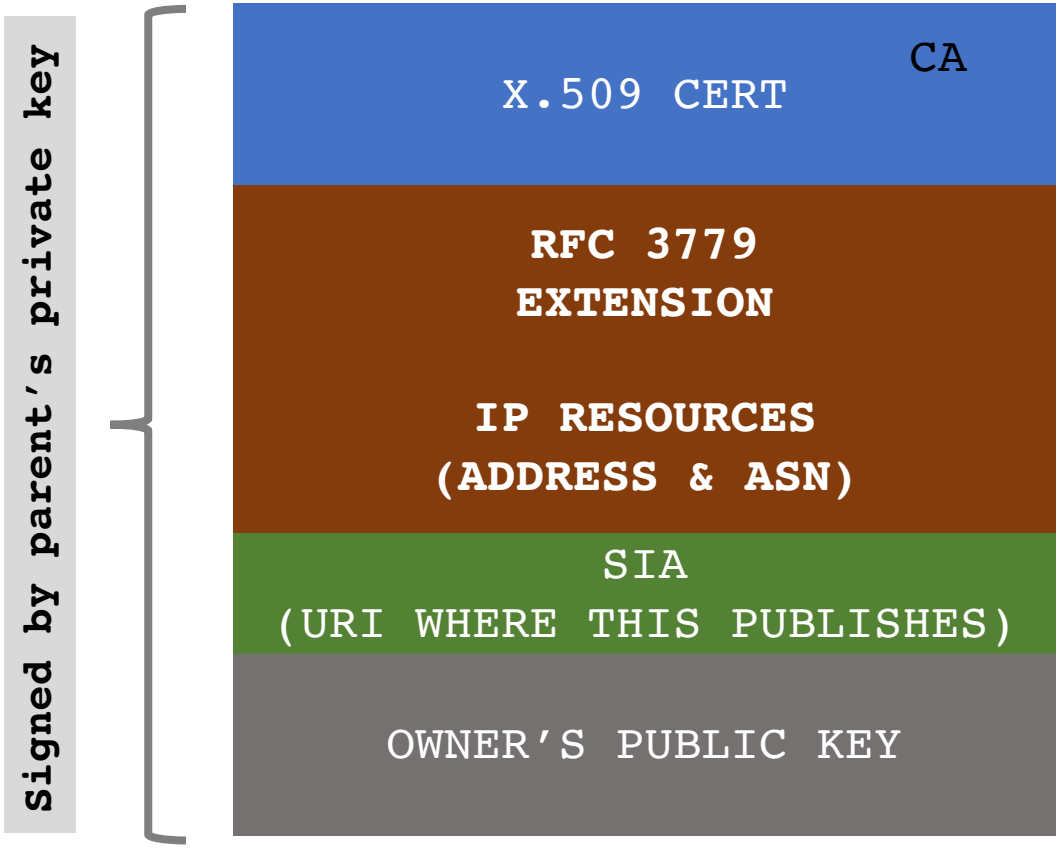

Image 4

# 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


v1.0

# RPKI profile ~ Resource Certificates




- RFC 3779 extensions – binds a list of resources (**IPv4/v6,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 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!*


v1.0

# Route Origin Authorization (ROA)

- (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, and **sign** this permission with its private key (~ROA)


v1.0

# Route Origin Authorization (ROA)

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

| **Prefix** | 203.176.32.0/19 |
|---|---|
| **Max-length** | /24 |
| **Origin ASN** | **AS17821** |

- ** *Multiple ROAs can exist for the same prefix*


v1.0

# What can RPKI do?

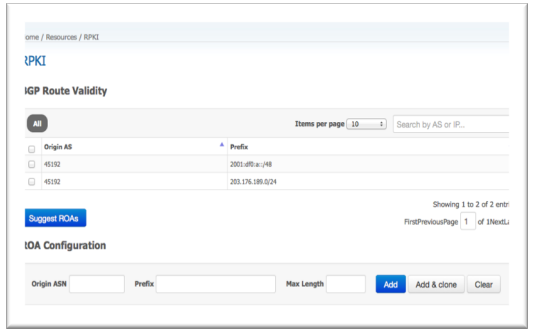
- Authoritatively proof:
  - Who is the legitimate owner of an address, and
  - Identify which ASNs have the permission from the holder to originate the address

- Can help:
  - prevent **route hijacks/mis-origination/misconfiguration**

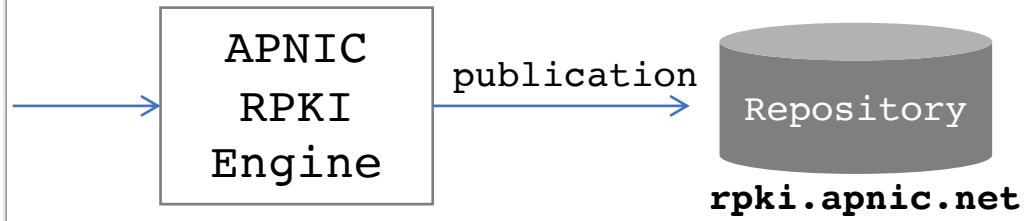

v1.0

# 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

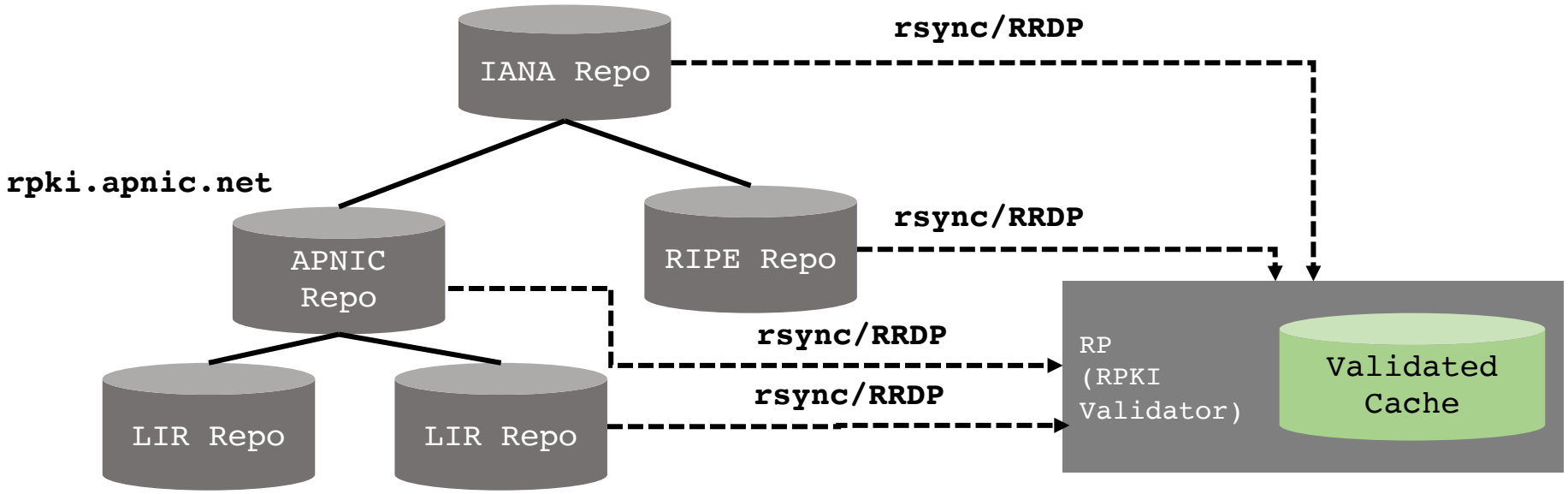



MyAPNIC GUI



# RPKI Components

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

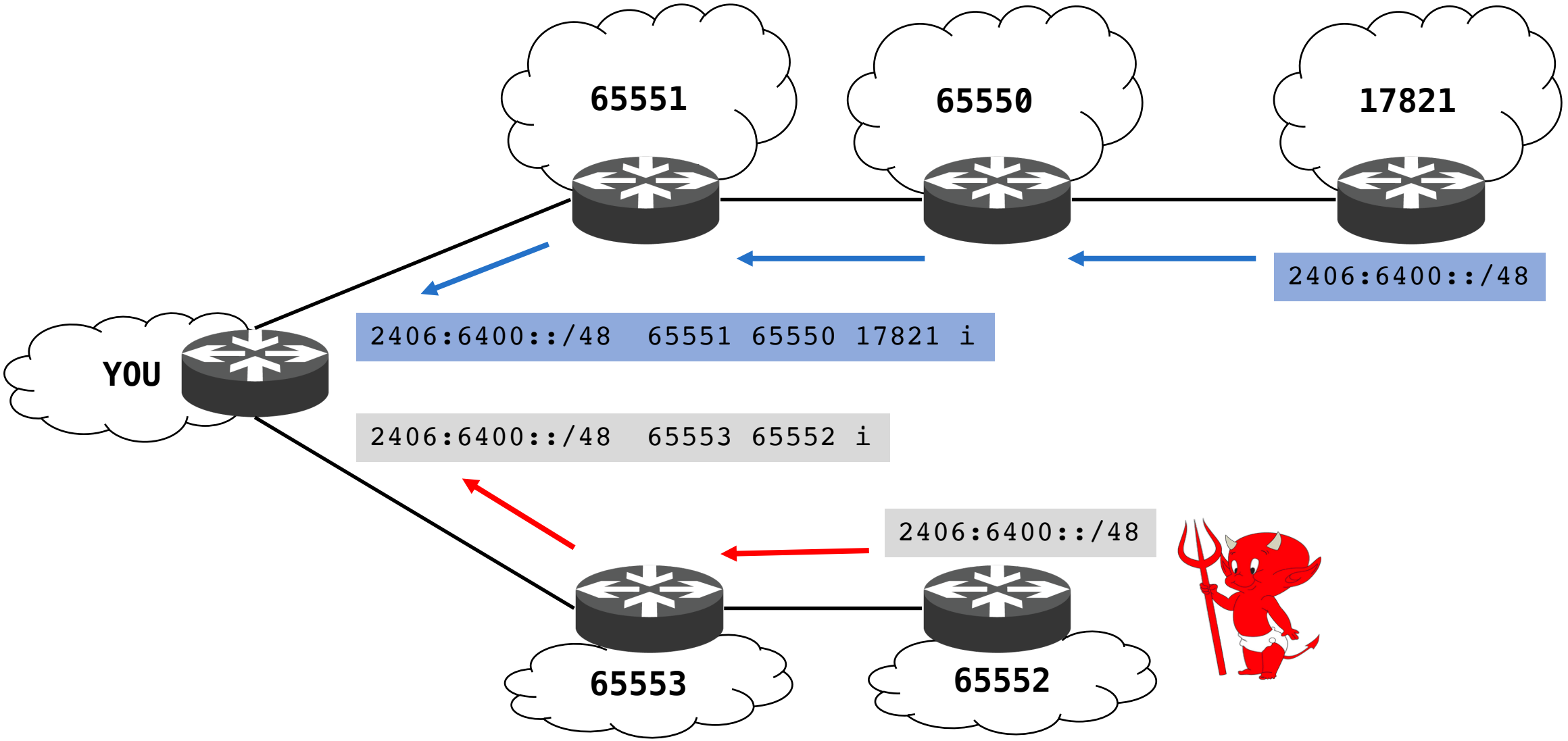


# RPKI Service Models

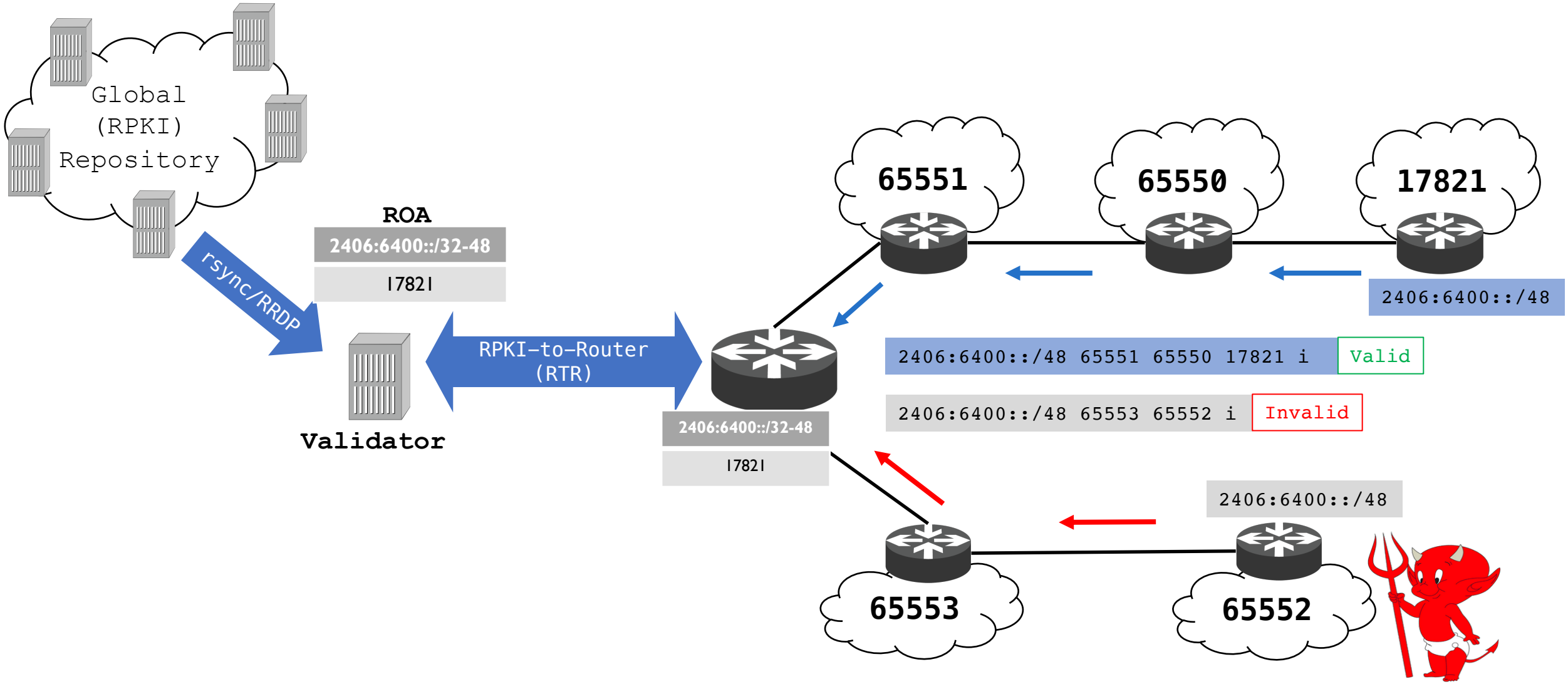
- Hosted model:
  - The RIR (APNIC) runs the CA functions on members' behalf
    - Manage keys, repo, etc.
    - Generate certificates for resource delegations
- Delegated model:
  - Member becomes the CA (delegated by the parent CA) and operates the full RPKI system
    - JPNIC, TWNIC, CNNIC (IDNIC in progress)


v1.0

# Route Origin Validation (ROV)



# Route Origin Validation (ROV)





v1.0

# 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


v1.0

# 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)


v1.0

# Validation States

**ROA**

| ASN | Prefix | Max Length |
|---|---|---|
| 65420 | 10.0.0.0/16 | 18 |

## BGP Routes

| ASN | Prefix | RPKI State |
|---|---|---|
| 65420 | 10.0.0.0/16 | VALID |
| 65420 | 10.0.128.0/17 | VALID |
| 65421 | 10.0.0.0/16 | INVALID |
| 65420 | 10.0.10.0/24 | INVALID |
| 65430 | 10.0.0.0/8 | NOT FOUND |


v1.0

# Possible actions - RPKI states

- Do Nothing
- Tag
  - If you have downstream customers or run a route server (IXP)
  - Ex:
    - `Valid (ASN:65XX1)`
    - `Not Found (ASN:65XX2)`
    - `Invalid (ASN:65XX3)`
- Modify preference values
  - RFC7115
- **Drop Invalids**
  - ~6K IPv4 & ~1.5K IPv6 routes (might want to check your top flows)
    https://rpki-monitor.antd.nist.gov/index.php?p=3&s=0

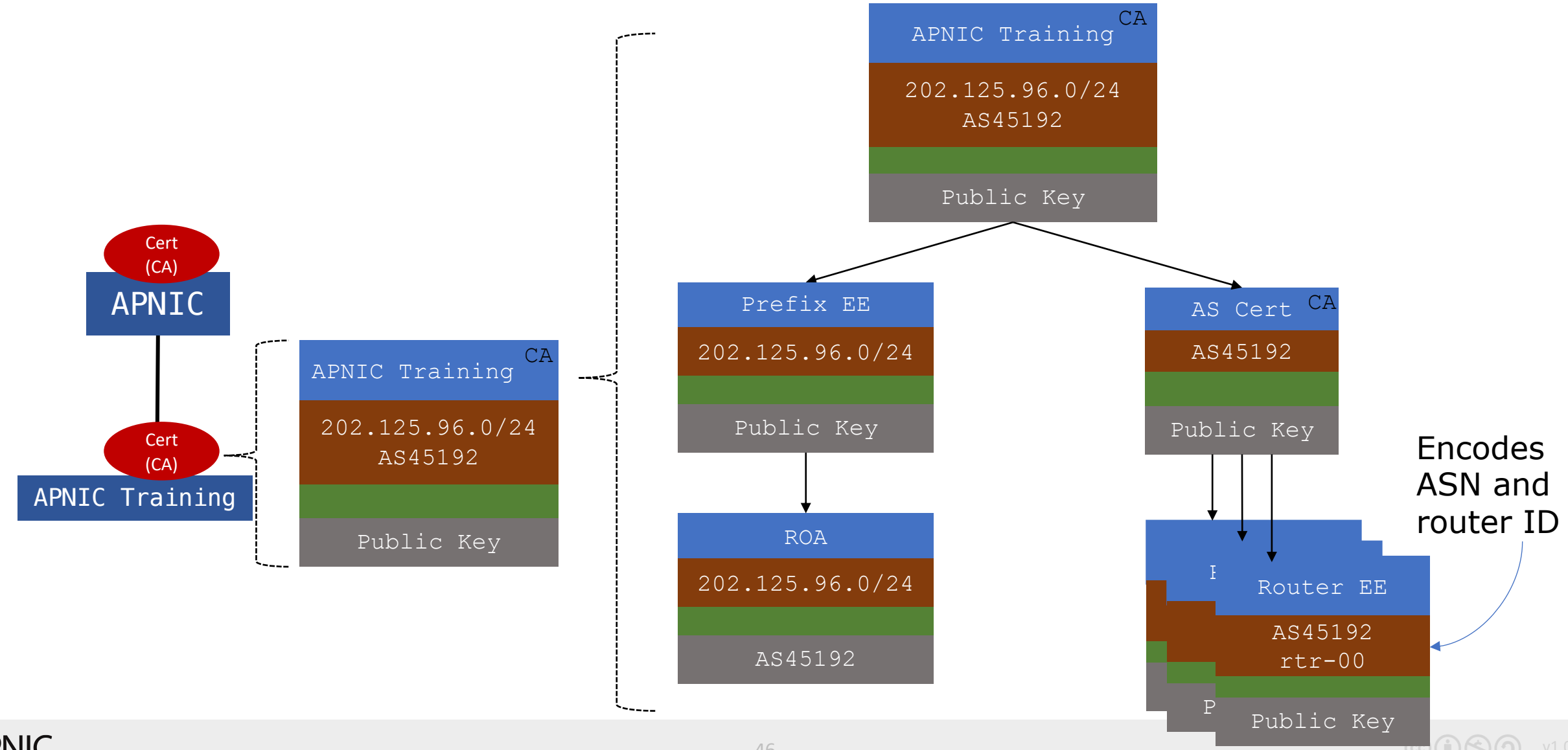

v1.0

# Are ROAs enough?

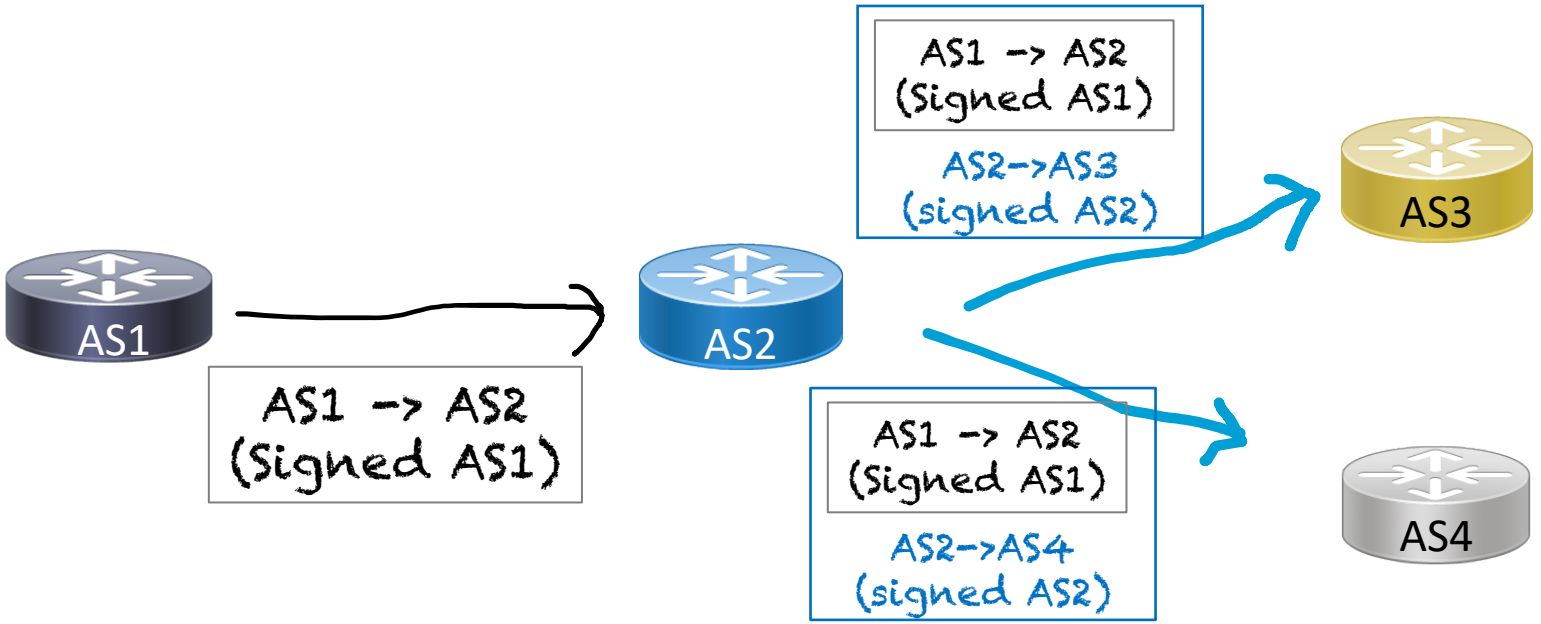
- What if I forge the origin AS in the AS path?
  - Would be accepted as good – pass origin validation!
- Which means, we need to secure the AS path as well
  - AS path validation (per-prefix)
- We can use RPKI certificates for this


v1.0

# AS keys (per-router keys)



# BGPsec (RFC8205)




- 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 NOT BGPsec?

- 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


v1.0

# AS Provider Authorization (draft but promising)

- ASPA is digitally signed object that binds
  - a Set of Provider ASNs to a Customer ASN (for a specific AFI),
- For Routing, the ASPA is an attestation
  - that the AS holder (Customer ASN) has authorized the set of provider ASNs to propagate its announcements onwards….


v1.0

# ASPA Validation/Verification ~ simplified

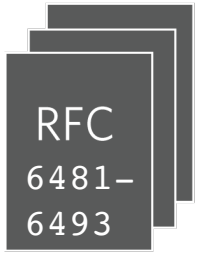
- For a received route (v4/v6):
  - If there is no valid ASPA for the Customer AS (AS(0)) UNKNOWN
  - If there is an ASPA with the customer AS, and:
    - if AS(I) in the AS_PATH attribute (AS_SEQ{AS(I), AS(I-1)}) is in the SPAS VALID
    - Else, INVALID


v1.0

# RPKI Further Reading

X.509 PKI Certificates

Extensions for IP Addresses and ASNs




Resource Public Key Infrastructure

# Acknowledgement


- Geoff Huston, APNIC
- Randy Bush, IIJ Labs/Arrcus

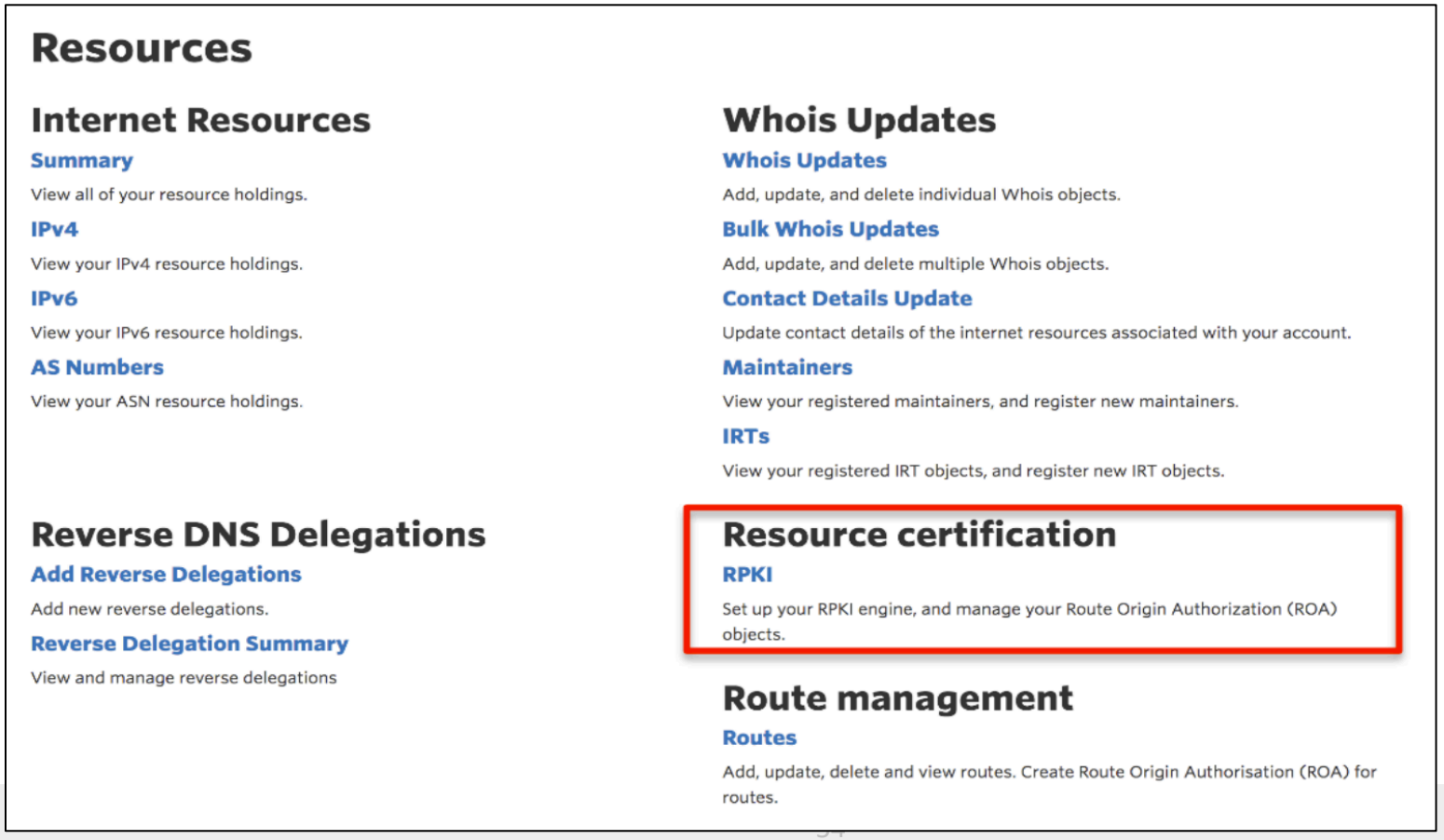


v1.0

# Implementation

# 1. Create & publish your ROA

- Login MyAPNIC
  - Need to activate the RPKI engine to create ROAs
  - Go to **Resources → Resource certification → RPKI** (see image below)

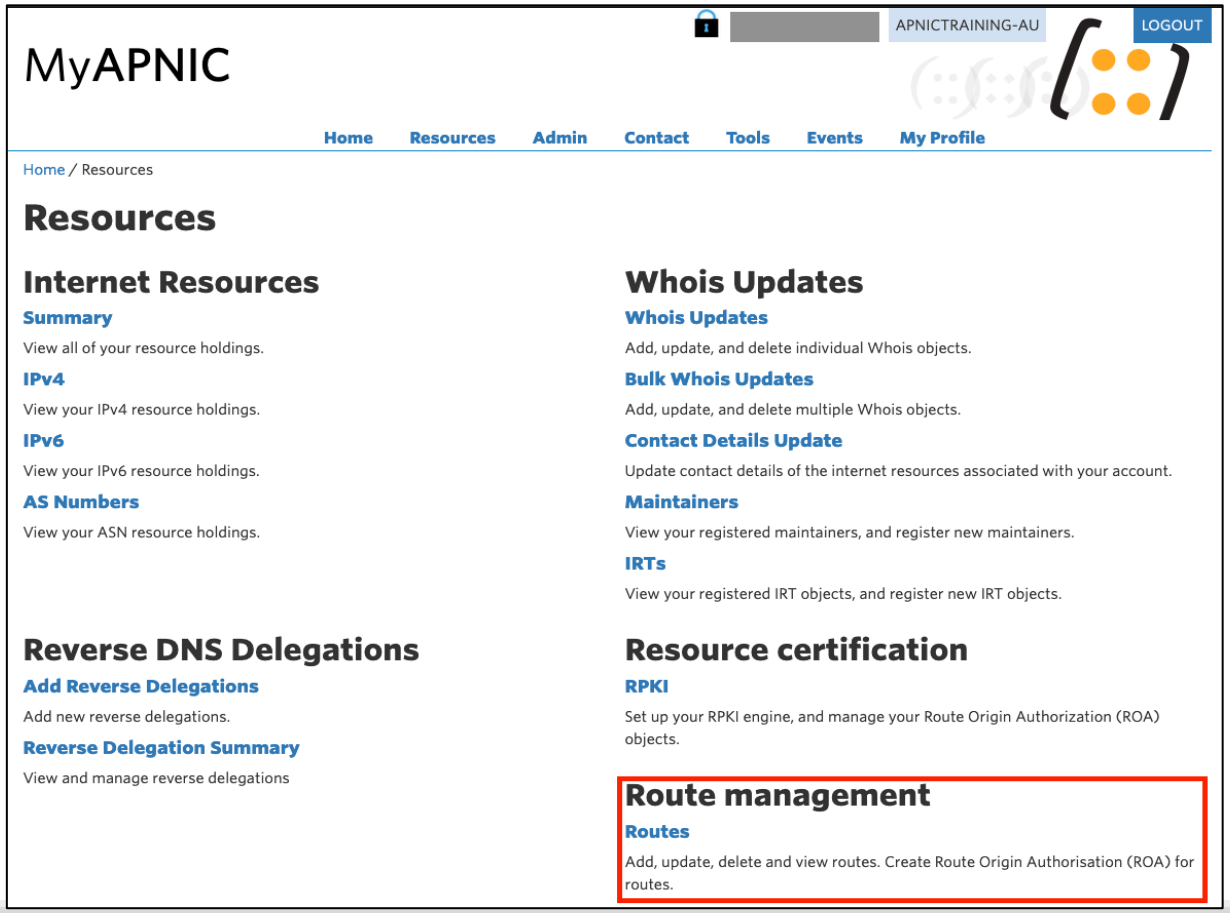




v1.0

# Create & publish your ROA

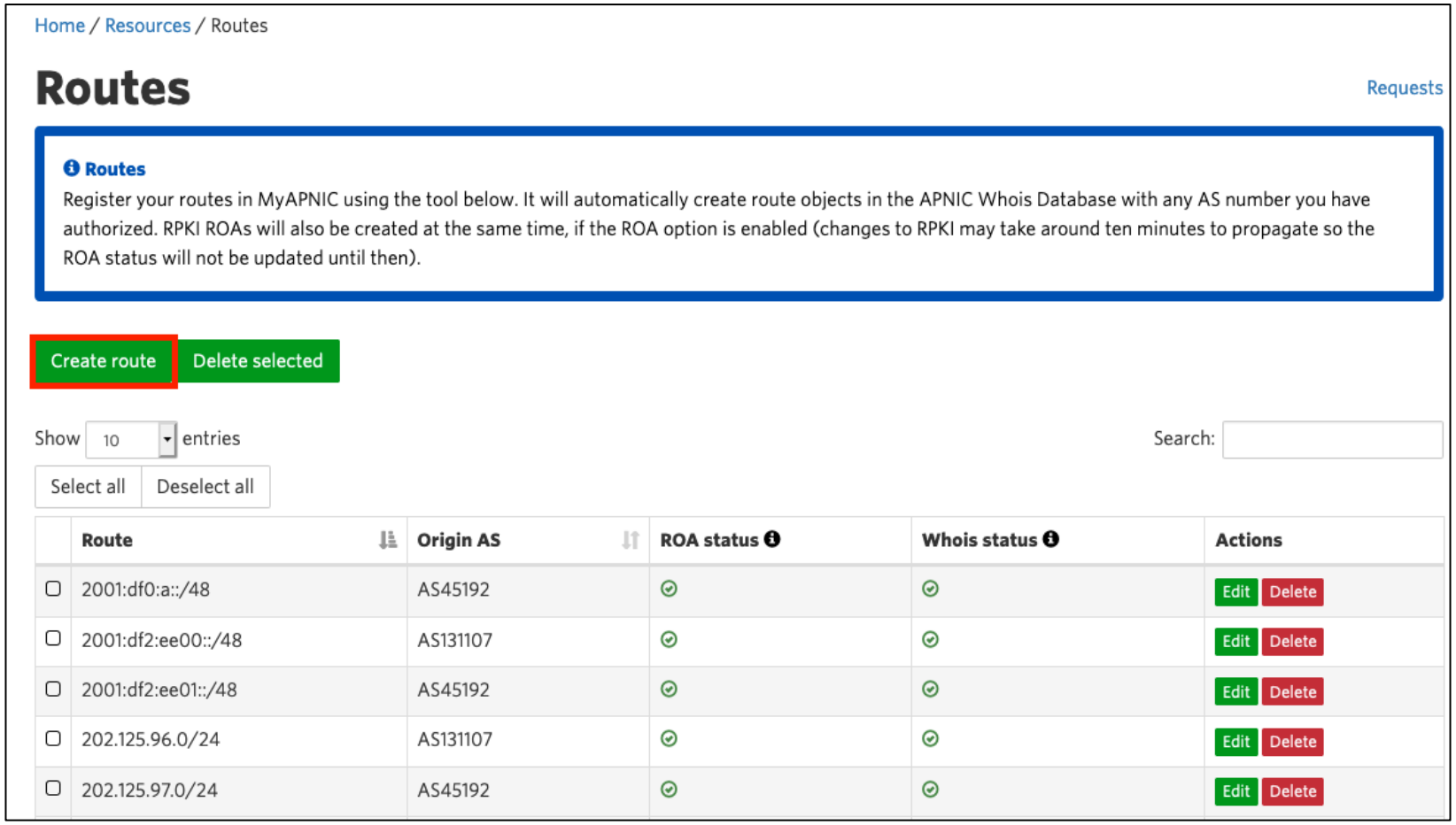
- Then go to the Routes page
  - Go to **Resources → Route Management → Routes** (see image below)




https://www.apnic.net/wp-content/uploads/2017/12/ROUTE_MANAGEMENT_GUIDE.pdf

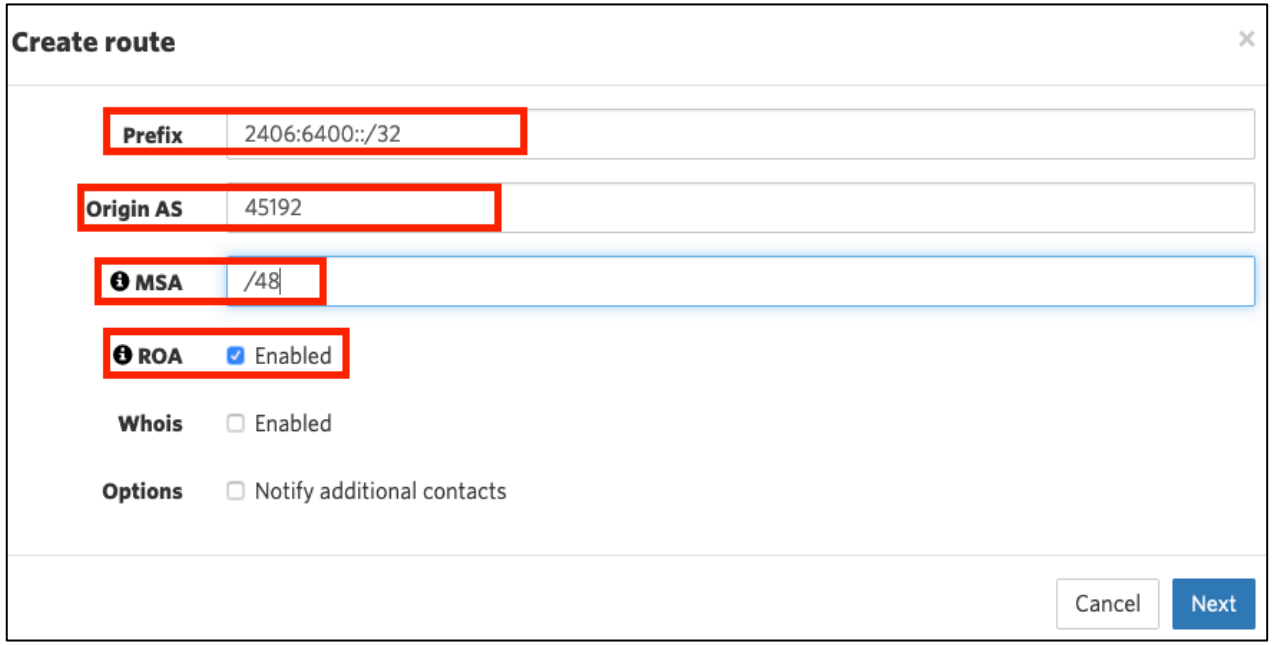
# Create (publish) your ROA

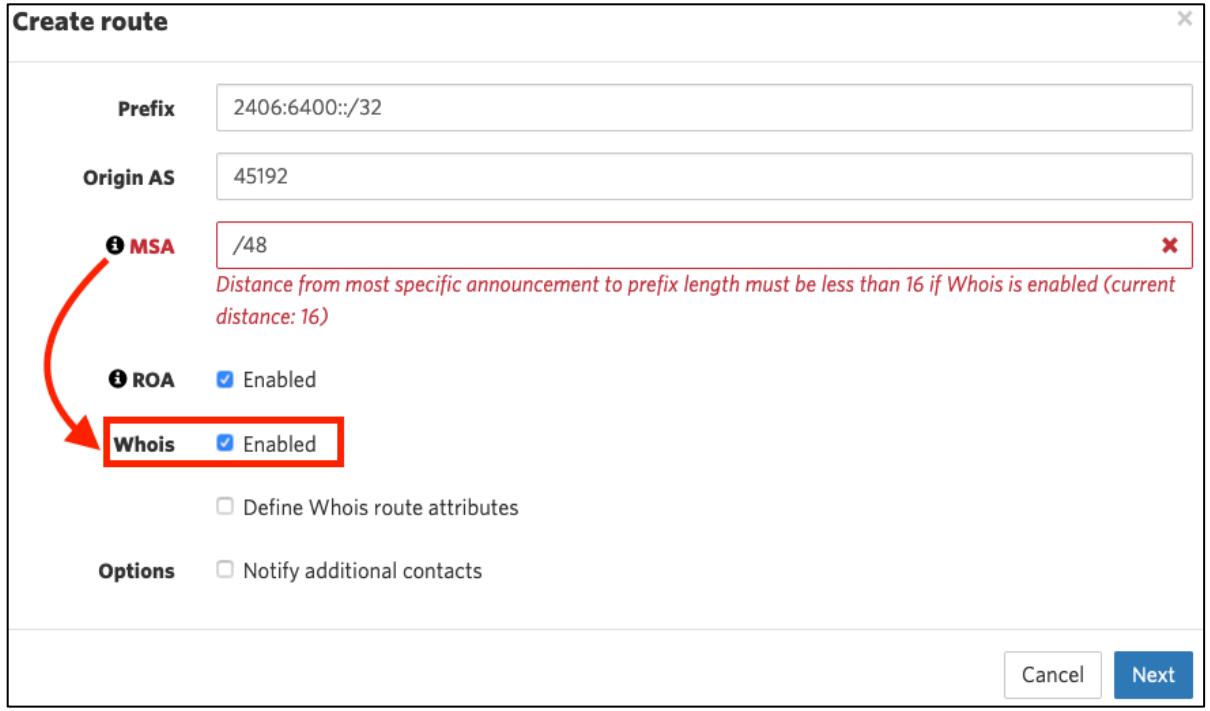
- Select **`Create route`** (as shown below)



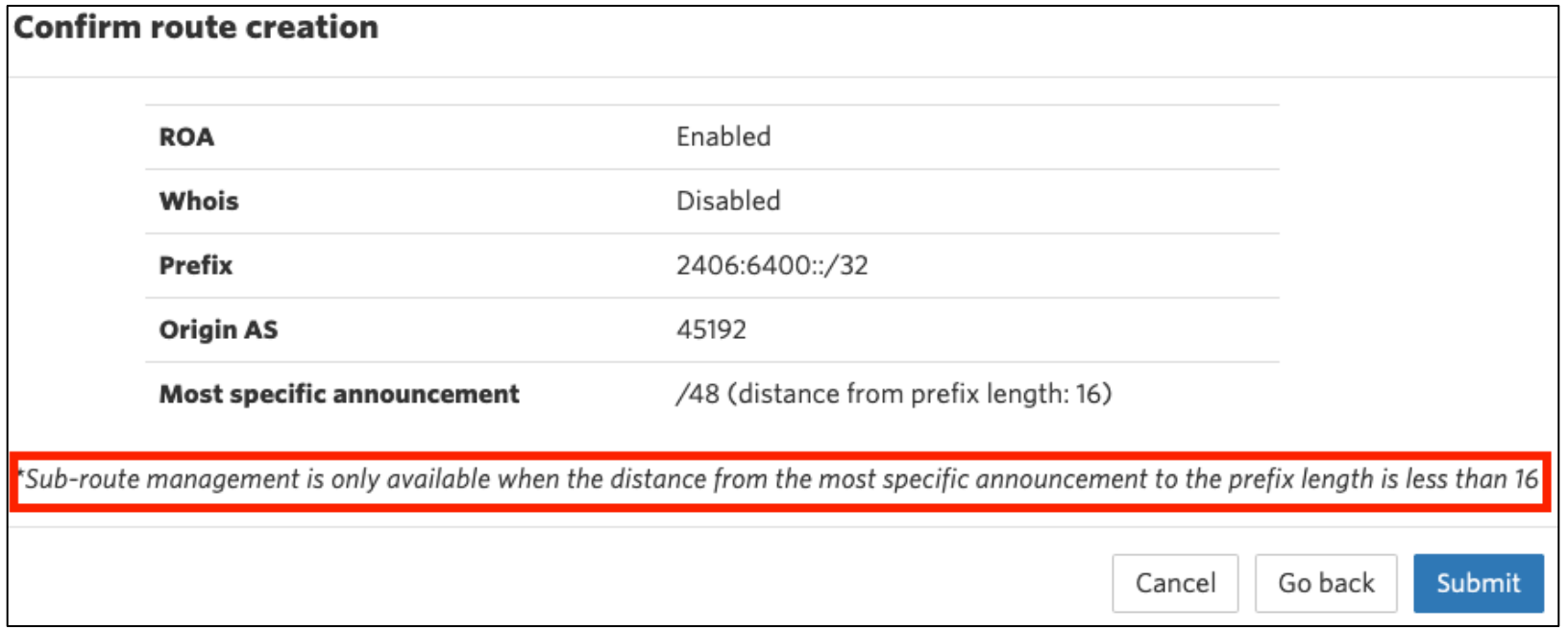
# Create (publish) your ROA

- Example for **IPv6** below



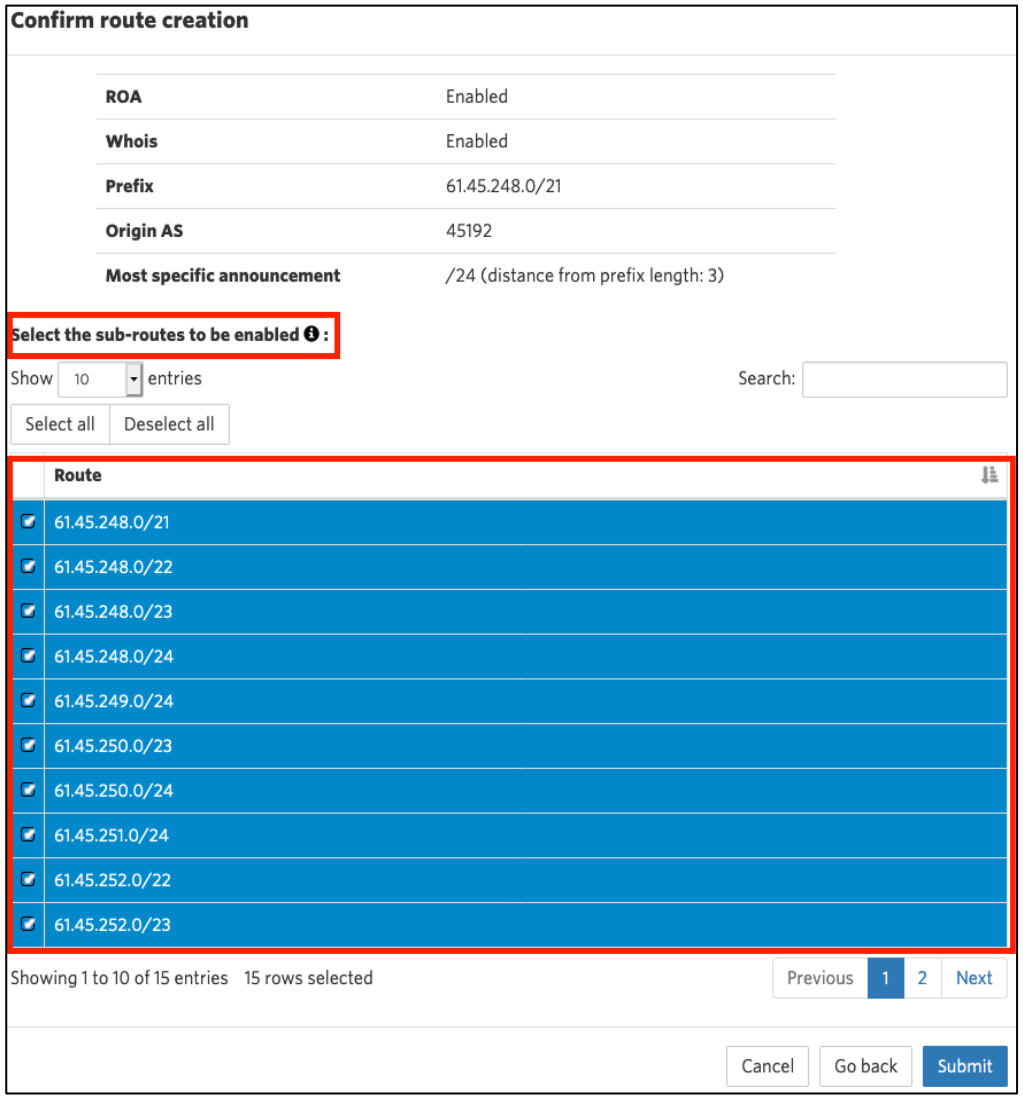



# Create (publish) your ROA



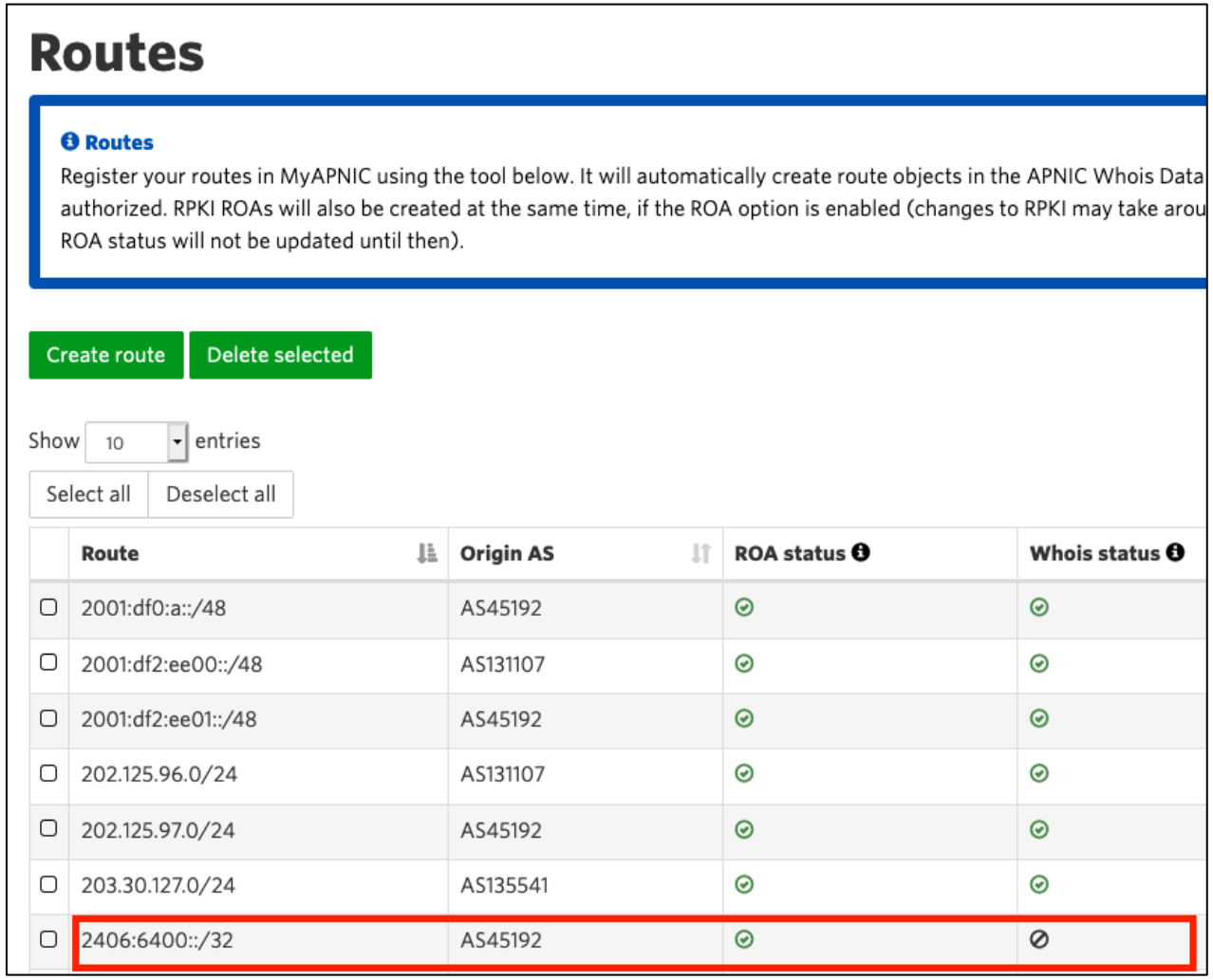
# Create (publish) your ROA

- 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**.

Show 10 entries

Search: 61.45.248.0

| ASN | Prefix | Maximum Length | Trust Anchor |
|---|---|---|---|
| 135533 | 61.45.248.0/24 | 24 | APNIC RPKI Root |

First Previous 1 Next Last

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


v1.0

# 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
```


v1.0

# 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"

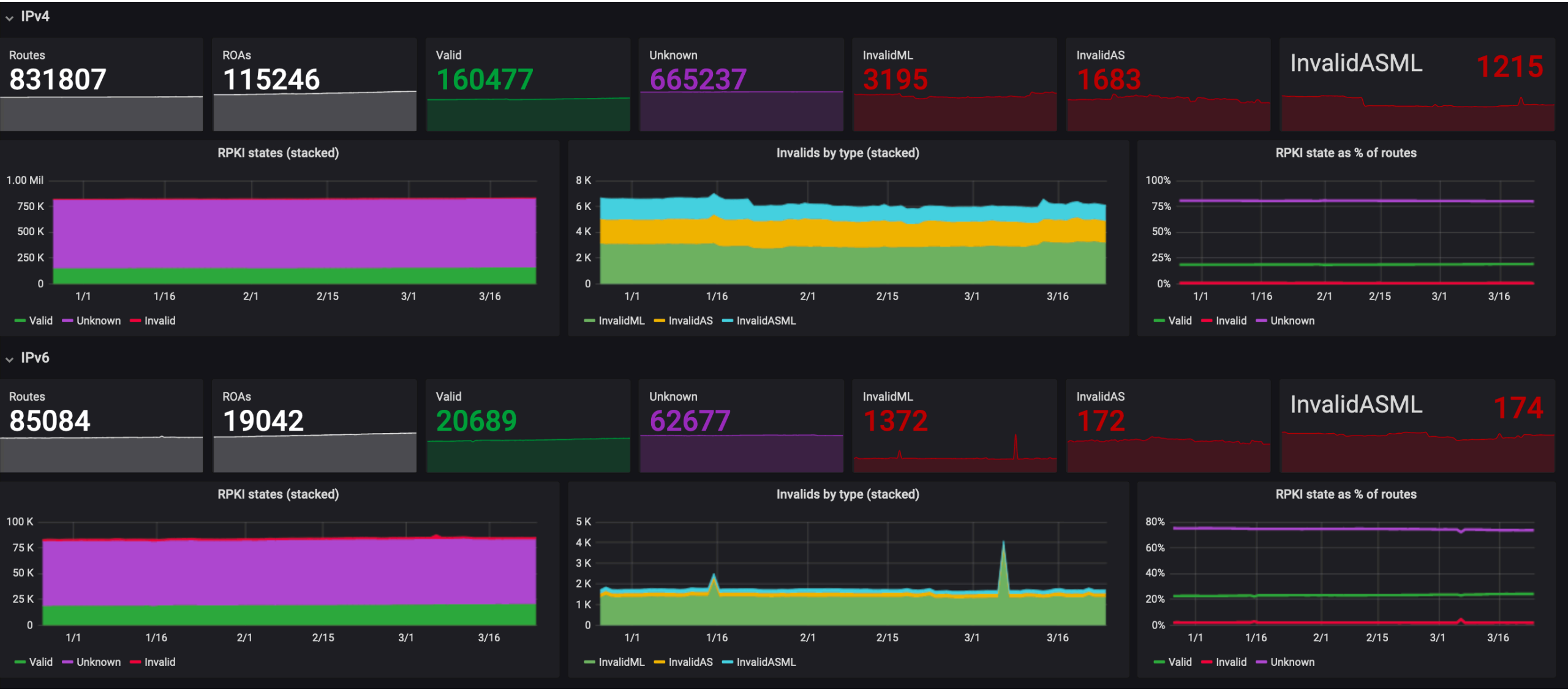
------------------------
ROA Details
------------------------
Origin ASN:       AS131107
Not valid Before: 2016-09-07 02:10:04
Not valid After:  2020-07-30 00:00:00  Expires in 2y190d9h34m23.2000000029802s
Trust Anchor:     rpki.apnic.net
Prefixes:         2001:df2:ee00::/48 (max length /48) 202.125.96.0/24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

# Check your ROA

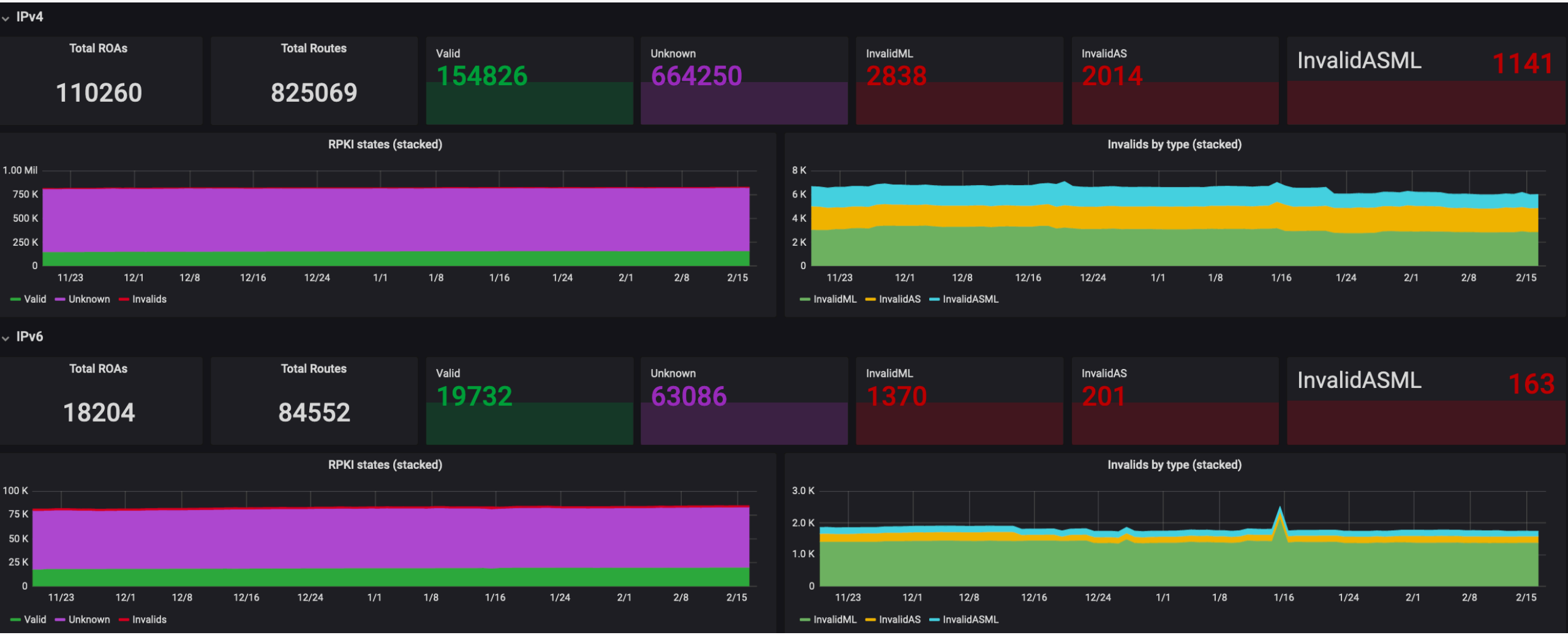
https://bgp.he.net/

| Announced By | | |
|---|---|---|
| **Origin AS** | **Announcement** | **Description** |
| AS131107 | 2001:df2:ee00::/48 | testing |

# Global ROA Status



# Global ROA Status ~ month ago



# Global ROA Status

### Total invalid counts by registry and protocol

| Registry | Proto | Count |
|---|---|---|
| afrinic | ipv4 | 63 |
| afrinic | ipv6 | 1 |
| apnic | ipv4 | 2295 |
| apnic | ipv6 | 706 |
| arin | ipv4 | 478 |
| arin | ipv6 | 56 |
| lacnic | ipv4 | 497 |
| lacnic | ipv6 | 199 |
| ripe | ipv4 | 1919 |
| ripe | ipv6 | 711 |

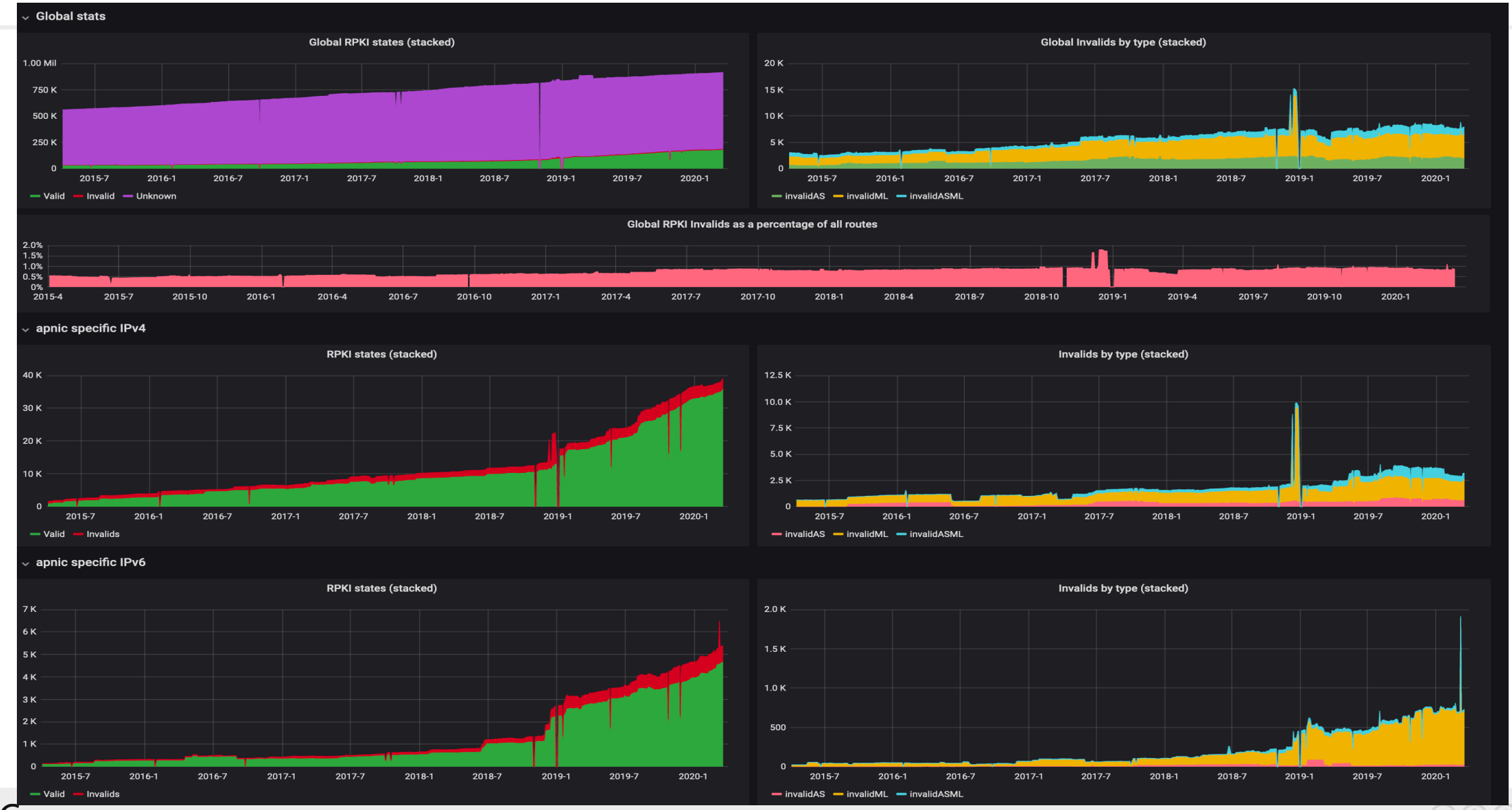
### Total invalid counts by NIR and protocol

| NIR | Proto | Count |
|---|---|---|
| cnnic | ipv4 | 43 |
| cnnic | ipv6 | 1 |
| idnic | ipv4 | 484 |
| idnic | ipv6 | 12 |
| irinn | ipv4 | 142 |
| irinn | ipv6 | 15 |
| jpnic | ipv4 | 3 |
| jpnic | ipv6 | 1 |
| twnic | ipv4 | 97 |
| twnic | ipv6 | 20 |
| vnnic | ipv4 | 125 |
| vnnic | ipv6 | 1 |



v1.0

# Rise of Invalids ☹



# 2. Deploy RPKI Validator

- Many options:
  - RIPE RPKI Validator

```
https://www.ripe.net/manage-ips-and-asns/resource-management/certification/tools-and-resources
```

  - Dragon Research Labs RPKI Toolkit

```
https://github.com/dragonresearch/rpki.net
```

  - Routinator

```
https://github.com/NLnetLabs/routinator
```

  - OctoRPKI & GoRTR (Cloudflare's RPKI toolkit)

```
https://github.com/cloudflare/cfrpki
```

  - Fort (NIC Mexico's Validator)

```
https://github.com/NICMx/FORT-validator
```

# 3. Router Configuration (IOS)

- Establishing session with the validator

```
router bgp 131107
 bgp rpki server tcp <validator-IP> port <323/8282/3323> refresh 120
```

- Note:
  - Cisco IOS by default does not include invalid routes for best path selection!
  - If you don’t want to drop invalids, we need explicitly tell BGP (under respective address families)

```
bgp bestpath prefix-validate allow-invalid
```


v1.0

# Configuration (IOS)

- 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:X6::151 activate
  neighbor X6:X6:X6:X6::151 route-map ROUTE-VALIDATION in
 exit-address-family
!
```

# 3. Router Configuration (JunOS)

- 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;
           }
       }
   }
}
```


v1.0

# Configuration (JunOS)

- Define policies based on the validation states

```
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;
           }
       }
   }
}
```


v1.0

# Router Configuration (JunOS)

- Apply the policy to inbound updates

```
protocols {
    bgp {
        group external-peers {
            #output-ommitted
            neighbor X.X.X.1 {
                import ROUTE-VALIDATION;
                family inet {
                    unicast;
                }
            }
        }
        group external-peers-v6 {
            #output-ommitted
            neighbor X6:X6:X6:X6::1 {
                import ROUTE-VALIDATION;
                family inet6 {
                    unicast;
                }
            }
        }
    }
```

# RPKI Verification (IOS)

- IOS has only

```
#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, Mininum 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

Network            Maxlen Origin-AS Source Neighbor
1.9.0.0/16         24     4788      0     202.125.96.47/323
1.9.12.0/24        24     65037     0     202.125.96.47/323
1.9.21.0/24        24     24514     0     202.125.96.47/323
1.9.23.0/24        24     65120     0     202.125.96.47/323
```

```
#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

Network             Maxlen Origin-AS Source Neighbor
2001:200::/32       32     2500      0     202.125.96.47/323
2001:200:136::/48   48     9367      0     202.125.96.47/323
2001:200:900::/40   40     7660      0     202.125.96.47/323
2001:200:8000::/35  35     4690      0     202.125.96.47/323
```


v1.0

# 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
     Community: 4826:1000 4826:2050 4826:2110 4826:2540 4826:2900 4826:5203
     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
```


v1.0

# RPKI Verification (JunOS)

- Check the RPKI cache

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

Prefix                Origin-AS       Session                   State  Mismatch
1.9.0.0/16-24              4788 202.125.96.46                valid
1.9.12.0/24-24            65037 202.125.96.46                valid
1.9.21.0/24-24            24514 202.125.96.46                valid
1.9.23.0/24-24            65120 202.125.96.46                valid

----------
2001:200::/32-32           2500 202.125.96.46                valid
2001:200:136::/48-48       9367 202.125.96.46                valid
2001:200:900::/40-40       7660 202.125.96.46                valid
2001:200:8000::/35-35      4690 202.125.96.46                valid
2001:200:c000::/35-35     23634 202.125.96.46                valid
2001:200:e000::/35-35      7660 202.125.96.46                valid
```

*Would have been nice if they had per AF!*


v1.0

# RPKI Verification (JunOS)

- Can filter per origin ASN

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


Prefix                Origin-AS       Session                  State        Mismatch
202.125.97.0/24-24       45192        202.125.96.46             valid
203.176.189.0/24-24      45192        202.125.96.46             valid
2001:df2:ee01::/48-48    45192        202.125.96.46             valid


 IPv4 records: 2
 IPv6 records: 1
```

*IOS should have something similar!*


v1.0

# 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

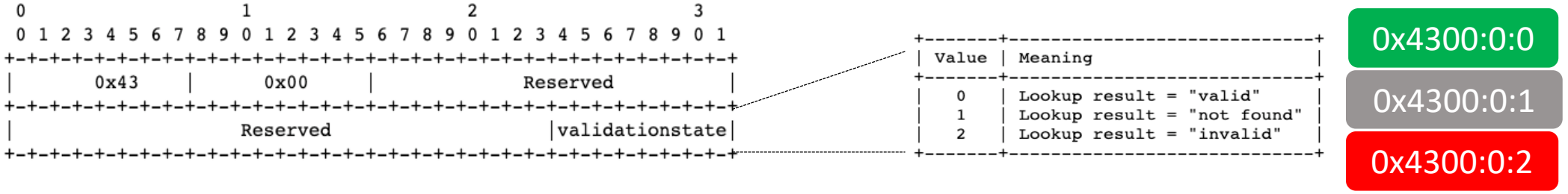
inet6.0: 93909 destinations, 93910 routes (93909 active, 0 holddown, 0
hidden)
+ = Active Route, - = Last Active, * = Both

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
```


v1.0

# Propagating RPKI states to iBGP peers

- To avoid every BGP speaker having an RTR session, and
- Ensure all BGP speakers have consistent information
  - Relies on non-transitive extended BGP community (RFC8097)




  - Sender (one with 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)

```
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::X6 activate
  neighbor X6:X6:X6:X6::X6 send-community both
  neighbor X6:X6:X6:X6::X6 announce rpki state
 exit-address-family
!
```


v1.0

# Propagating RPKI states (IOS)

- Receiver (iBGP peer)

```
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!


v1.0

# Propagating RPKI states (JunOS)

- Sender (one with RTR session)

```
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)

```
 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;
     }

}
```

# Operational Considerations

- Max-length
  - Make sure the value covers your BGP announcements

- minimal ROAs
  - Reduce spoofed origin-AS attack surface
    - https://tools.ietf.org/html/draft-ietf-sidrops-rpkimaxlen-03
    - ROAs should cover only those prefixes announced in BGP


v1.0

# Operational Considerations

- When RTR session goes down?
  - Validation state changes to Not Found for all routes (including Invalids) after a while (JunOS/SROS: 1hr, IOS-XE: 5mins)
    - at least two RTR sessions and careful filtering policies
  - What happens if the two validators don’t agree??


v1.0

# Operational Considerations

- During a router reload?
  - Will we receive ROAs first or BGP updates first?
    - If BGP update is faster than ROA, invalid routes to iBGP peers


v1.0

# Operational Considerations

- Default routes?
  - Even if you drop Invalids, default route will match anything


v1.0

# Operational Considerations

- iBGP State Propagation? ~ Multivendor
  - IOS as BR, propagating states to JunOS iBGP peers
    `unknown iana 4300`
  - Options(*<JunOS 17.4R3, 18.2R3, 18.4R2*):
    - Either act on the states at the border, or
    - Match and tag them with custom communities before propagating

# Operational Considerations

- iBGP State Propagation?
  - What if one eBGP router sends Valid tag, and the other sends Not Found:
    - If the RTR session is down on one eBGP speaker, and/or
    - Different ROA cache entry flush timers for different router vendors (JunOS/SROS: 1hr, IOS-XE: 5mins)
  - What is the router best path selection?
    - XE behaviour seems to pick validation state over other BGP attributes


v1.0

# Other developments

- ROA with AS-0 origin (RFC6483/RFC7607)
  - Reserved by IANA for non-routed networks
  - Negative attestation: no valid ASN has been granted authority
    - Not to be routed (Ex: IXP LAN prefixes)
    - Overridden by another ROA (with an origin-AS other than AS-0)
    - APNIC ~ Nov 2018
  - Prop-132: unallocated/unassigned APNIC space
    - ~ RFC6491 for special use/reserved/unallocated


v1.0

https://www.apnic.net/community/security/resource-certification/#routing


v1.0

# Any questions?