Securing Internet Routing

Tashi Phuntsho (<u>tashi@apnic.net</u>) Senior Network Janitor/Technical Trainer





Why should we bother?

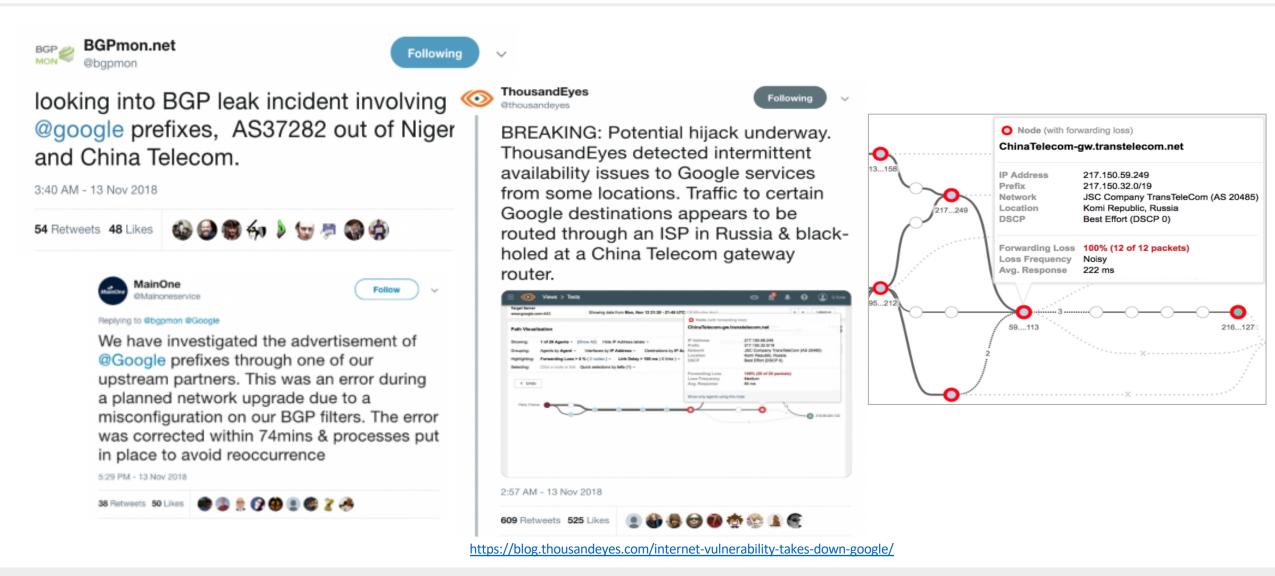


- As a Manager
 - I don't want to be front page news of a IT paper, or an actual newspaper for routing errors









Headlines

(::**ʃ**::ʃ::ʃ::ʃ::ʃ:/

ars **TECHNICA**

BIZ & IT TECH SCIENCE POLICY CARS GAMING & CULTURE

BORDER GATEWAY PROTOCOL ATTACK -

Suspicious event hijacks Amazon traffic for 2 hours, steals cryptocurrency

Almost 1,300 addresses for Amazon Route 53 rerouted for two hours.

Kevin Beaumont 👔 🤣 @GossiTheDog · Apr 24, 2018

so users clicked through certificate errors.

Maybe related to this: twitter.com/InternetIntel/...

Doug Madory

InternetIntelligence @InternetIntel

routes from 11:05 to 13:03 UTC today:

@DougMadory

205.251.192.0/24

205.251.193.0/24 205.251.195.0/24

205.251.197.0/24

205.251.199.0/24

⑦ 2 9:23 PM - Apr 24, 2018

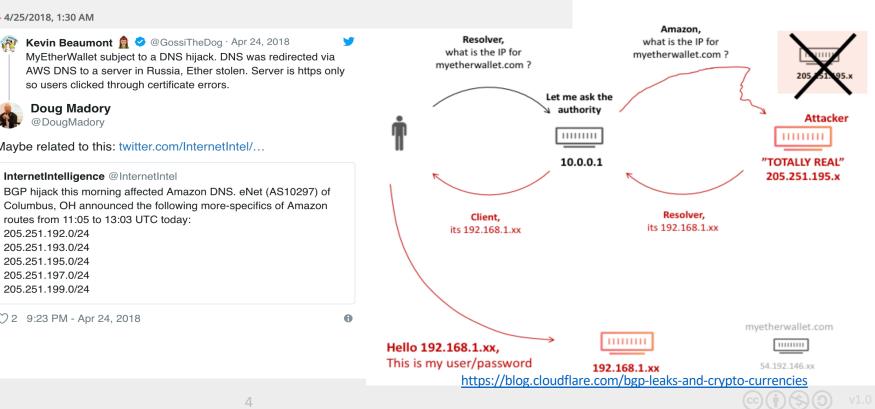
DAN GOODIN - 4/25/2018, 1:30 AM

Follow



BGP hijack this morning affected Amazon DNS. eNet (AS10297) of Columbus, OH announced the following more-specifics of Amazon routes from 11:05 to 13:03 UTC today: 205.251.192.0/24 205.251.193.0/24 205.251.195.0/24 205.251.197.0/24 205.251.199.0/24

7:52 AM - 24 Apr 2018



Headlines

(::*f*::*f*::*f*::*f*::*f*)

Large BGP Leak by Google Disrupts Internet in Japan

通信が安定致しました。

Research // Aug 28, 2017 // Doug Madory N光及びOCNでNTT東西のフレッツ

2 183.177.32.145	Foundation Desifie	Televe	lanan	0.240
	Equinix Asia Pacific	Tokyo	Japan	0.249
	IIJ IPv4 BLOCK (AS2497)	Tokyo	Japan	0.618
4 58.138.102.109		Tokyo	Japan	0.877
5 58.138.88.86	sjc002bb12.IIJ.Net		United States	
6 152.179.48.117 7 *	TenGigE0-3-0-8.GW6.SJC7.ALTER.NET	San Jose	United States	97.869
8 152.179.105.110	<pre>google-gw.customer.alter.net</pre>	Chicago	United States	337.19
9 108.170.243.197	Google Inc.	Chicago	United States	246.325
10 *	Coordia Tao		United Chattan	256 100
11 209.85.241.43	Google Inc.	11	United States	256.188
12 72.14.238.38	Google Inc.	Vancouver		247.849
13 209.85.245.110 14 *	Google Inc.	Vancouver	Canada	249.291
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 18 *	OCN (AS4713) CIDR BLOCK 81	Tokyo	Japan	246.426
19 153.149.218.10	OCN (AS4713) CIDR BLOCK 93	Ōsaka-shi	Japan	256.027
20 125.170.96.38 21 *	OCN (AS4713) CIDR BLOCK 77		Japan	255.683
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 * 5 211.11.83.160	OCN (AS4713) CIDR BLOCK 23	Inuyama	Japan	256.212

trac 1 1	ce from London, I *	England to Nürnberg, Germany at 03	30 Aug 25, 2017	7	
2 1	195.66.248.190	fe0-2.tr2.linx.net	London	United Kingdom	0.327
3 1	195.66.249.10	ge0-2-502.tr5.linx.net	London	United Kingdom	0.441
4 1	195.66.249.13	ge0-2-501.tr4.linx.net	London	United Kingdom	0.477
5 1	195.66.248.10	uunet-uk-transit.thn.linx.net	London	United Kingdom	0.507
6 1	158.43.193.245	POS0-0.CR2.LND6.ALTER.NET	London	United Kingdom	0.497
7 1	140.222.239.41	0.xe-0-0-0.IL1.NYC50.ALTER.NET	New York	United States	108.146
8 1	146.188.4.197	xe-0-0-1.IL1.NYC41.ALTER.NET	New York	United States	75.719
9 1	140.222.234.221	0.et-10-1-0.GW7.CHI13.ALTER.NET	Chicago	United States	94.793
10 1	152.179.105.110	<pre>google-gw.customer.alter.net</pre>	Chicago	United States	224.352
11 *	*				
12 2	216.239.40.189	Google Inc.	Northlake	United States	202.193
13 2	216.239.58.255	Google Inc.			203.995
14 2	216.239.58.12	Google Inc.			207.026
15 2	209.85.253.184	Google Inc.	Luxembourg	Luxembourg	212.944
16 2	209.85.252.215	Google Inc.			213.112
17 1	108.170.252.71	Google Inc.			213.265
18 7	72.14.222.53	Google Inc.		Germany	212.061
19 1	188.111.165.169	Vodafone GmbH		Germany	227.077
20 1	178 7 128 117	Vodafone D2 CmbH	Nürnhorg	Germany	JJV JJE

After leak (EU->EU)

After leak (JP->JP)

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





YouTube blames Pakistan network for 2-hour outage

Company appears to confirm reports that Pakistan Telecom was responsible for routing traffic according to erroneous Internet Protocols.

BY GREG SANDOVAL | FEBRUARY 24, 2008 10:15 PM PST



Why should we bother?



- As a Engineer
 - I don't want to be told at 3AM my routing is broken
 - Or while on a holiday





- Because NO ONE is in charge?
 - No single authority model for the Internet
 - No reference point for what's right in routing





- 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?





- Routing is VARIABLE
 - □ The view of the network depends on where you are
 - Different routing outcomes at different locations
 - ${\scriptstyle \Box}~\sim$ no reference view to compare the local view ${\scriptstyle \textcircled{\sc o}}$



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





- And as always, there is no E-bit
 A bad routing update does not identify itself as BAD
- So tools/techniques try to identify GOOD updates



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

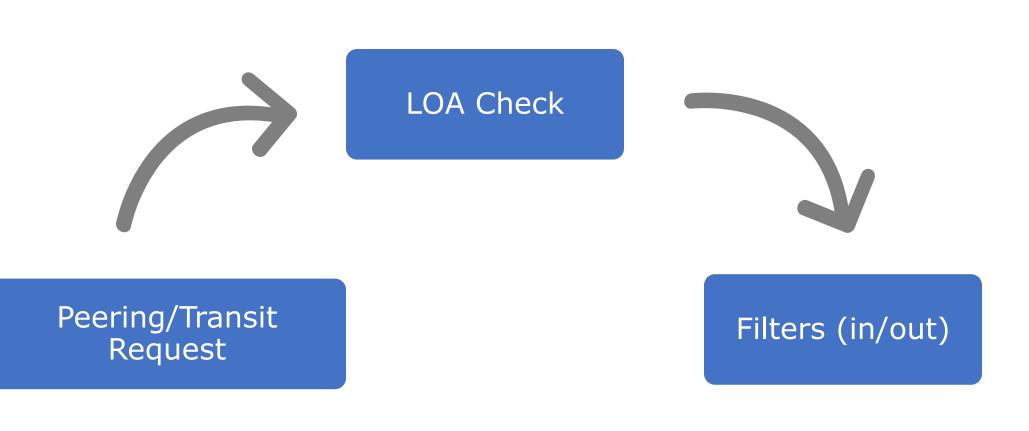




APNIC

Current practice

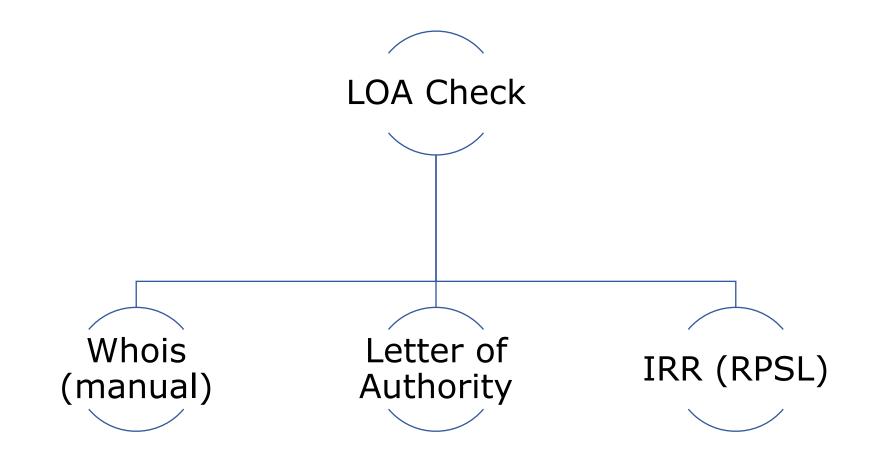






Tools & Techniques







Tools & Techniques

(::**ʃ**::ʃ::ʃ::ʃ::ʃ

 Look up whois
 verify holder of a resource

<pre>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 r</pre>				
% 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	country phone:			
netname: APNICTRAINING-AP	fax-no:			
descr: Prefix for APNICTRAINING LAB DC	e-mail:			
country: AU				
admin-c: AT480-AP	admin-c			
tech-c: AT480-AP	tech-c:			
status: ALLOCATED NON-PORTABLE	nic-hdl			
mnt-by: MAINT-AU-APNICTRAINING	mnt-by:			
mnt-irt: IRT-APNICTRAINING-AU	last-mc			
last-modified: 2016-06-17T00:17:28Z	source:			
source: APNIC	5001 66.			
	% Infor			
irt: IRT-APNICTRAINING-AU	∕₀ 111101			
address: 6 Cordelia Street				
address: South Brisbane	route:			
address: QLD 4101	descr:			
e-mail: training@apnic.net	origin:			
abuse-mailbox: training@apnic.net	mnt-by:			
admin-c: AT480-AP	country			
tech-c: AT480-AP	last-mo			
auth: # Filtered				
mnt-by: MAINT-AU-APNICTRAINING	source:			
last-modified: 2013-10-31T11:01:10Z				
source: APNIC				

ole:	APNIC Training
ddress:	6 Cordelia Street
ddress:	South Brisbane
ddress:	QLD 4101
ountry:	AU
none:	+61 7 3858 3100
ax-no:	+61 7 3858 3199
-mail:	training@apnic.net
dmin-c:	JW3997-AP
ech-c:	JW3997-AP
ic-hdl:	AT480-AP
nt-by:	MAINT-AU-APNICTRAINING
ast-modified:	2017-08-22T04:59:14Z
ource:	APNIC

% Information related to '202.125.96.0/24AS131107'

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

APNIC

Tools & Techniques

Ask for a Letter of Authority

Absolve from any liabilities



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





APNIC

Tools & Techniques

- Look up/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-modifie	ed: 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:	<pre>* http://www.ripe.net/whois</pre>
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 ASEASS appounde AS-DPUKNET-TRANSTT

AS-DRUKNET-TRANSIT

announce

to AS6453

export:

Tools & Techniques

(::**ʃ::**ʃ::**ʃ**::**ʃ**::

• 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@tash	i 🗠 hana3 -t	F 38195 _1 SUP		IN AS-SUPERLOOP
		ist SUPERLOOP-		
				^38195(_38195)*\$
				^38195(_[0-9]+)*_(681 4647 4749 4785)\$
				^38195(_[0-9]+)*_(4846 4858 7477 7578)\$
				^38195(_[0-9]+)*_(7585 7604 7628 7631)\$
				^38195(_[0-9]+)*_(7699192901929719336)\$
				^38195(_[0-9]+)*_(9499 9544 9549 10143)\$
				^38195(_[0-9]+)*_(10145 11031 12041 15133)\$
				^38195(_[0-9]+)*_(15967 17462 17498 17766)\$
				^38195(_[0-9]+)*_(17829 17907 17991 18000)\$
				^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)\$
				^38195(_[0-9]+)*_(55455 55506 55575 55707)\$
				^38195(_[0-9]+)*_(55752 55766 55803 55845)\$
				^38195(_[0-9]+)*_(55884 55931 55954 56037)\$
				^38195(_[0-9]+)*_(56098 56135 56178 56225)\$
				^38195(_[0-9]+)*_(56271 56287 58422 58443)\$
				^38195(_[0-9]+)*_(58511 58606 58634 58676)\$
				^38195(_[0-9]+)*_(58712 58739 58750 58868)\$
				^38195(_[0-9]+)*_(58914 59256 59330 59339)\$
				^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)?

Back to basics – identifying GOOD



- Using digital signatures to convey the "*authority to use*"?
 - A private key to sign the authority, and
 - the public key to validate that authority



How about trust in this framework?



• Follows the resource allocation/delegation hierarchy



RPKI Chain of Trust



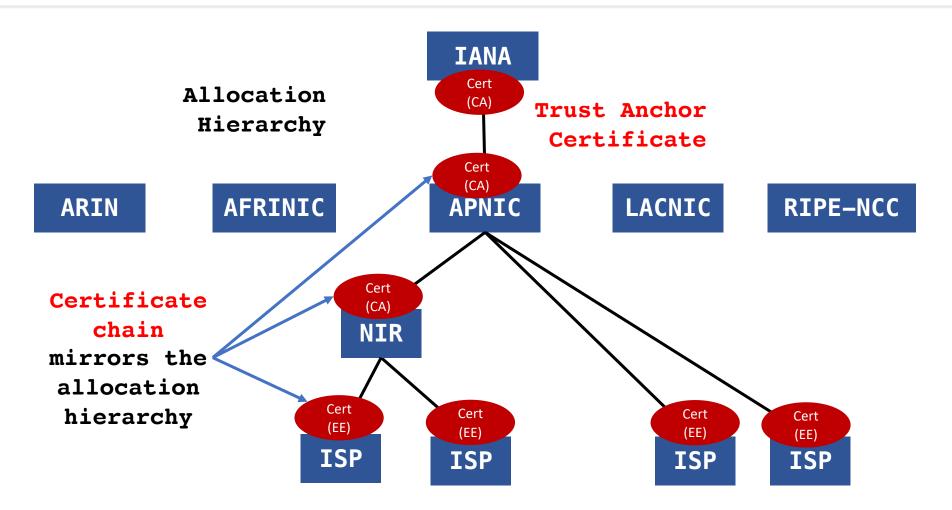


Image 4







- 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
 - proves the holder of the private key (B) is the legitimate holder of the resource!



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)

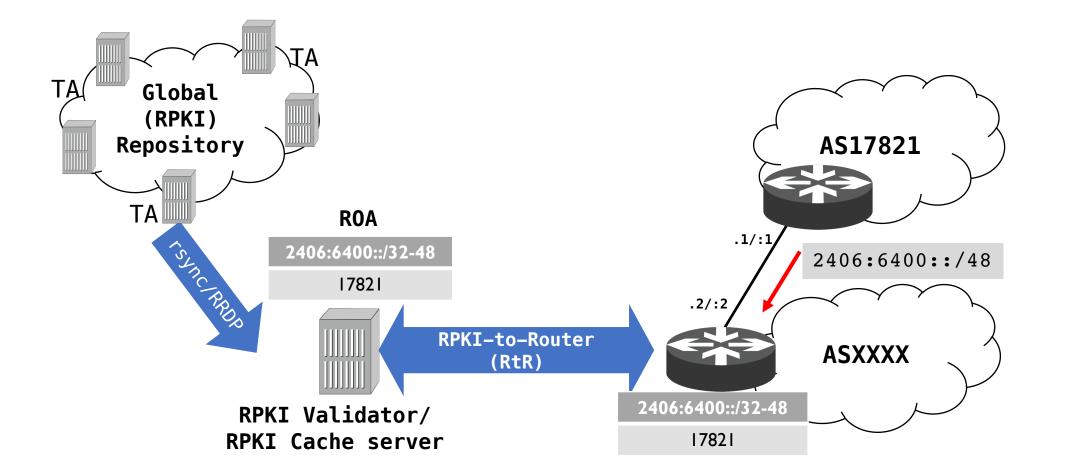
Prefix	203.176.32.0/19
Max-length	/24
Origin ASN	AS17821





Route Origin Validation (ROV)







Are ROAs enough?

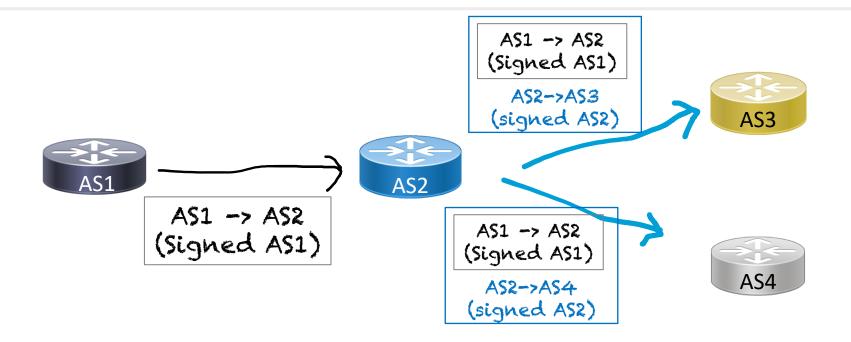


- What if I forge the origin AS in the AS path?
 Would be accepted as "good" pass origin validation!
- Which means, we need to secure the AS path as well
 Need AS path validation (per-prefix)



AS path validation - BGPsec





□ 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

AS path validation issues...



- More resources
 - CPU high crypto overhead to validate signatures, and
 - Memory
 - Updates in BGPsec would be per prefix
 - New attributes carrying signatures and certs/key-id for every AS in the AS path
- How do we distribute the certificates required?
- Can we have partial adoption?
- Given so much overhead, can it do more Route leaks?

What can we do?

- Basic BGP OpSec hygiene RFC7454/RFC8212
 - RFC 8212 BGP default reject or something similar
 - Filters with your customers and peers
 - Prefix filters, Prefix limit
 - AS-PATH filters, AS-PATH limit
 - Use IRR objects (source option) or ROA-to-IRR
 - Filters with your upstream(s)
 - Create ROAs for your resources
 - Filter based on ROAs -> ROV
- Join industry initiatives like MANRS
 - https://www.manrs.org/







APNIC

ROV – Industry trends

AT&T/as7018 now drops invalid prefixes from peers

Jay Borkenhagen jayb at braeburn.org Mon Feb 11 14:53:45 UTC 2019

[apops] RPKI ROV & Dropping of Invalids - Africa

- To: apops@apops.net
- Subject: [apops] RPKI ROV & Dropping of Invalids Africa
- Previous message (by thread): BGP topological vs centra From: Mark Tinka <mark.tinka@seacom.mu>
- Next message (by thread): AT&T/as7018 now drops investor protection from protection of the protectio
- Messages sorted by: [date] [thread] [subject] [auf Hello all.

FYI:

In November 2018 during the ZAPF (South Africa Peering Forum) meeting in Cape Town, 3 major ISP's 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.

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

We continue to accept invalid route announcements at least for now. We are communicating with our c

invalid announcements we are propagating, informin We are still standing by for the 3rd ISP to complete their implementation, and we are certain they will communicate with the community accordingly. routes will be accepted by fewer and fewer network

Thanks to those of you who are publishing ROAs in also like to encourage other networks to join us i prefixes on our networks 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 to improve the guality of routing information in t 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.

Thanks!

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:

Ja - 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).



MMIX is

dropping

Invalids



Acknowledgement

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



Any questions?







