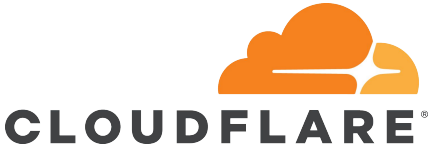


RPKI: our approach for deploying at scale

Louis Poinsignon - RIPE77

Introduction to Cloudflare

Some numbers...

- 155+ PoPs and growing
- 72+ countries
- 186+ Internet exchanges

- >600B Web requests a day ~10% of all web requests
- Regular DDoS attacks larger than 500Gbps, 300M PPS
- >100bn DNS requests a day

CLOUDFLARE

Who am I?

- Louis Poinsignon
- Network, data and software @ Cloudflare London and SF
- Built a network data pipeline (flows and BGP) for Cloudflare scale, open-source:
  https://github.com/cloudflare/goflow
  https://github.com/cloudflare/fgbgp

CLOUDFLARE

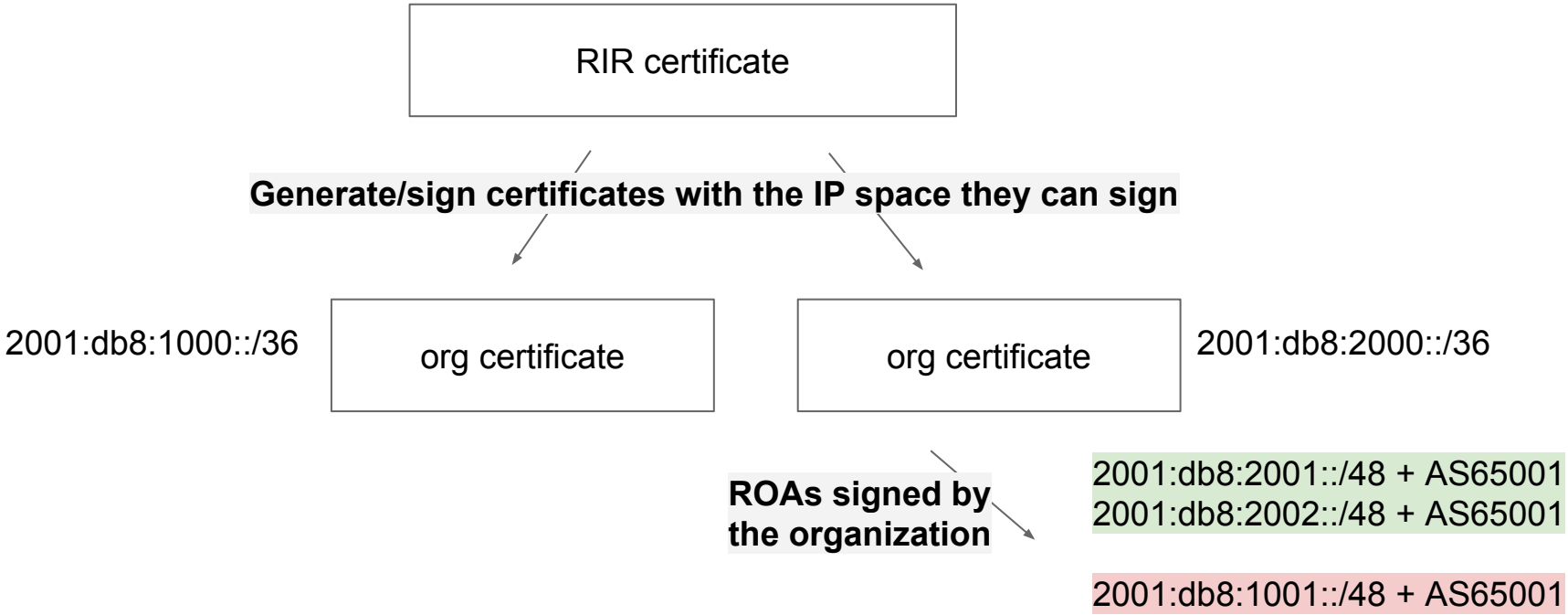
RPKI

What is RPKI

RFC6480: defines a way of cryptographically signing:
route + length + origin ASN

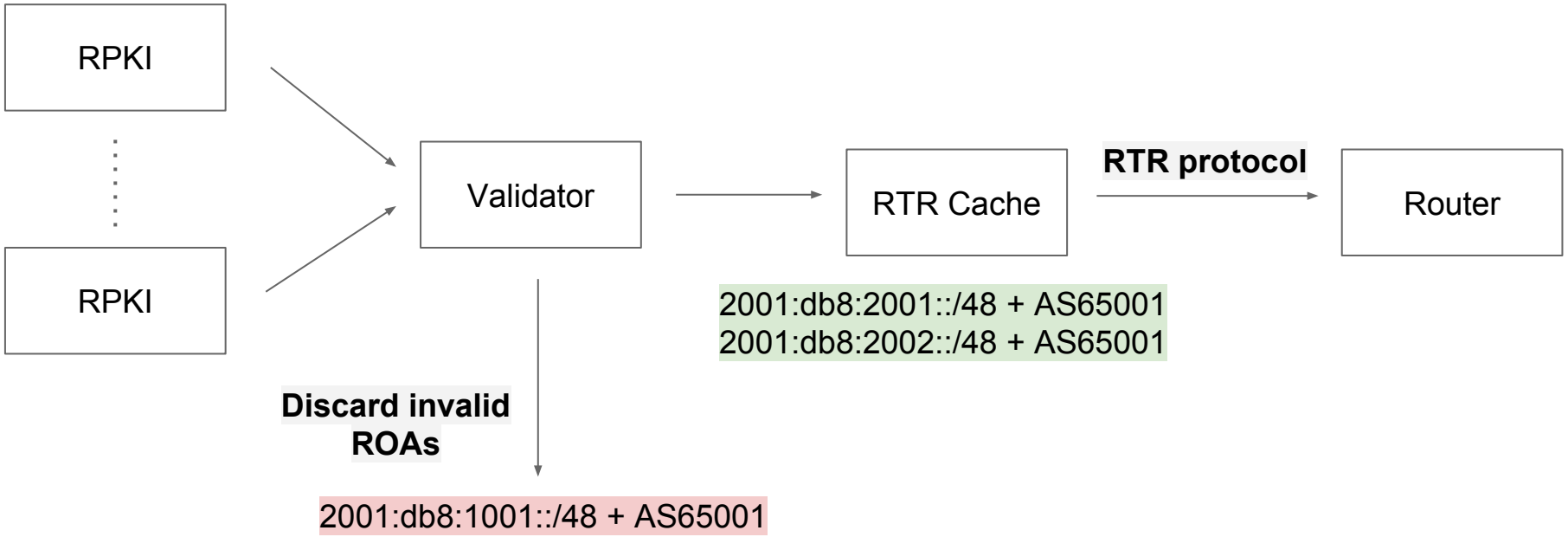
RFC6810: defines a communication method between router and validation system

How RPKI works



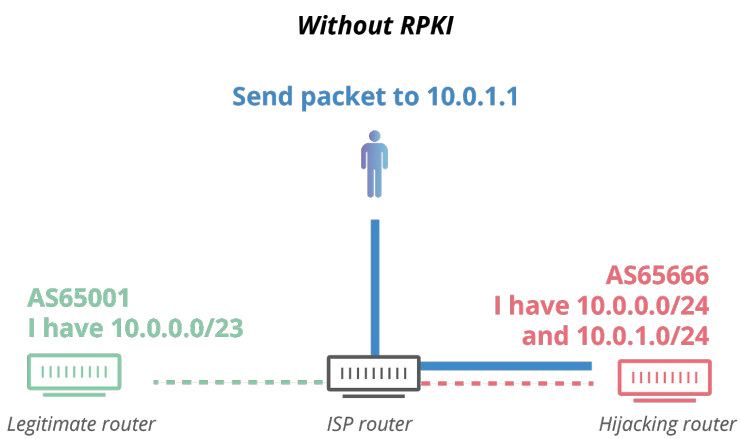
CLOUDFLARE

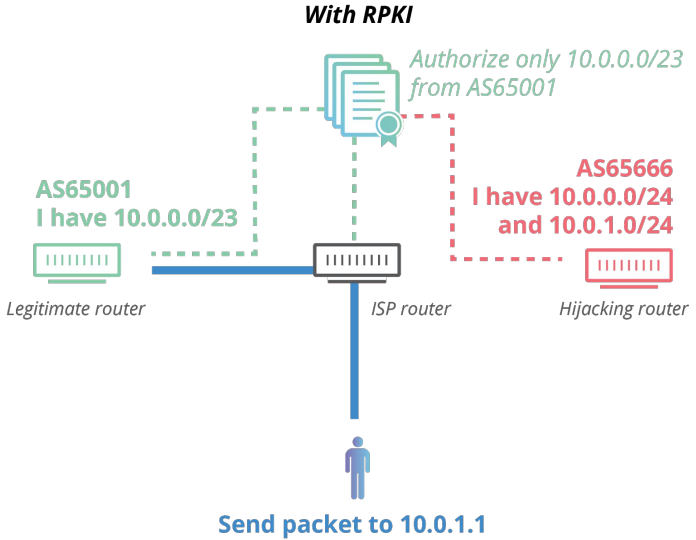
How validation works



CLOUDFLARE

Summary





CLOUDFLARE

Use-cases

- Filter out bad announcements
- For “Bring your own IP” services → make sure your clients are the true owner of a range

BGP leaks and hijacks

Why signing?

BGP leaks and cryptocurrencies - The Cloudflare Blog
https://blog.cloudflare.com/bgp-leaks-and-crypto-currencies/
Apr 24, 2018 - The broad definition of a **BGP leak** would be IP space that is announced by somebody not ... Those IPs are for Route53 **Amazon** DNS servers.

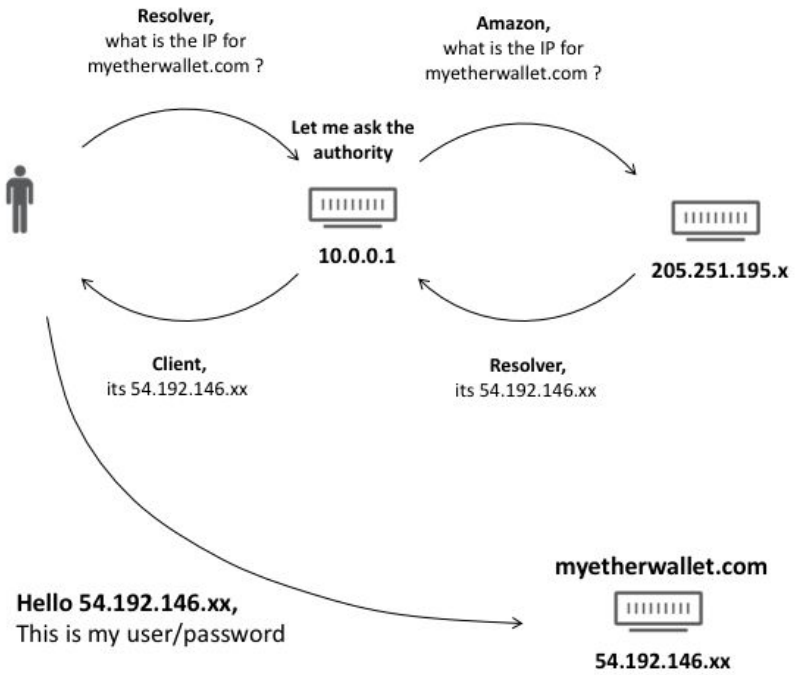
Amazon Route 53 DNS and BGP Hijack - ThousandEyes Blog
https://blog.thousandeyes.com/amazon-route-53-dns-and-bgp-hijack/
Apr 24, 2018 - Anatomy of a **BGP** Hijack on **Amazon's** Route 53 DNS Service .... blog posts reviews some best practices for combating **BGP leaks** and hijacks.

BGP routing security flaw caused Amazon Route 53 incident
https://searchsecurity.techtarget.com/.../BGP-routing-security-flaw-caused-Amazon-Ro...
Apr 25, 2018 - A long-standing flaw in **BGP** routing security that allows attackers to ... to eliminate **BGP** route hijacking, route **leaks** and forwarding of traffic with ...

Suspicious Event Hijacks Amazon Traffic For 2 hours, Steals ...
https://it.slashdot.org/.../suspicious-event-hijacks-amazon-traffic-for-2-hours-steals-cry...
Apr 24, 2018 - **Amazon** lost control of some of its widely used cloud services for two ... 'Kernel Memory **Leaking**' Intel Processor Design Flaw Forces Linux, Windows Redesign ..... We have yet to see a **BGP** session be hijacked, or an external ...

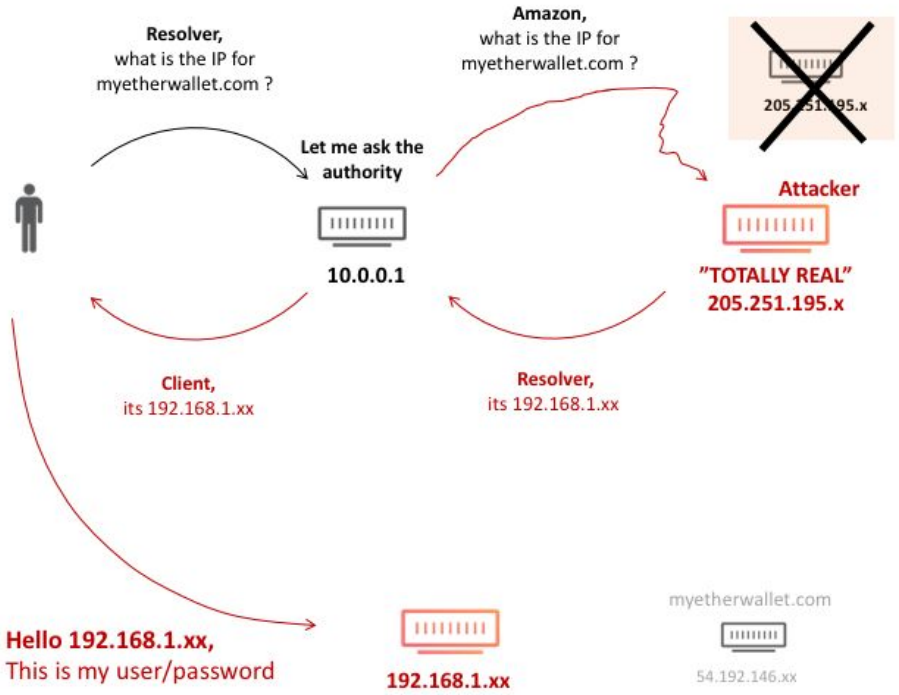
CLOUDFLARE

What happened?



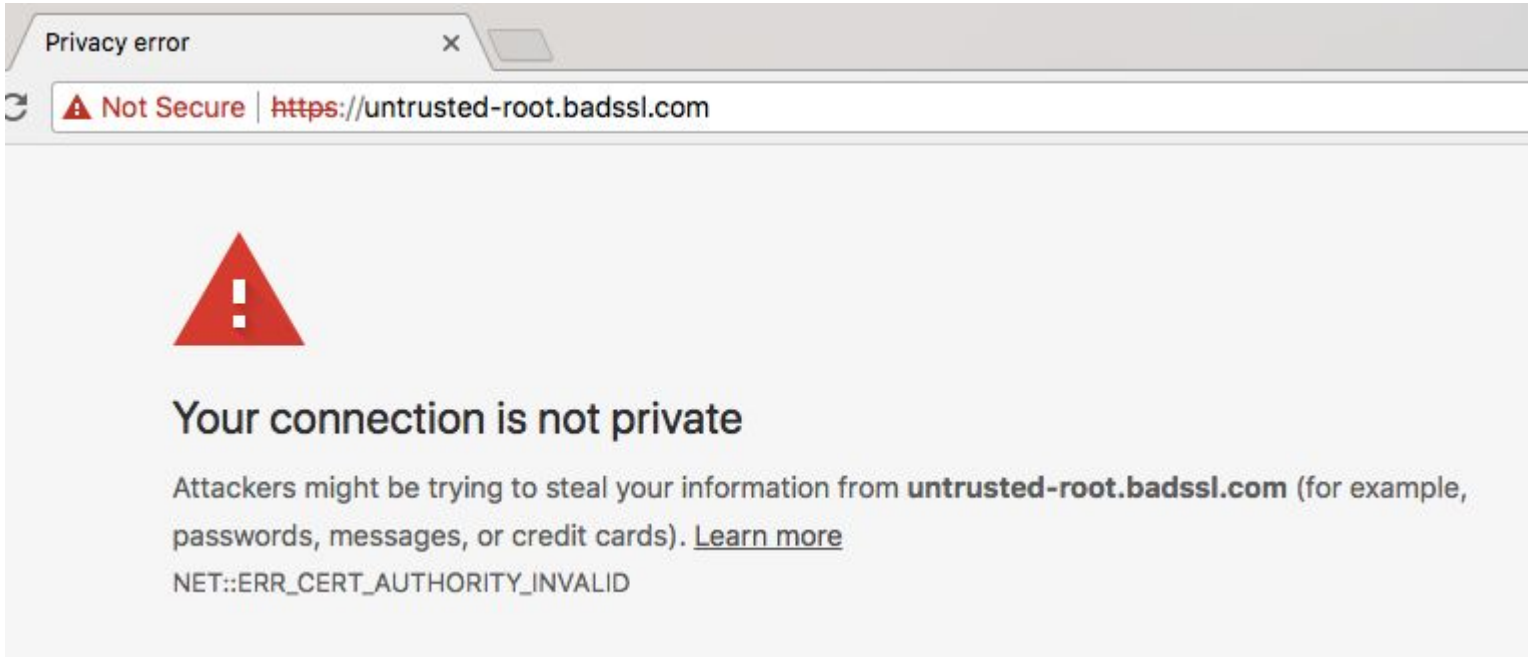
CLOUDFLARE

What happened?



CLOUDFLARE

What happened?



CLOUDFLARE

BGP leaks/hijacks

“CIA Triad”: *Confidentiality, integrity, availability*

- Rendering a ressource unreachable (availability)

or

- Impersonating
  - Protocols at risk: DNS/UDP due to no confidentiality nor integrity checks
  - HTTPS and DNSSEC offers a layer of security: reduce availability in exchange of integrity

BGP leaks/hijacks

Someone controlling **65002** wants to hijack **2001:db8:3000::/32** originally announced by **65001**
Possible types:

#	Announcement	AS Path	Effect
1	2001:db8:3000::/32	AS65002	May become shortest AS Path. BGP origin validation/RPKI could filter it out. Sensitive on IXes.
2	2001:db8:3000::/32	AS65002 AS65001	BGP origin validation out of scope. But AS Path longer so less risks. Sensitive on IXes.
3	2001:db8:3000::/48	either	Most specific prefix: will be preferred as long as accepted. BGP origin validation/RPKI could filter it out.

BGP leaks/hijacks

From the previous table: very localized attacks.

While waiting on RPKI:

- IRR filtering: but no guarantees the owner of the prefix actually wrote the information.
- Announcing max-length /24 IPv4 or /48 IPv6 for critical ressources like authority DNS

At scale

At scale

What does “at scale” mean?

- Addresses in all 5 regions (LACNIC, Afrinic, APNIC, ARIN, RIPE)
- Automate prefixes signing and invalidation + long term maintenance
- Strict validation at scale
- Monitoring and failure models

Choice of mode

Hosted or delegated?

- **Hosted**: the certificate and ROA signing is maintained by the RIR
- **Delegated**: a certificate indicates the location of the PKI of the organization. ROAs are generated and signed by the organization.

RIR	Status
Afrinic	Both
APNIC	Both
ARIN	Both
LACNIC	Hosted
RIPE	Hosted and on-demand delegated

CLOUDFLARE

Hosted

We chose **hosted** because:

- We do not allocate IP addresses
  - Very few changes, made by the network team

- Only a handful of software for maintaining RPKI CA
  - rsync to maintain

- Not all RIR offer delegated
- If the RIR certificate is compromised: similar to any CA compromised

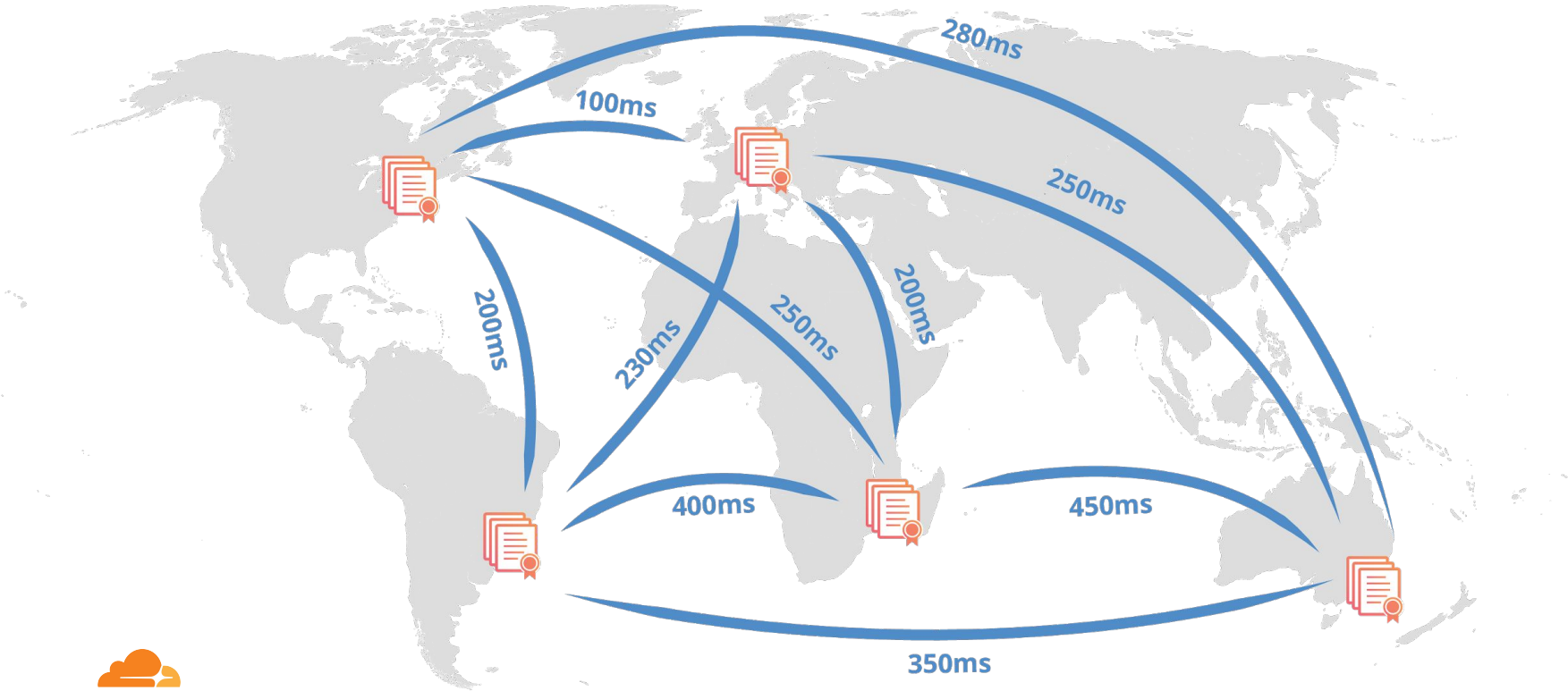
CLOUDFLARE

APIs

- With automation, we want **APIs (GET/PUT/UPDATE/DELETE)**.
- Cloudflare announces many prefixes. We have our provisioning databases/IPAM.

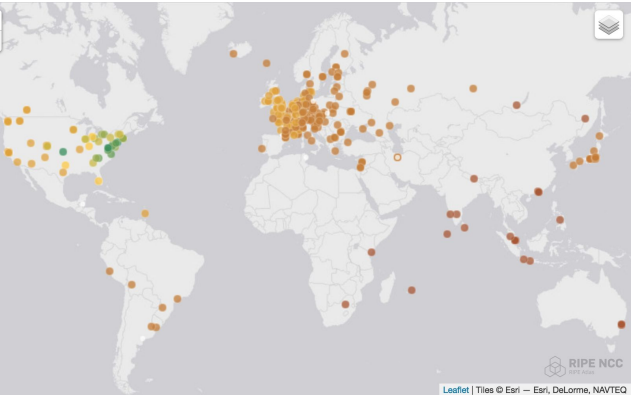
RIR	Status
Afrinic	Uses APNIC software
APNIC	Draft
ARIN	Insertion only (not listing, updating, deleting)
LACNIC	No (but easier to batch)
RIPE	No (but easier to batch)

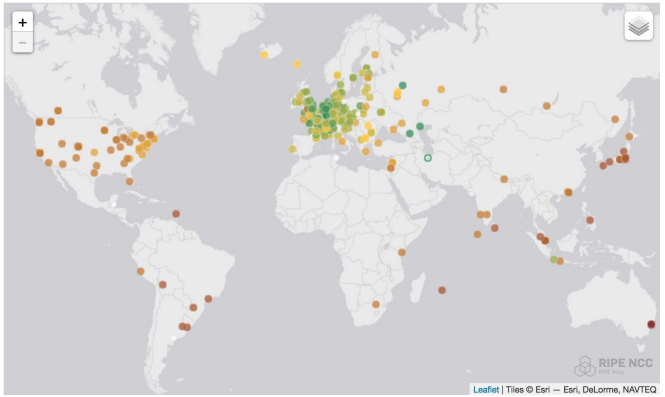
Availability

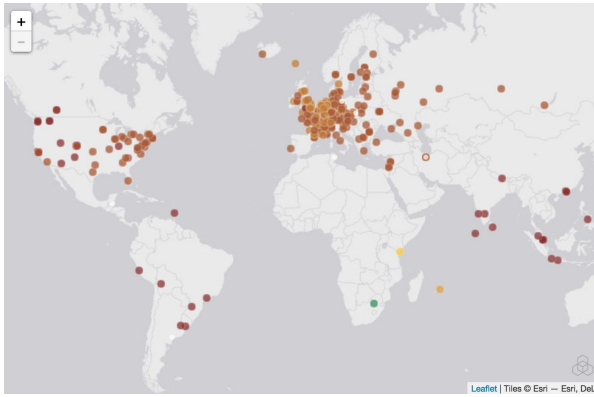


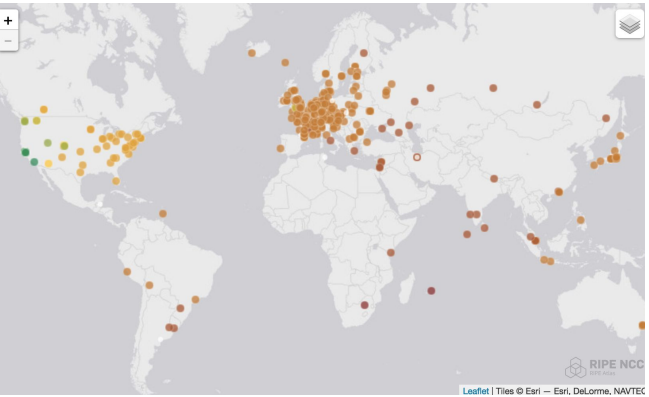
CLOUDFLARE

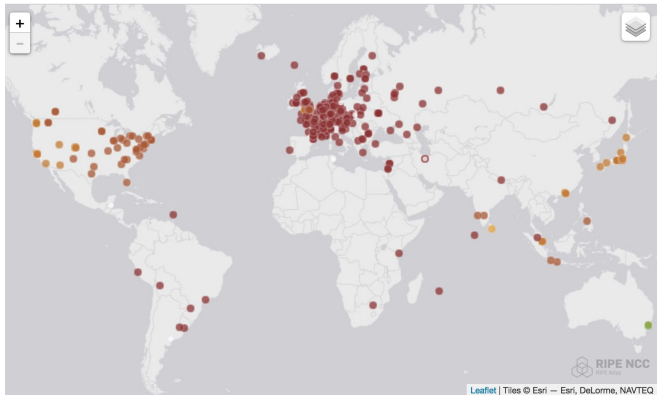
RIPE Atlas → RPKIs

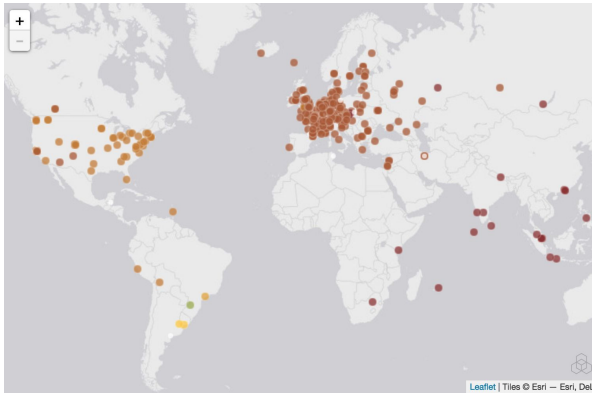












Availability

Low throughput for ⅘ RPKIs (>80ms)

East Asia = no local RPKI

Rsync protocol

- Caching?
- High usage?

CLOUDFLARE

Availability

From Sydney

- RIPE: 90MB, took 5 min (2.4Mbps)
- ARIN: 9MB, took 5 seconds (14Mbps)
- APNIC: 5MB, took 1 second (40Mbps)
- LACNIC: 19MB, took 10 seconds (15Mbps)
- Afrinic: 2MB, took 11 seconds (1.45Mbps)

Future?

- A bit more than 10% of the routes.
- If everything was signed, 1 GB to download at 2-4Mbps (30mn-1 hour)
  - Painful updates/refresh
  - Database could be filled with random records

ASN	Prefix	Max Length
AS0	2001:7fa:0:3::/64	128

CLOUDFLARE

Security and performance

We have 150+ PoPs.

How to do validation on every single one of them?

Security and performance

- RTR from central point to each router
  - Single point of failure
  - Latency/packet loss
  - **No encryption** (only TCP supported, no TLS or SSH)

Security and performance

- Validator software on every PoP
  - Wasted resources (10GB disk/RAM, 1-2 CPU)
  - Harder monitoring and maintenance
  - Latency to rsync from faraway places

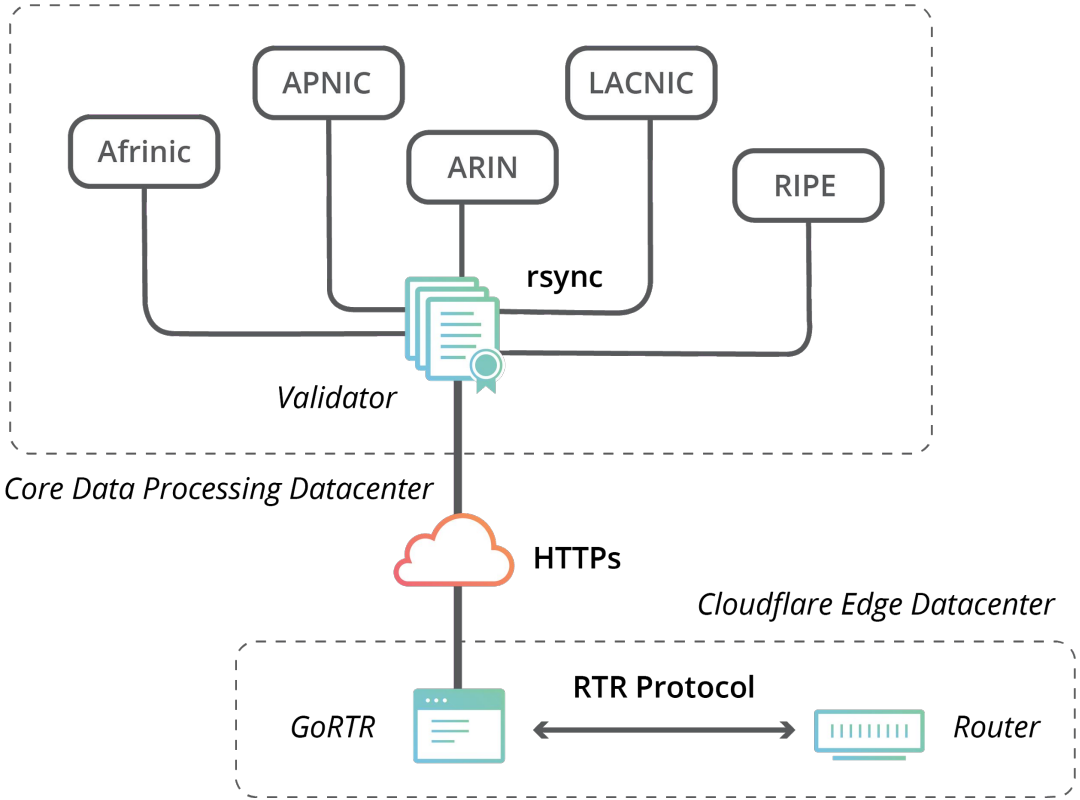
Security and performance

- Our solution
  - Have a local cache in each PoP using our CDN and HTTPs
  - Central validation having authority
  - Custom RTR software to communicate with routers

  - Integration with Salt and our pipelines

CLOUDFLARE

Security and performance



CLOUDFLARE

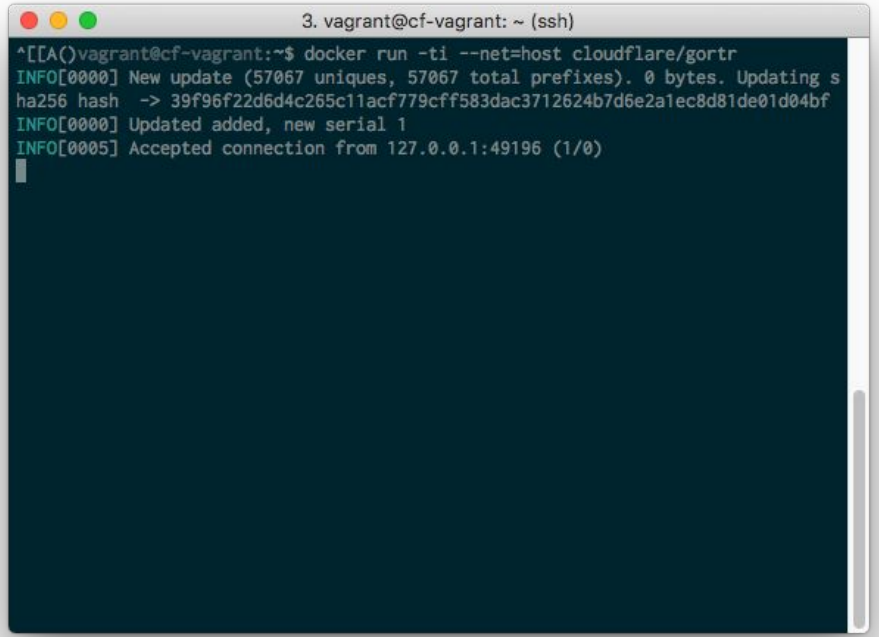
Also

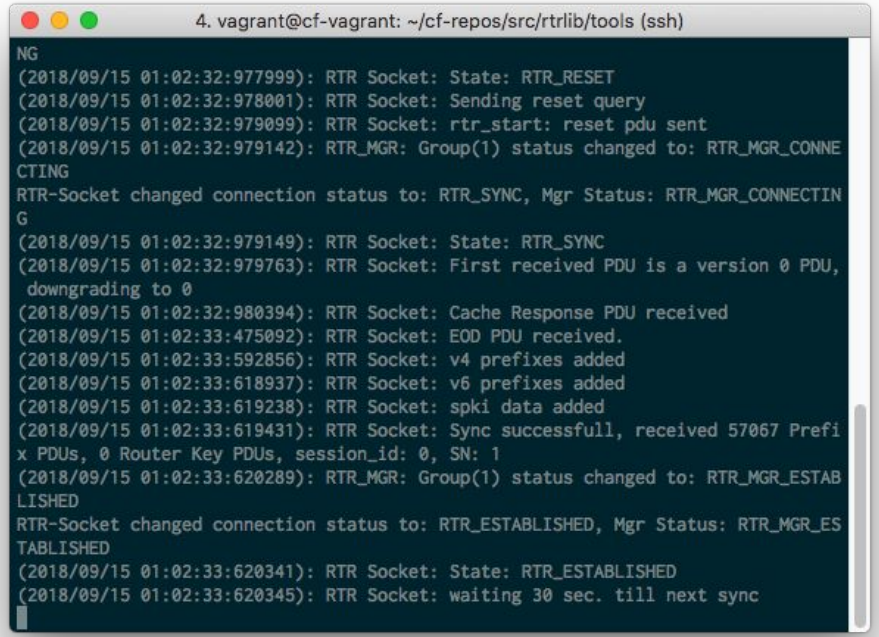
- Use our list of validated prefixes **signed** (RIPE Validator Format):
  - **https://rpki.cloudflare.com/rpki.json**

- Use our implementation of RTR Cache
  - **https://github.com/cloudflare/gortr**

CLOUDFLARE

GoRTR

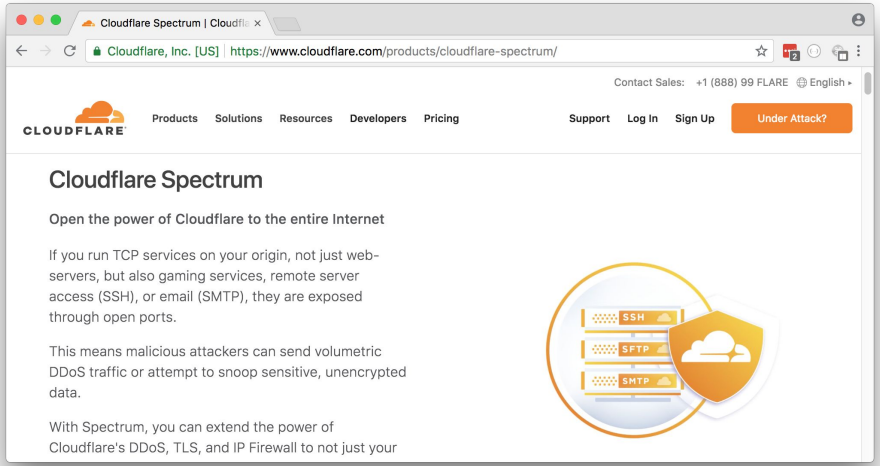




CLOUDFLARE

Also

- Soon™:
  - A RTR Server service on Cloudflare Spectrum
  - Nothing to install
  - *If you want to run tests*

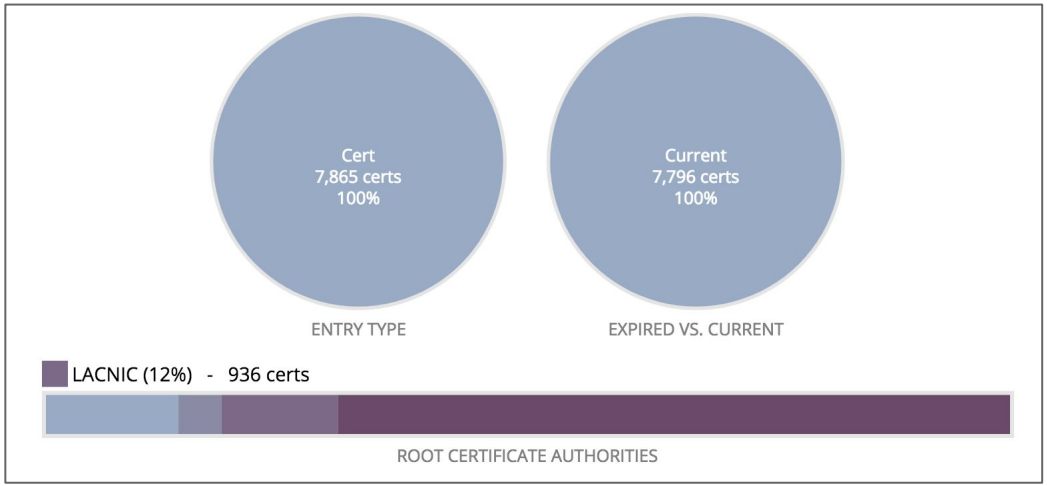


CLOUDFLARE

Monitoring

Monitoring of PKI

- Cloudflare's Certificate Transparency
  - https://ct.cloudflare.com/logs/cirrus



LOG DETAILS

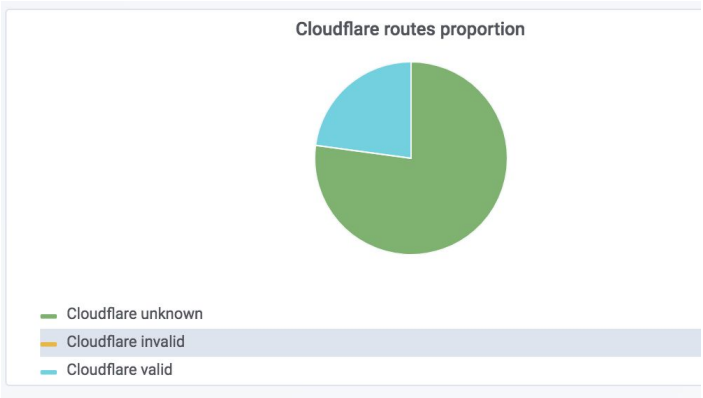
Cloudflare Cirrus

ct.cloudflare.com/logs/cirrus
Last Update: 2018-09-03 21:19 UTC
Avg. Throughput: 0 certs/hr
Contains: 7,886 certificates
Unsubmitted: 0 certificates (100% full)

CLOUDFLARE

Monitoring of validation

- Coming from our validator:
  - Number of ROAs
  - Distribution



- Coming from our edge
  - Number of invalids/valids
  - Number of filtered routes

- Online
  - https://rpki-monitor.antd.nist.gov/

CLOUDFLARE

Monitoring of filtering

- Project @ Cloudflare:
  - With Cloudflare’s presence in more than 180 IX
  - Announce a prefix /24 IPv4 and /48 IPv6 which should be invalid
  - Have the enclosing prefix announced somewhere.
  - Probe the equipments + prefixes announced

CLOUDFLARE

Questions?

Thank you!

louis@cloudflare.com
@lpoinsig