# Securing Internet Routing: The Puzzle Pieces

BPF2024 | 30 May 2024 | Bangkok

Tashi Phuntsho

Network Engineer/Trainer @NSRC





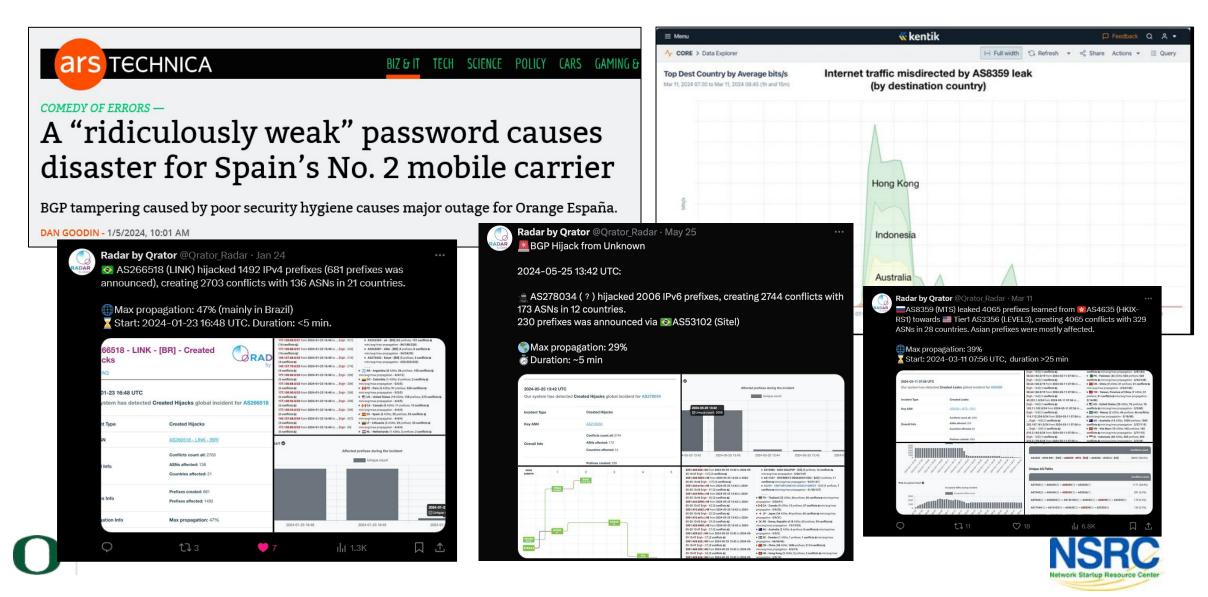
# Acknowledgment

- Slides/ideas from
  - Randy Bush (IIJ Labs/Arrcus)
  - Geoff Huston (APNIC)
  - Aftab Siddiqui (ISOC)
  - Job Snijders (Fastly)
  - Alexander Azimov (Yandex)
  - Alexander Lyamin (Qrator)
  - Yoshinobu Matsuzaki (IIJ/APNIC)





# Headlines/Incidents



- NO ONE is in charge?
  - No single authority point for the Internet
  - No REFERENCE point for what's RIGHT in routing





- Routing works by RUMOUR
  - TELL what you know to your neighbours/LEARN what your neighbours know
  - Assume everyone is CORRECT/HONEST
    - *Is the originating network the rightful owner?*





- Routing works in REVERSE
  - What you TELL others (outbound adv) affects inbound traffic
  - What you TRUST and ACCEPT (inbound adv) affects outbound traffic





- And sadly, there is no EVIL (E-bit) bit
  - RFC3514 was a humorous attempt

- Since a bad routing update does not identify itself as BAD:
  - Can we identify GOOD updates?
  - How do we identify what is GOOD?





# Identifying GOOD

- Back to basics can we use Digital Signatures 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/authorise the use of the resource(s)!





# Identifying GOOD

 Ok, let us use digital signatures, but how do we establish TRUST in this framework?

- Follow the numbered resource allocation

hierarchy

Regional Internet Registries
(RIRs)

CERT (CA)

National IRs

CERT (EE)

ISP

ISP



• WHOIS lookup - to verify the holder of a resource(s)

```
whois -h whois.apnic.net 202.144.128.0
% [whois.apnic.net]
% Whois data copyright terms http://www.apnic.net/db/dbcopyright.html
% Information related to '202.144.128.0 - 202.144.129.255'
% Abuse contact for '202.144.128.0 - 202.144.129.255' is 'systems@bt.bt'
                202.144.128.0 - 202.144.129.255
inetnum:
netname:
                DRUKNET
                DrukNet System
descr:
descr:
                DrukNet
                Bhutan Telecom
descr:
descr:
                Thimphu
country:
                JT106-AP
admin-c:
tech-c:
                JT106-AP
abuse-c:
                AB1276-AP
status:
                ASSIGNED NON-PORTABLE
mnt-by:
                MAINT-BT-DRUKNET
                IRT-BTTELECOM-BT
mnt-irt:
last-modified: 2021-01-14T06:15:57Z
                APNIC
source:
```

```
% Information related to 'AS18024'
% Abuse contact for 'AS18024' is 'systems@bt.bt'
                AS18024
aut-num:
                BTTELECOM-AS-AP
as-name:
                Bhutan Telecom Ltd
descr:
country:
                ORG-BTL2-AP
                DN01-AP
admin-c:
tech-c:
                DN01-AP
                AB1276-AP
abuse-c:
                MAINT-BT-DRUKNET
mnt-lower:
mnt-routes:
                MAINT-BT-DRUKNET
                APNIC-HM
mnt-by:
                IRT-BTTELECOM-BT
mnt-irt:
last-modified: 2021-01-14T06:16:00Z
source:
                APNIC
```

```
% Information related to '202.144.128.0/20AS18024'
                202.144.128.0/20
route:
descr:
                 DRUKNET-BLOCK-A1
                 BT
country:
notify:
                 ioc@bt.bt
mnt-by:
                MAINT-BT-DRUKNET
origin:
                 AS18024
last-modified: 2018-09-18T09:37:40Z
                 APNIC
source:
                                                               Network Startup Resource Center
```



# • <u>IRR</u> (Internet Routing Registry) lookup

- Publish my routing intent (route origination) and in some cases, inter-AS routing policies

```
whois -h whois.radb.net 202.144.128.0
            202.144.128.0/23
route:
descr:
           DRUKNET-VSNL Route Object
            AS17660
origin:
mnt-by:
           MAINT-VSNL-IN
changed:
            ip.admin@vsnl.co.in 20070102
            RADB
source:
                202.144.128.0/20
route:
                DRUKNET-BLOCK-A1
descr:
                BT
country:
notify:
                ioc@bt.bt
mnt-by:
                MAINT-BT-DRUKNET
                AS18024
origin:
last-modified: 2018-09-18T09:37:40Z
source:
                APNIC
```

```
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:
               0THER
mnt-bv:
               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
               * 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:
                *********
remarks:
aut-num:
               AS17660
               DRUKNET-AS
as-name:
               DrukNet ISP
descr:
descr:
               Bhutan Telecom
descr:
               Thimphu
country:
               BT
import:
               from AS6461 action pref=100; accept ANY
               to AS6461 announce AS-DRUKNET-TRANSIT
export:
import:
               from AS2914 action pref=150; accept ANY
               to AS2914 announce AS-DRUKNET-TRANSIT
export:
               from AS6453 action pref=100; accept ANY
import:
               to AS6453 announce AS-DRUKNET-TRANSIT
export:
               from AS42 action pref=250; accept AS42
import:
```

- IRR (Internet Routing Registry) entries
  - Helps craft route filters (prefix/as-path)

with RPSL tools (rtconfig/bgpq3-4)

```
bgpq4 -bl PEERv4-IN AS17660
PEERv4-IN = [
    45.64.248.0/22,
    103.245.240.0/22,
    103.245.242.0/23,
    119.2.96.0/19,
    202.144.128.0/19,
    202.144.128.0/20,
    202.144.128.0/23,
    202.144.144.0/20,
    202.144.148.0/22
   bgpg4 -S APNIC -bl PEERv4-IN AS17660
PEERv4-IN = [
    45.64.248.0/22,
    103.245.240.0/22,
    103.245.242.0/23,
    119.2.96.0/19,
    202.144.128.0/19
     bgpq4 -6bl PEERv6-IN AS17660
    PEERv6-IN = [
        2405:d000::/32.
        2405:d000:7000::/36

→ bgpq4 -S APNIC -6bl PEERv6-IN AS17660

     PEERv6-IN = [
        2405:d000::/32.
        2405:d000:7000::/36
```

```
bapa4 -l BTv4-IN AS-DRUKNET-TRANSIT

    bgpg4 −6l BTv6−IN AS−DRUKNET−TRANSIT

no ip prefix-list BTv4-IN
                                                    no ipv6 prefix-list BTv6-IN
ip prefix-list BTv4-IN permit 27.123.224.0/19
                                                    ipv6 prefix-list BTv6-IN permit 2001:df3:e180::/48
ip prefix-list BTv4-IN permit 27.123.224.0/22
                                                    ipv6 prefix-list BTv6-IN permit 2001:df5:a300::/48
ip prefix-list BTv4-IN permit 27.124.64.0/20
                                                    ipv6 prefix-list BTv6-IN permit 2400:1440::/32
ip prefix-list BTv4-IN permit 27.124.64.0/22
                                                    ipv6 prefix-list BTv6-IN permit 2400:4e60::/32
ip prefix-list BTv4-IN permit 27.124.68.0/22
                                                    ipv6 prefix-list BTv6-IN permit 2400:4e60::/33
ip prefix-list BTv4-IN permit 27.124.72.0/22
                                                    ipv6 prefix-list BTv6-IN permit 2400:4e60:8000::/33
ip prefix-list BTv4-IN permit 27.124.76.0/22
                                                    ipv6 prefix-list BTv6-IN permit 2403:580::/32
ip prefix-list BTv4-IN permit 43.230.208.0/24
                                                    ipv6 prefix-list BTv6-IN permit 2403:580::/33
ip prefix-list BTv4-IN permit 45.64.248.0/22
                                                    ipv6 prefix-list BTv6-IN permit 2403:580:8000::/33
ip prefix-list BTv4-IN permit 45.64.248.0/23
                                                    ipv6 prefix-list BTv6-IN permit 2403:8700::/32
ip prefix-list BTv4-IN permit 45.64.250.0/24
                                                    ipv6 prefix-list BTv6-IN permit 2404:5540::/32
ip prefix-list BTv4-IN permit 45.64.251.0/24
                                                    ipv6 prefix-list BTv6-IN permit 2404:5540::/33
ip prefix-list BTv4-IN permit 103.7.252.0/22
  prefix-list BTv4-IN permit 103.10.236.0/22
                                                    ipv6 prefix-list BTv6-IN permit 2404:5540::/34
               bapa3 -3f 17660 -l BT-IN AS-DRUKNET-TRANSIT
            no ip as-path access-list BT-IN
            ip as-path access-list BT-IN permit ^17660(_17660)*$
            ip as-path access-list BT-IN permit ^17660(_[0-9]+)*_(18024|18025|59219|132232)$
            ip as-path access-list BT-IN permit ^17660(_[0-9]+)*_(134715|135666|137925|137994)$
            ip as-path access-list BT-IN permit ^17660( [0-9]+)* (140695)$

    □ bgpg4 -f 1/660 -l BT-IN AS-DRUKNET-TRANSIT

            no ip as-path access-list BT-IN
```

ip as-path access-list BT-IN permit ^17660(\_[0-9]+)\*\_(18024|18025|59219|132232)\$

ip as-path access-list BT-IN permit ^17660( [0-9]+)\* (134715|135666|137925|137994)\$

ip as-path access-list BT-IN permit ^17660( 17660)\*\$

ip as-path access-list BT-IN permit ^17660( [0-9]+)\* (140695)\$

- Issues with IRR
  - No single authority model
    - Is an entry genuine/correct?
  - Too many RRs
    - If two RRs contain conflicting data which one to use/trust?
  - Incomplete data
    - If a route is not in a RR ~ invalid or is the RR just missing data?

- Issues with IRR Filters
  - Your filters ONLY as good as the correctness of the IRR entries!
    - GOOD idea to rely on authoritative sources:
      - -S in bgpq3/4, or -s in rtconfig





# Aside – IRR improvements

- prop-151 (Aftab): <u>restricting</u> non-hierarchical as-set
  - Helps fix name collision issues
  - as-set can ONLY be created by the maintainer of the ASN in the object

- Hierarchical as-set (RFC2622)
  - AS-DRUKNET-TRANSIT
    - non-hierarchical as-set
  - AS4826:AS-VOCUS
    - hierarchical as-set
    - <AS#>:AS-<as\_set\_name>

```
AS-AMAZON
as-set:
            Amazon ASNs
descr:
            AS-AMAZON-NA, AS-AMAZON-AP, AS-AMAZON-EU, AS16509:AS-AMAZON
members:
admin-c:
            AC6-ORG-ARIN
            AC6-ORG-ARIN
tech-c:
notify:
            noc@amazon.com
mnt-by:
            MAINT-AS16509
changed:
            noc@amazon.com 20230420 #17:54:10Z
source:
                AS-AMAZON
as-set:
tech-c:
                DUMY-RIPE
admin-c:
                DUMY-RIPE
mnt-by:
                KATERINA-MNT
                2022-10-23T19:05:59Z
created:
last-modified: 2022-10-23T19:05:59Z
                RTPF
```

```
as-set:
                AS4826: AS-VOCUS
                Vocus Communications AS4826 AS-SET
descr:
members:
                AS4826, AS4826: AS-CUSTOMERS
admin-c:
                VPL1-AP
tech-c:
remarks:
                For queries please email the below contacts
remarks:
                NOC - ******
remarks:
                IRR Data - ******
remarks:
                Peering enquiries - ******
mnt-by:
                MAINT-AU-VOCUS
last-modified:
                2022-05-29T00:28:23Z
                APNIC
source:
```





# Aside – IRR improvements

#### RADB & RPKI

- RADB migrated to IRRDv4 on 13<sup>th</sup> November 2023
- New RPKI based features implemented
  - route/route6 objects that is inconsistent with a corresponding ROA will be rejected
  - RPKI Invalid objects will no longer be visible in a query
  - Not Found or Valid will not be affected

Prefix: 1.1.1.0/24

ASN: 13335

Route: 1.1.1.0/24 Origin: AS13335 Source: RADB

Route: 1.1.1.0/25 Origin: AS13335 Source: RADB

Route: 1.1.1.0/25 Origin: AS12345 Source: RADB





#### Route Origin Authorization (ROA)

- Binding of prefixes & nominated ASN
- Can be verified crypto-magically

• Multiple ROAs can exist for the same prefix

Prefix	202.144.128.0/20
Max-length	/20
Origin ASN	AS18024

route: 202.144.128.0/20

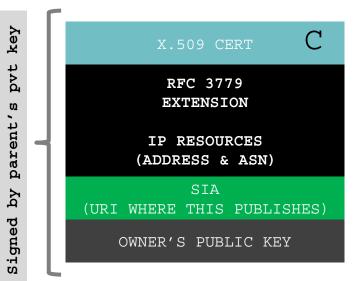
descr: RPKI ROA for 202.144.128.0/20 / AS18024

remarks: This AS18024 route object represents routing data retrieved from the RPKI. This route object is the result of an automated RPKI-to-IRR conversion process performed by IRRd.

max-length: 20

origin: AS18024

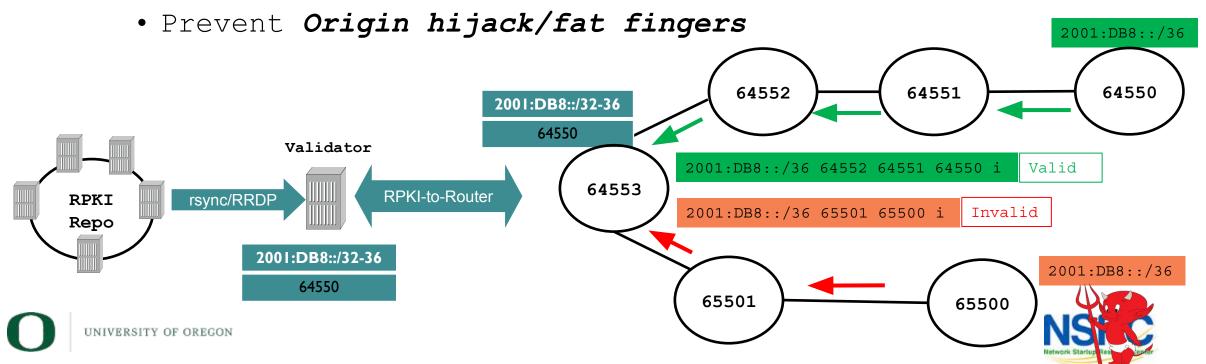
source: RPKI # Trust Anchor: apnic







- Route Origin Validation (ROV)
  - Validating received routes against validated ROAs
  - What can it help with?
    - Validate if an ASN is permitted to originate a route



#### ROA BCPs

- Use <u>max-length</u> judiciously
  - Only cover those prefixes announced in BGP ~ minimal ROA RFC9319
- Multi-ASN network?
  - Aggregates/sub-aggs: Transit ASN
  - More specifics: Access ASN
- ROA with **ASO** origin (RFC7607)
  - Not to be confused with undelegated/unassigned ASO ROA

- ROV BCPs
  - Default routes?
  - Secure the RTR session
    - SSH/MD5/TLS/TCP-AO/TLS
  - iBGP propagation RFC8097
  - Know your platform:
    - RTR refresh timer □ route refresh (Adj RIB In or soft reconfig in)

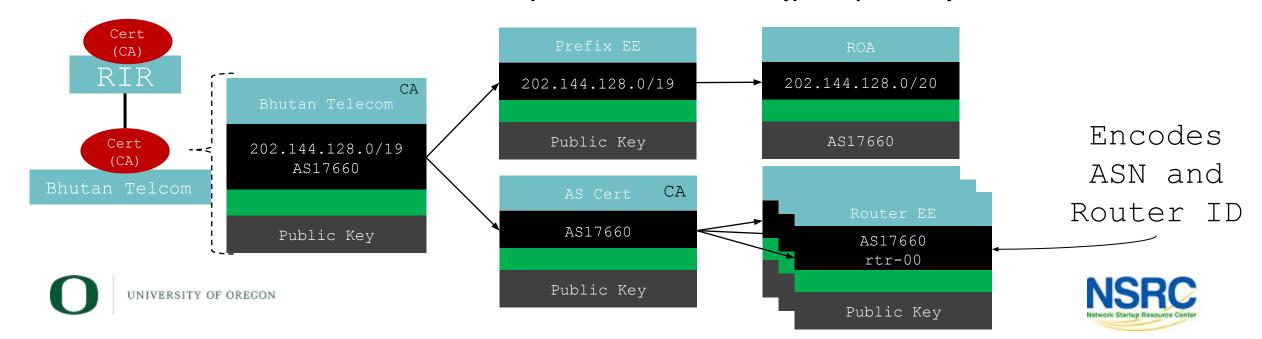
https://blog.apnic.net/2022/04/04/rpki-2021-retrospective/

https://blog.apnic.net/2020/04/10/rise-of-the-invalids/





- Are ROAs and ROV enough?
  - Forged origin ASN: will PASS the ROV test & will be accepted as GOOD
- Ideas?
  - Secure the PATH ~ <u>AS path validation</u> (per prefix) □ BGPsec



AS1 -> AS2

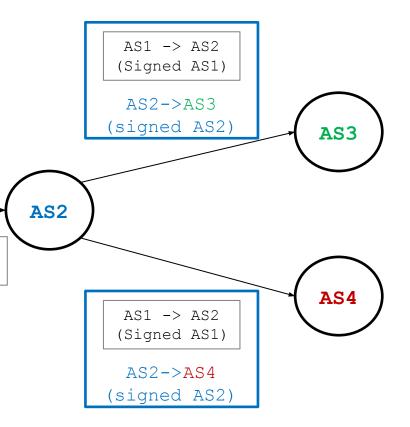
(Signed AS1)

#### BGPsec (RFC8205)

- Forward Path Signing
  - AS1 signs the message to AS2
  - AS2 signs the message to AS3/AS4, encapsulating AS1's message

#### Validation

- ROA check for the prefix and origin AS
- validate the received AS path against the chain of signatures (for each AS in the AS path) with AS key





- BGPsec (RFC8205) <u>Challenges</u>
  - Cannot jump across non-BGPsec routers/networks
    - traditional BGP (no BGPsec UPDATE messages)
  - Complex crypto & key distribution mechanism
    - CPU intensive (validate signatures)
    - Memory intensive (per prefix BGPsec UPDATE; new attributes to carry signatures and certs/key IDs for every AS in the AS path)
  - Possible hack
    - Routers could generate key pair -> send cert request to RPKI for signing
  - Lack of clarity
  - distributing the collection of certs required to yalidate path signature





- Route leak prevention
  - We already talked whitelist of customer/peer prefixes under IRR filtering
  - Don't announce routes/prefixes learned from your peers to other peers
  - Apply max prefix limits ~ doesn't help against partial leaks.



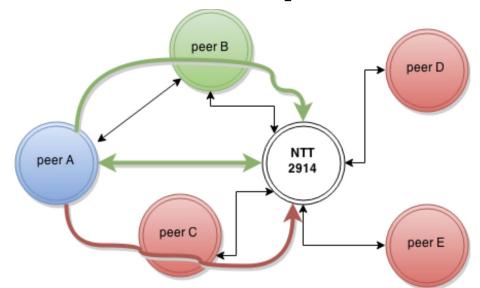


- Peerlock-lite ~ adapted from Job's NANOG67
  - Wikipedia says [7018, 7922, 3320, 3257, 6830, 3356, 2914, 5511, 3491, 1239, 6453, 6762, 1299, 12956, 701, 6461]
    - https://en.wikipedia.org/wiki/Tier 1 network
  - Will you sell transit to these networks?
    - REJECT any prefixes you receive from your customers which contains a big network ASN anywhere in the AS\_PATH





- Peerlock~ adapted from Job's NANOG67 talk
  - Given ASNs A, B, C, D, and E as NTT's peers.
  - Peer A subscribes to the peerlock idea (Protected ASN) and indicates that peer B is an "Allowed Upstream"



OK: ^A\_

OK: ^B A

NOT OK: ^C\_A\_

NOT OK: ^D\_A\_

NOT OK: ^E\_A\_





#### BGP Roles

- Update to the BGP OPEN message ~ BGP Role Capability
- Must be advertised to and received from a peer
  - If advertised and but not received: SHOULD ignore and establish traditional session
  - Strict mode: if advertised and not received REJECT

#### • Roles:

- Provider | Customer | Peer | RS | RS-client

#### Allowed relationship pairs

- Provider <-> Customer
- Customer <-> Provider
- RS <-> RS-Client
- RS-Client <-> RS
- Peer <-> Peer

#### **BIRD**

```
protocol bgp {
    local as 65001;
    neighbor 127.20.0.1 as 65000;
    multihop;
    source address 127.20.0.2;
    strict bind on;
    ipv4 {
        import all;
        export all;
    };
    local role customer;
}
```

#### FRR

```
router bgp 64502
neighbor 172.16.200.101 remote-as 64501
neighbor 172.16.200.101 ebgp-multihop
neighbor 172.16.200.101 passive
neighbor 172.16.200.101 local-role customer
```

https://blog.grator.net/en/route-leak-prevention-and-detection-rfc9234 162/





#### BGP Roles

- Only to Customer (OTC) attribute
  - Optional non-transitive attribute
- Ingress procedure:
  - If a route with the OTC Attribute is received from a Customer or an RS-Client, then it is a route leak and MUST be considered ineligible.

#### - Egress procedure:

• If a route contains the OTC
Attribute, it MUST NOT be
propagated to Providers, Peers,

Solution Status Version

BIRD + Appeared in 2.0.11

FRR + Appeared in 8.4

OpenBGPD + 7.5

Mikrotik Reduced functionality Appeared before RFC

https://blog.grator.net/en/route-leak-prevention-and-detection-rfc9234 162/





- ASPA (AS Provider Authorization)
  - Looks at malformed AS\_PATHs from customers and peers to detect malicious hijacks and route leaks
  - ASPA is a digitally signed object that binds
    - Set of Provider ASNs (SPAS) to a Customer ASN (CAS) for a specific AFI signed by the holder of the Customer ASN
  - For Routing, the ASPA is an attestation
    - that the AS holder (CAS) has authorized the SPAS to propagate its announcements onwards (upstreams/peers)





• ASPA (AS Provider Authorization) object

```
ASPA := {
    customer_asn (signer)
    providers (authorized to propagate to peers/upstreams)
    AFI (IPv4/IPv6)
}
```





#### Pair Verification (AS1, AS2)

- Retrieve cryptographically valid ASPA in a selected AFI with a customer value of AS1.
- If there is no valid ASPA record for AS1 the procedure exits with an outcome of **unknown**
- If AS2 is included in the SPAS, then the procedure exits with an outcome of **valid**
- Otherwise, the procedure exits with an outcome of invalid





#### • ASPA in ACTION - 26 January'23

```
Hi all,
                                                                             Subject info access:
                                                                                                      rsync://rpki.august.tw/repo/AS945/0/AS945.asa
Since a few days OpenBGPD is able to do ASPA verification and filtering
                                                                                                      Sun 17 Dec 2023 14:17:12 +0000
                                                                              ASPA valid until:
based on the outcome. Right now my system detected one ASPA invalid path
                                                                              Customer AS:
                                                                                                      945
that is an actuall route leak. So it seems ASPA is working :)
                                                                              Provider Set:
                                                                                 1: AS: 1299
                                                                                 2: AS: 6939
    --- begin terminal transcript ---
                                                                                 3: AS: 32097
    $ bgpctl show rib in avs invalid as 945
                                                                                 4: AS: 50058
    flags: * = Valid, > = Selected, I = via IBGP, A = Announced,
                                                                             01/26/23 01:54:24 A 2606:b0c0:b00b::/48 13830 3356 6939 61138 945
           S = Stale, E = Error
                                                                             01/26/23 01:54:24 A 2606:b0c0:b00b::/48 13830 50058 50058 50058 50058 945
    origin validation state: N = not-found, V = valid, ! = invalid
                                                                             01/26/23 01:54:24 A 2606:b0c0:b00b::/48 14907 6939 61138 945
    aspa validation state: ? = unknown, V = valid, ! = invalid
                                                                             01/26/23 01:54:24 A 2606:b0c0:b00b::/48 14907 50058 50058 50058 50058 945
    origin: i = IGP, e = EGP, ? = Incomplete
                                                                             01/26/23 01:54:24 A 2606:b0c0:b00b::/48 206499 6939 61138 945
    flags vs destination
                                                             med aspath origin
                                    gateway
                                                     lpref
          V-! 2606:b0c0:b00b::/48 2001:4bf8::253
                                                                0 8271 6939 61138 945 i
                                                       100
    --- end terminal transcript ---
```

https://www.manrs.org/2023/02/unpacking-the-first-route-leak-prevented-by-aspa/





• ASPA ~ Timeline [BGP, RP, RTR, Signer]

```
    OpenBSD rpki-client and OpenBGPD
    Routinator, Krill and RTRTR, StayRTR, rpki-prover, and RIPE NCC have either released ASPA-capable software or are in advanced stages to do so.
    APNIC signer demo - <a href="https://github.com/APNIC-net/rpki-aspa-demo">https://github.com/APNIC-net/rpki-aspa-demo</a>
    6-10 months for IETF to ratify ASPA
    SIDROPS in later stages of specifying the ASPA standard
    Tom Harrison (APNIC RPKI Lead): will start hosted in 2024
    RIRs make Signers available
    COTS BGP Speakers implementations
    https://www.manrs.org/2023/05/estimating-the-timeline-for-aspa-deployment/
```





# Need Help?

- Want to learn more about:
  - crafting route filters,
  - securing Internet routing best practices/tools
    - RPKI
    - ROV
    - MANRS
- Refer to NSRC's free training videos at:
  - https://learn.nsrc.org/bgp







# Troubleshooting Tools

- How/where do engineers, researchers, and analysts find the data about the incidents discussed so far?
  - Many network operators (ISPs) run their own looking glass.
  - Many of us rely on globally distributed collectors like:
    - RouteViews (the original looking glass since 1995), and
    - RIPE's RIS (routing information service)





#### RouteViews

- A collaborative router looking glass to share BGP views among network operators and researchers.
  - RouteViews was founded at the University of Oregon's Advanced Network Technology Center (ANTC) in 1995.
     Data archives (*every 2 hours*) began in 1997 and amount to 50TBs (compressed) today.



 The group is currently led by the Network Startup Resource Center (NSRC) group engineering team at the University of Oregon.



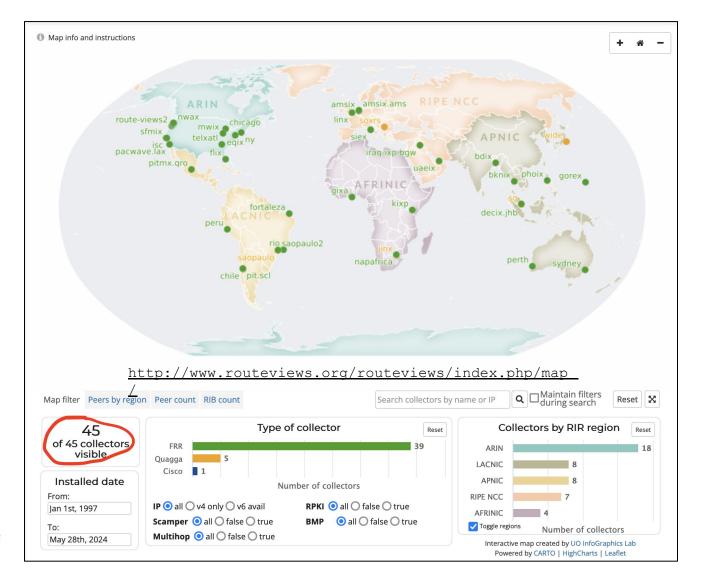
# Why RouteViews?

- Originally conceived in 1995 as a tool for Internet Operators to look at the BGP table from different locations/backbones around the world to <u>troubleshoot</u> and assess:
  - reachability, hijacks, peer visibility, mass withdrawals, and RPKI status
- The 27-year data-set of BGP information archived by RouteViews since 1997 has become an invaluable research resource
  - RouteViews data has been used in over 1000 research papers.
  - http://www.routeviews.org/routeviews/index.php/papers/





# RouteViews Collector Map







## Peering with RouteViews

- Send full table (if you can)
- Remove default routes
- Remove NULL routes
- Remove RFC1981 addresses
- RouteViews don't accept/want ADD-PATH (TX/RX)
- RouteViews don't send routes to you (ONLY collects)
- When peering with multi-hop collectors, set ebgp-multihop

https://www.routeviews.org/routeviews/index.php/peering-request-form/





# khàawp Khun Kráp ขอบคุณ ครับ





