RPKI Ecosystem Measurement

NANOG / Seattle

2023.06.14

Romain Fontugne, Amreesh Phokeer, Cristel Pelsser,
Kevin Vermeulen, & Randy Bush

Agenda

- Experimental Set-Up
- Measuring RPKI Data Propagation
- Measuring Effect on BGP (grotty CDFs)
- Some Problems and Anomalies
- What Needs to be Done
  - Faster Publication
  - Increase Fetching Frequency
- A Warning and a Plea for Simplicity

Romain had to present at PAM, an academic audience

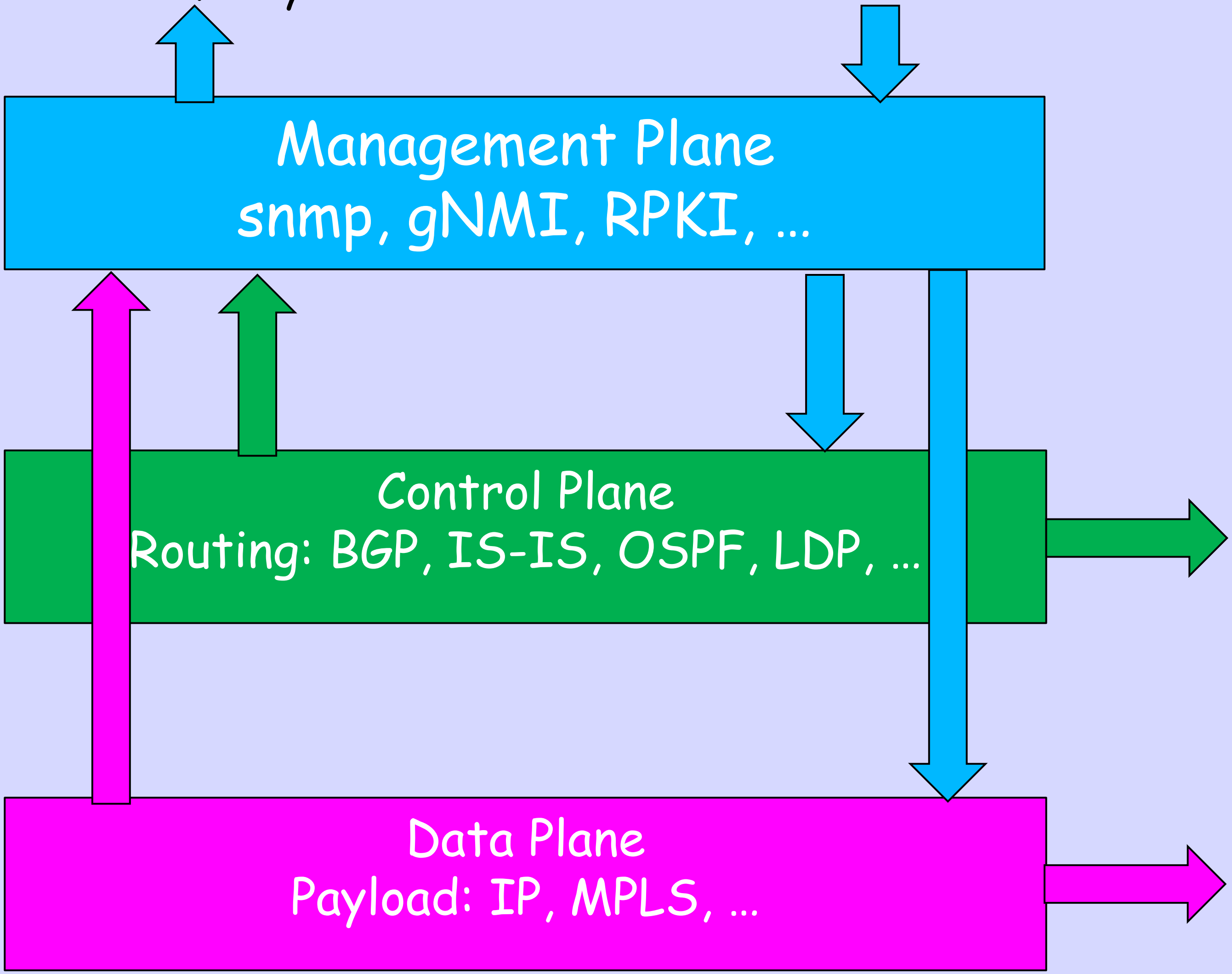
So first he had to describe the RPKI

RPKI is Complex

Into the Future with the Internet Vendor Task Force A Very Curmudgeonly View or Testing Spaghetti — A Wall's Point of View
ACM SIGCOMM Computer Communication Review Volume 35, Number 5, October 2005

Measurement
Monitoring
Telemetry

Configuration
RPKI/ROV
etc



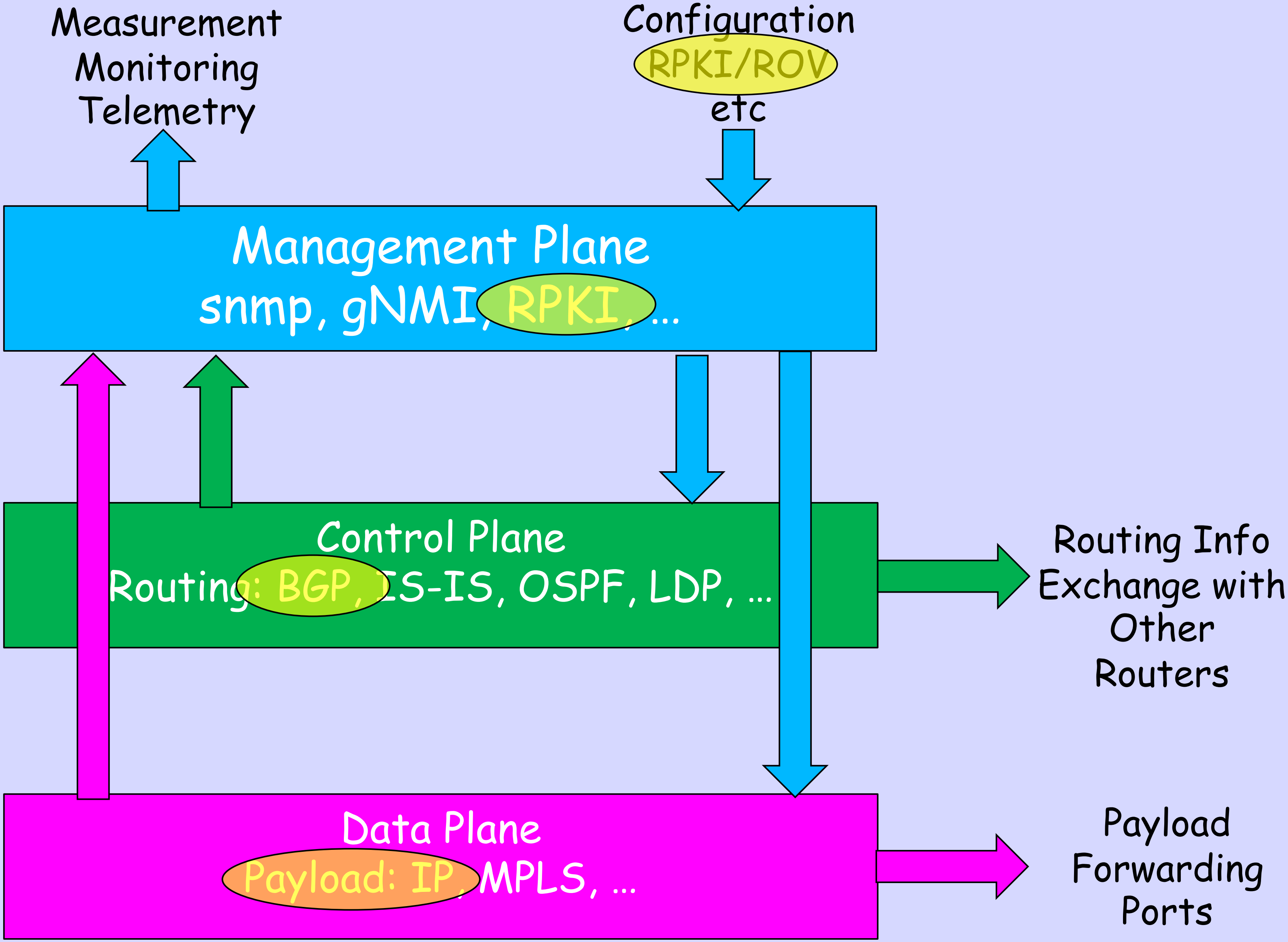
Management Plane
snmp, gNMI, RPKI, …

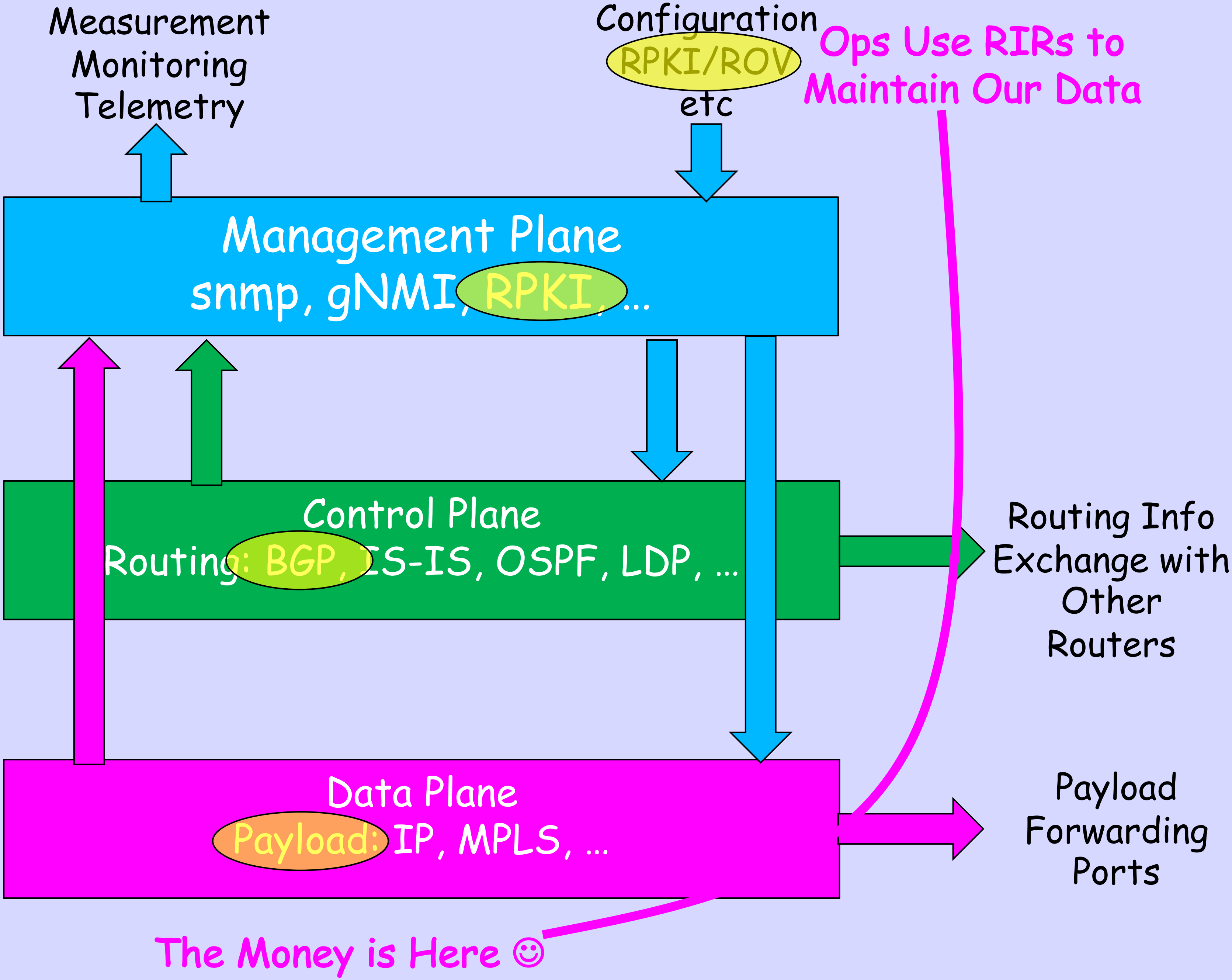
Control Plane
Routing: BGP, IS-IS, OSPF, LDP, …

Routing Info
Exchange with
Other
Routers

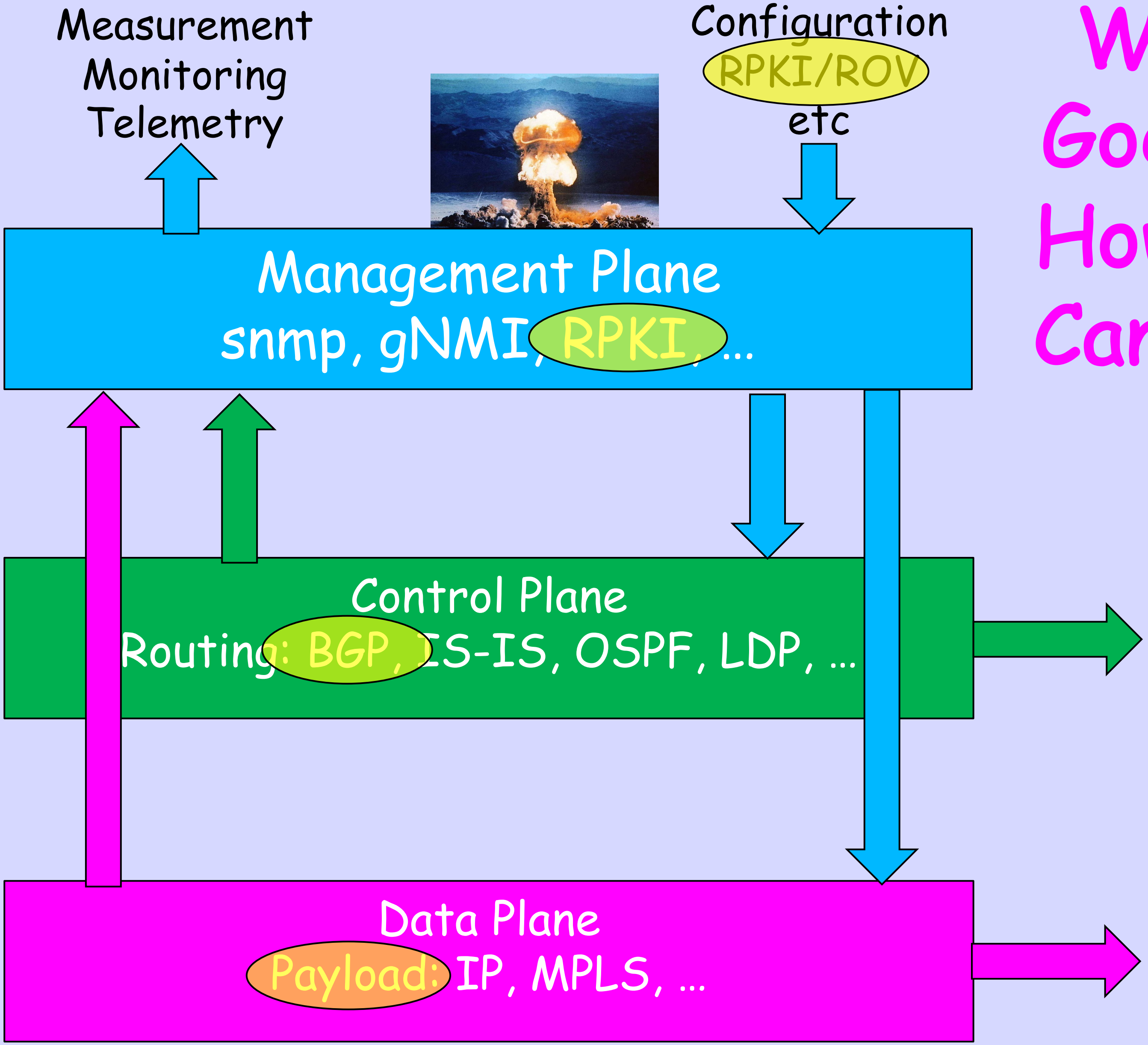
Data Plane
Payload: IP, MPLS, …

Payload
Forwarding
Ports


Measurement
Monitoring
Telemetry
Configuration
RPKI/ROV
etc
Management Plane
snmp, gNMI, RPKI, ...
Control Plane
Routing: BGP, IS-IS, OSPF, LDP, ...
Routing Info
Exchange with
Other
Routers
Data Plane
Payload: IP, MPLS, ...
Payload
Forwarding
Ports


Measurement
Monitoring
Telemetry
Configuration
RPKI/ROV
etc
Ops Use RIRs to
Maintain Our Data
Management Plane
snmp, gNMI, RPKI, …
Control Plane
Routing: BGP, IS-IS, OSPF, LDP, …
Routing Info
Exchange with
Other
Routers
Data Plane
Payload: IP, MPLS, …
Payload
Forwarding
Ports
The Money is Here ☺

When this Goes Wrong How Quickly Can I Fix it?



Measurement
Monitoring
Telemetry

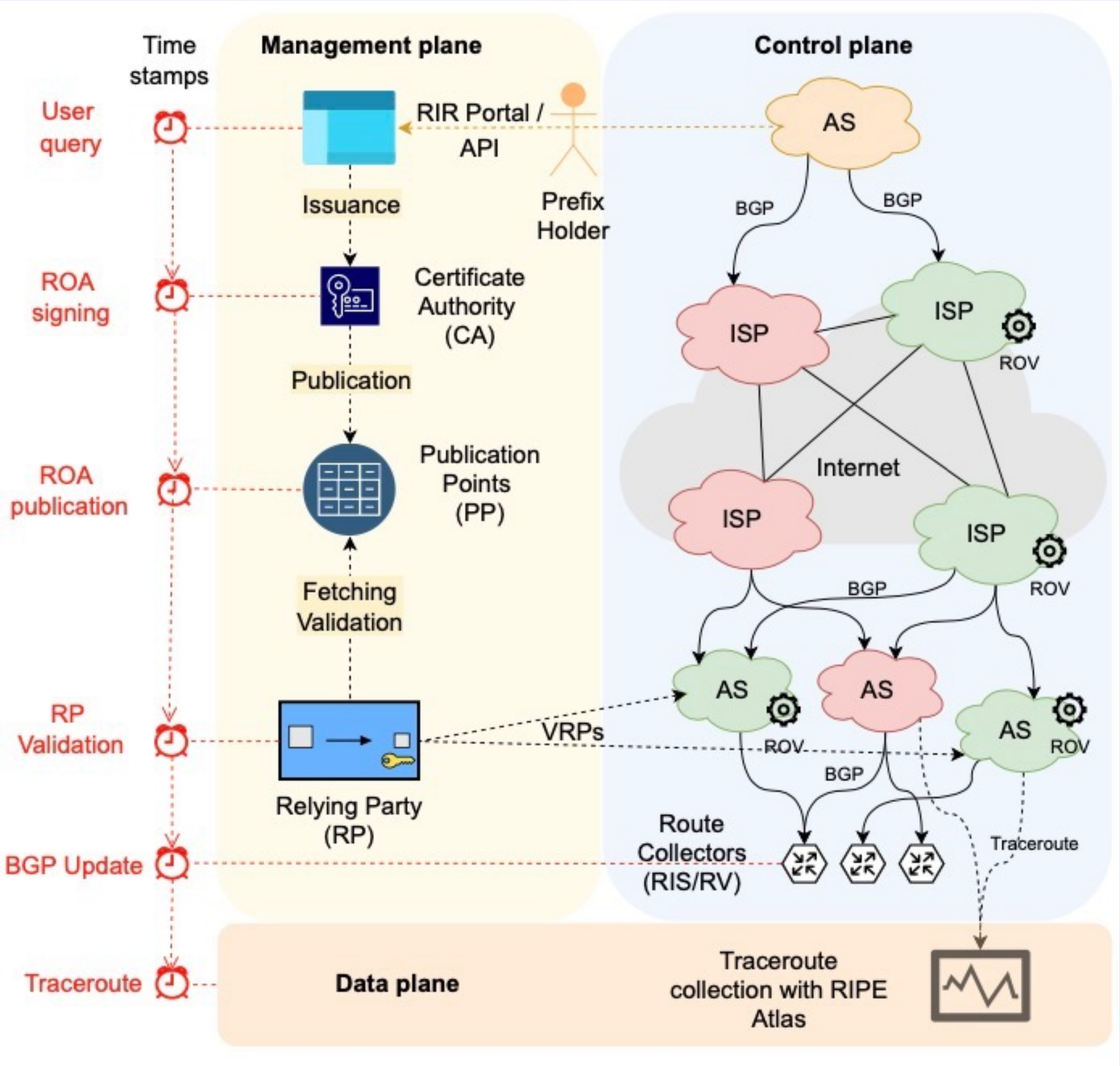
Configuration
RPKI/ROV
etc

Management Plane
snmp, gNMI, RPKI, …

Control Plane
Routing: BGP, IS-IS, OSPF, LDP, …

Data Plane
Payload: IP, MPLS, …

Experimental Set-Up


Time stamps
Management plane
Control plane
User query
ROA signing
ROA publication
RP Validation
BGP Update
Traceroute
RIR Portal / API
Prefix Holder
Issuance
Certificate Authority (CA)
Publication
Publication Points (PP)
Fetching Validation
Relying Party (RP)
VRPs
AS
BGP
BGP
ISP
ISP
ROV
Internet
ISP
ISP
ROV
BGP
AS
AS
AS
ROV
ROV
BGP
Traceroute
Route Collectors (RIS/RV)
Data plane
Traceroute collection with RIPE Atlas

The Experiment(s)

Prefixes

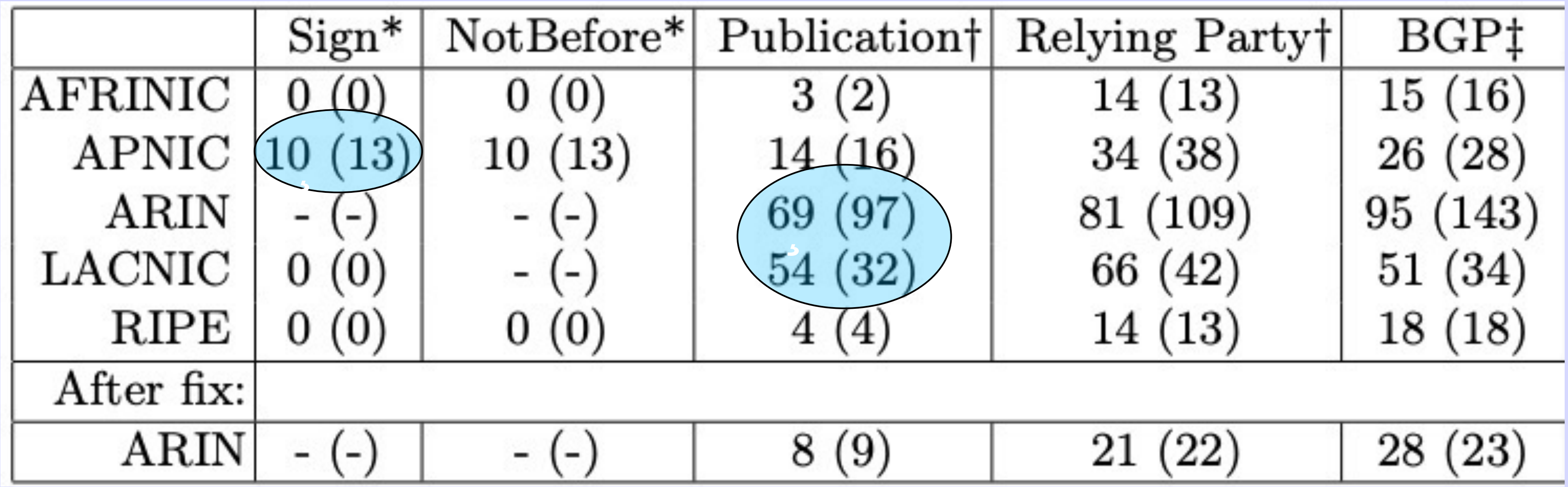
- Each of the Five RIRs loaned us a set of IPv4 /24s and IPv6 /48s
- Prefixes were announced from one AS with ROV upstreams and some direct IX peers which were non-ROV
- Another set of RIPE prefixes from 3 ASs fed by non-ROV upstreams
- Measurements taken over eleven months

ROA Beacons

- Used API or GUI screen-scraper at each RIR to Create and Delete ROAs
- Control /24s and /48s have non-varying 'good' ROAs, always Valid
- Test /24 and /48 always have an Invalidating ROA
- Then We Announced a Validating ROA once per day for half a day

Measuring RPKI Data Propagation

ROA Creation Delay (min)



	Sign*	NotBefore*	Publication†	Relying Party†	BGP‡
AFRINIC	0 (0)	0 (0)	3 (2)	14 (13)	15 (16)
APNIC	10 (13)	10 (13)	14 (16)	34 (38)	26 (28)
ARIN	- (-)	- (-)	69 (97)	81 (109)	95 (143)
LACNIC	0 (0)	- (-)	54 (32)	66 (42)	51 (34)
RIPE	0 (0)	0 (0)	4 (4)	14 (13)	18 (18)
After fix:					
ARIN	- (-)	- (-)	8 (9)	21 (22)	28 (23)

- APNIC always waits for 20 minute batches (mean 10 min)
- ARIN and LacNIC were signing in GMT (HSM)
  - But publishing in Local Time; therefore
  - NotBefore appeared to be hours before publication
  - We reported, they hacked a work-around

ROA Creation Delay

- Creation times vary significantly across RIRs, with medians ranging from a few minutes to over an hour for new ROAs to reach the publication points
- And we know of at least one NIR (not RIR) that only publishes once per day!
- Originally, APNIC only committed to once a day

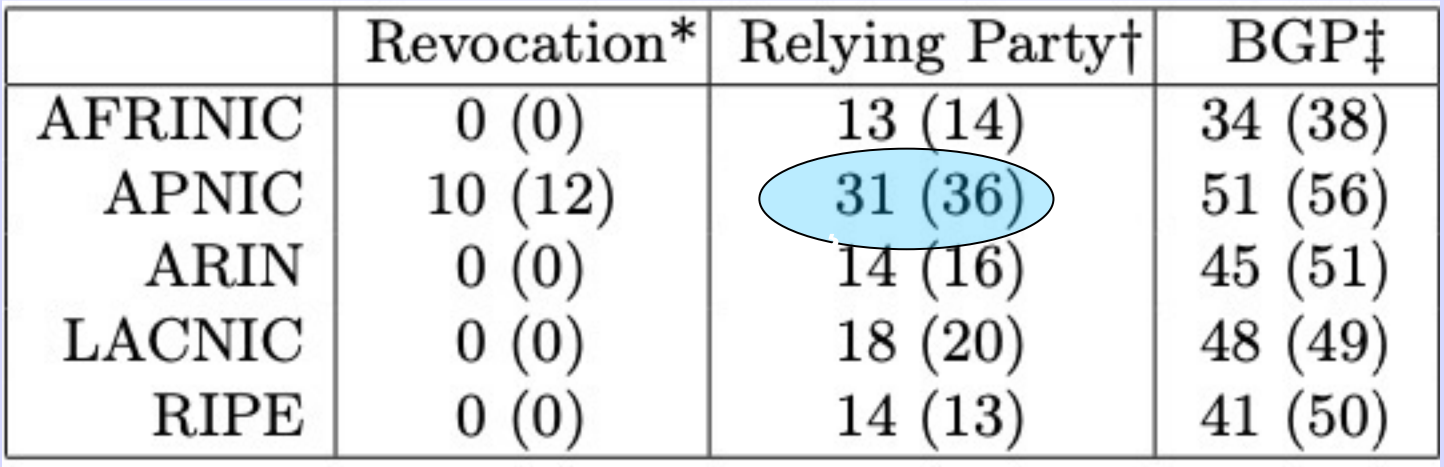
Measurement Relying Party

- One instance of RP software
- See Philip Smith’s measurements on how RPs vary ☹
- We did not run RPKI-Rtr, because we were more interested in effect on BGP
- Some RPs have HeadOfLineBlocking trying to fetch from bad Publication Points

RIPE/RIS Collectors

- Recorded Control and Test at RIPE/RIS
- If Control missing, that measurement is discarded
- This measures control plane, BGP, effect
- Used two collectors, RRC00 and RRC01. Studies have shown that's enough
- Has all the biases discussed for years

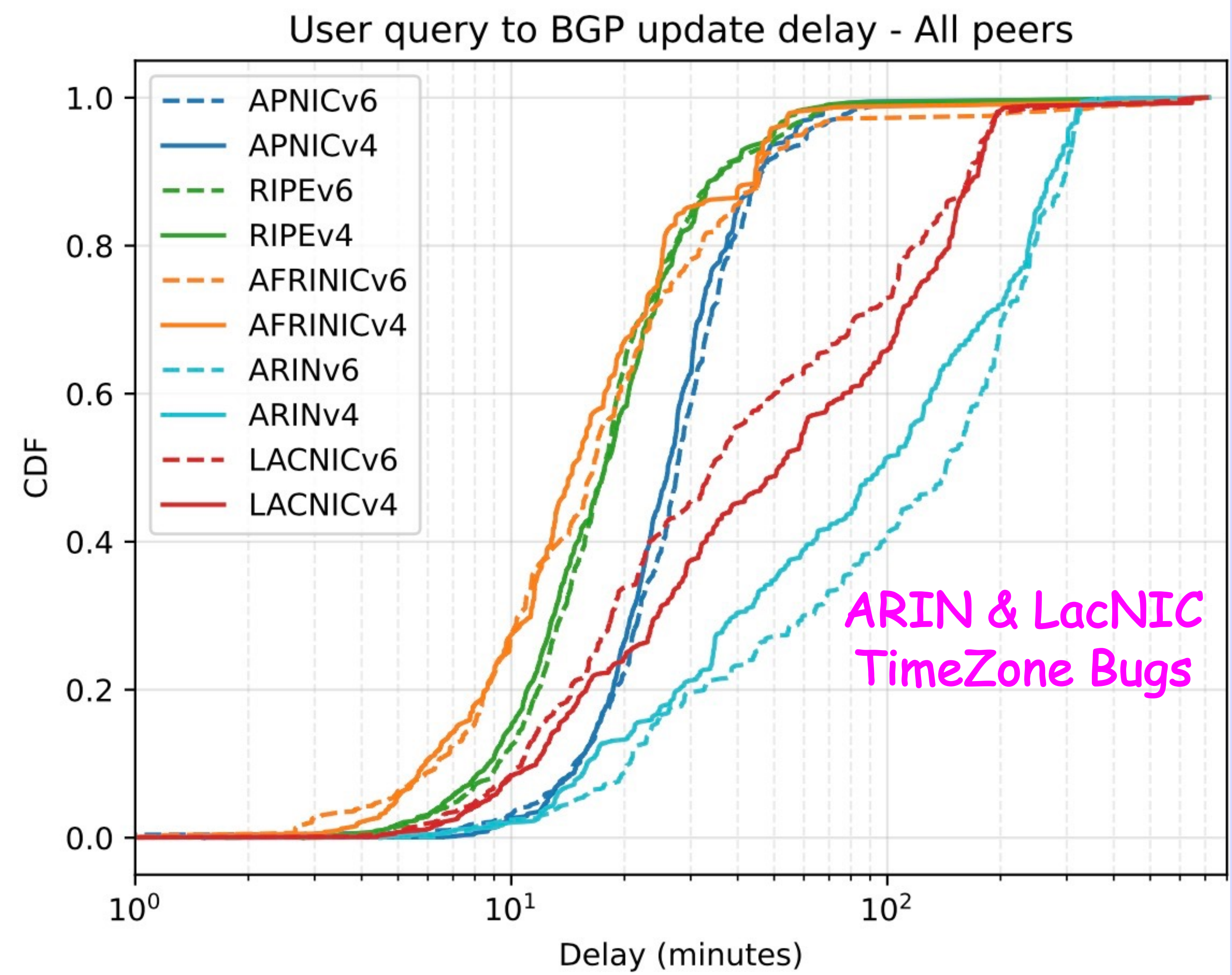
ROA Revoke Delay (min)



	Revocation*	Relying Party†	BGP‡
AFRINIC	0 (0)	13 (14)	34 (38)
APNIC	10 (12)	31 (36)	51 (56)
ARIN	0 (0)	14 (16)	45 (51)
LACNIC	0 (0)	18 (20)	48 (49)
RIPE	0 (0)	14 (13)	41 (50)

Additional APNIC delay possibly due to RP hanging
Plus APNIC has that 20 minute batching delay

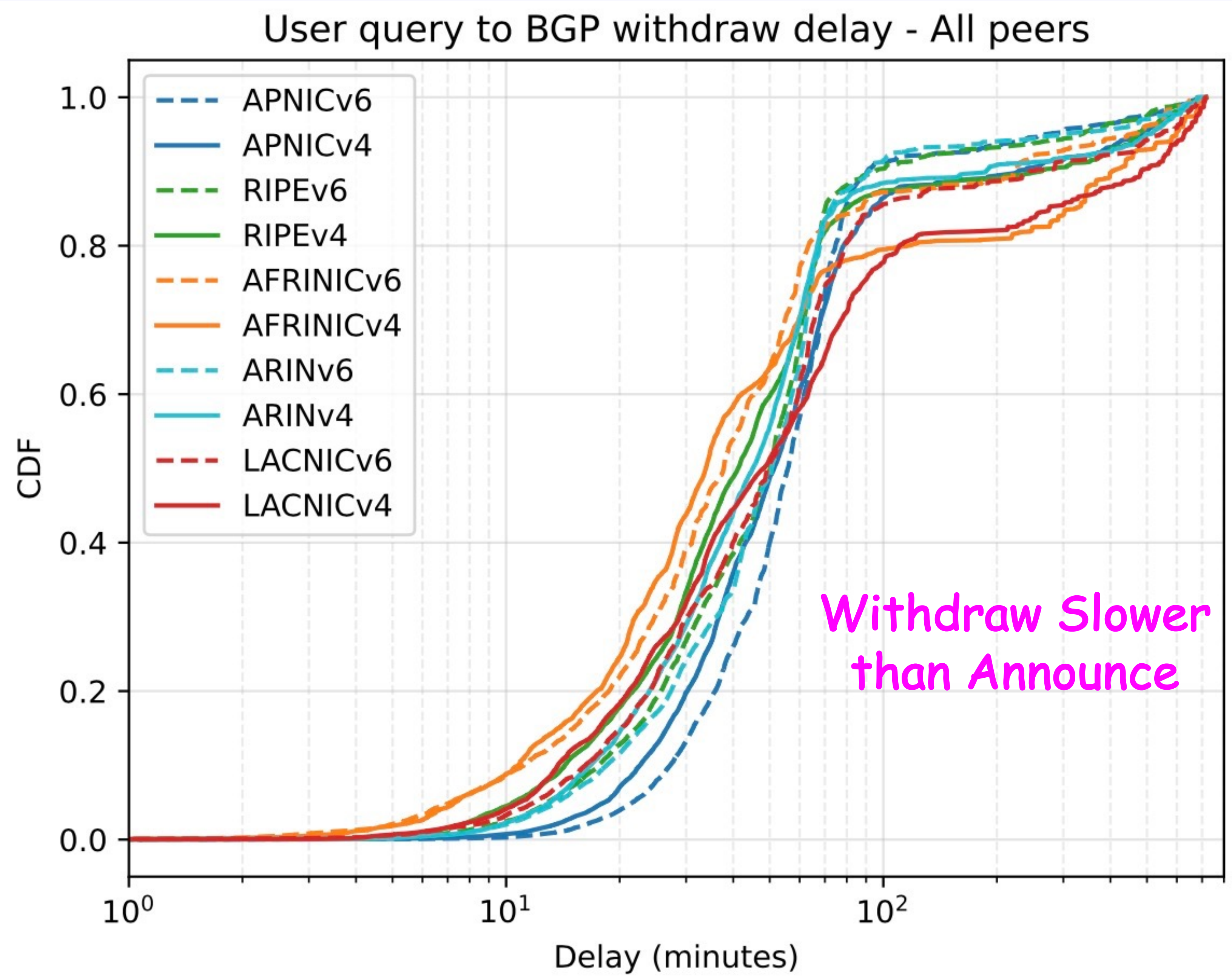
Measuring the Effect on BGP (grotty CDFs)



ARIN & LacNIC TimeZone Bugs

TimeZone Bugs

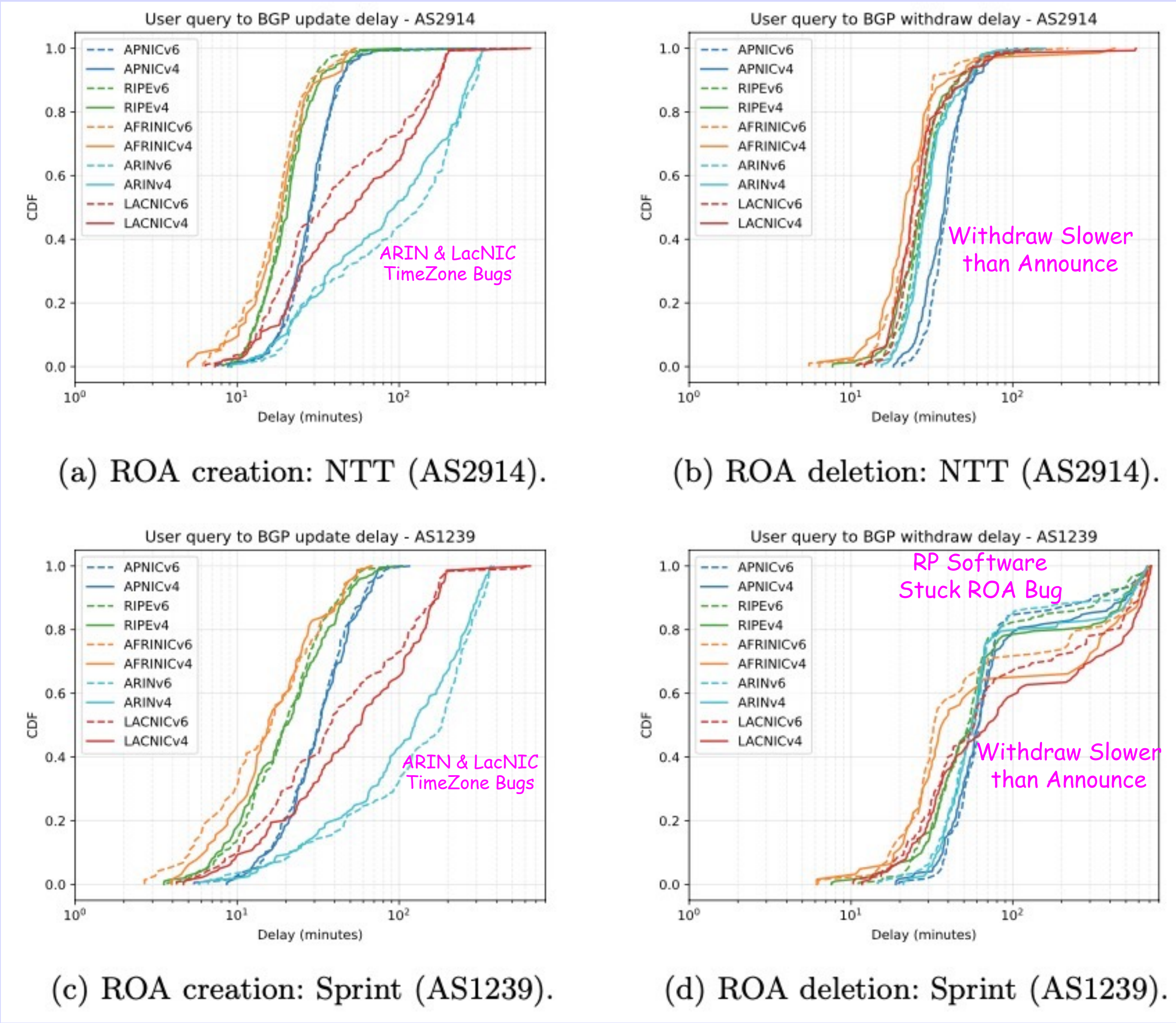
- Your HSM is Going to be in GMT
- That is a Fact of Life
- So Do Not Run Your Signer in any Time Zone other than GMT
- Or Your Signing Time and the NotBefore will be Mis-Aligned



Withdraw Slower than Announce

Withdraws are Slower

- ROV only needs one Validating ROA
- So only one cache needs to have a ROA for the router to Validate
- But all of the router's caches must have received the Withdraw from the PPs to Invalidate



(a) ROA creation: NTT (AS2914).

ARIN & LacNIC TimeZone Bugs

(b) ROA deletion: NTT (AS2914).

Withdraw Slower than Announce

(c) ROA creation: Sprint (AS1239).

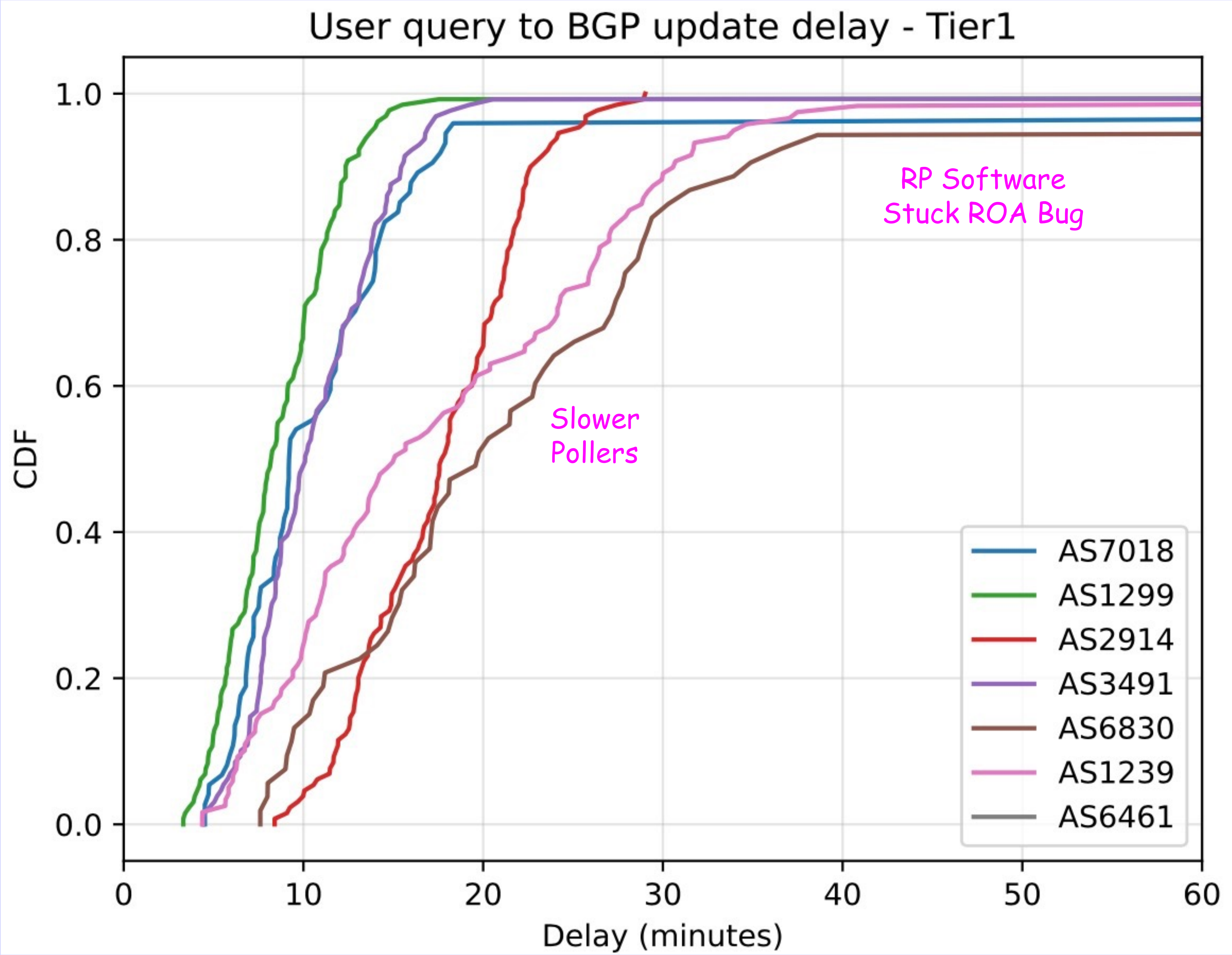
ARIN & LacNIC TimeZone Bugs

(d) ROA deletion: Sprint (AS1239).

RP Software Stuck ROA Bug

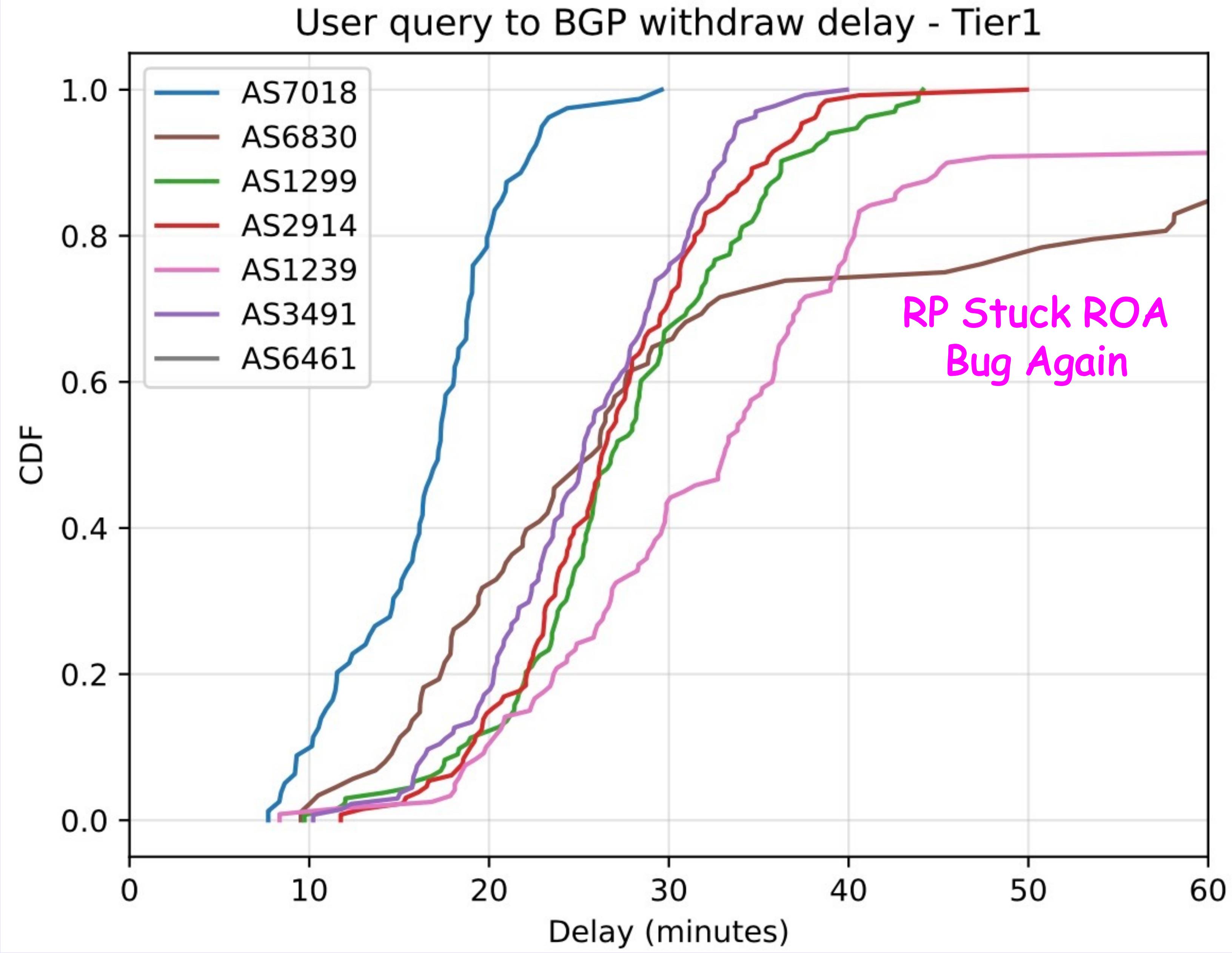
Withdraw Slower than Announce

Sprint's slower curve because RPs pull less frequently than NTT's, &/or sucky RP software
Sprint starts a bit earlier because routers poll RP caches more frequently than NTT's
Confirmed with friends at Sprint and NTT



RP Software Stuck ROA Bug

Slower Pollers

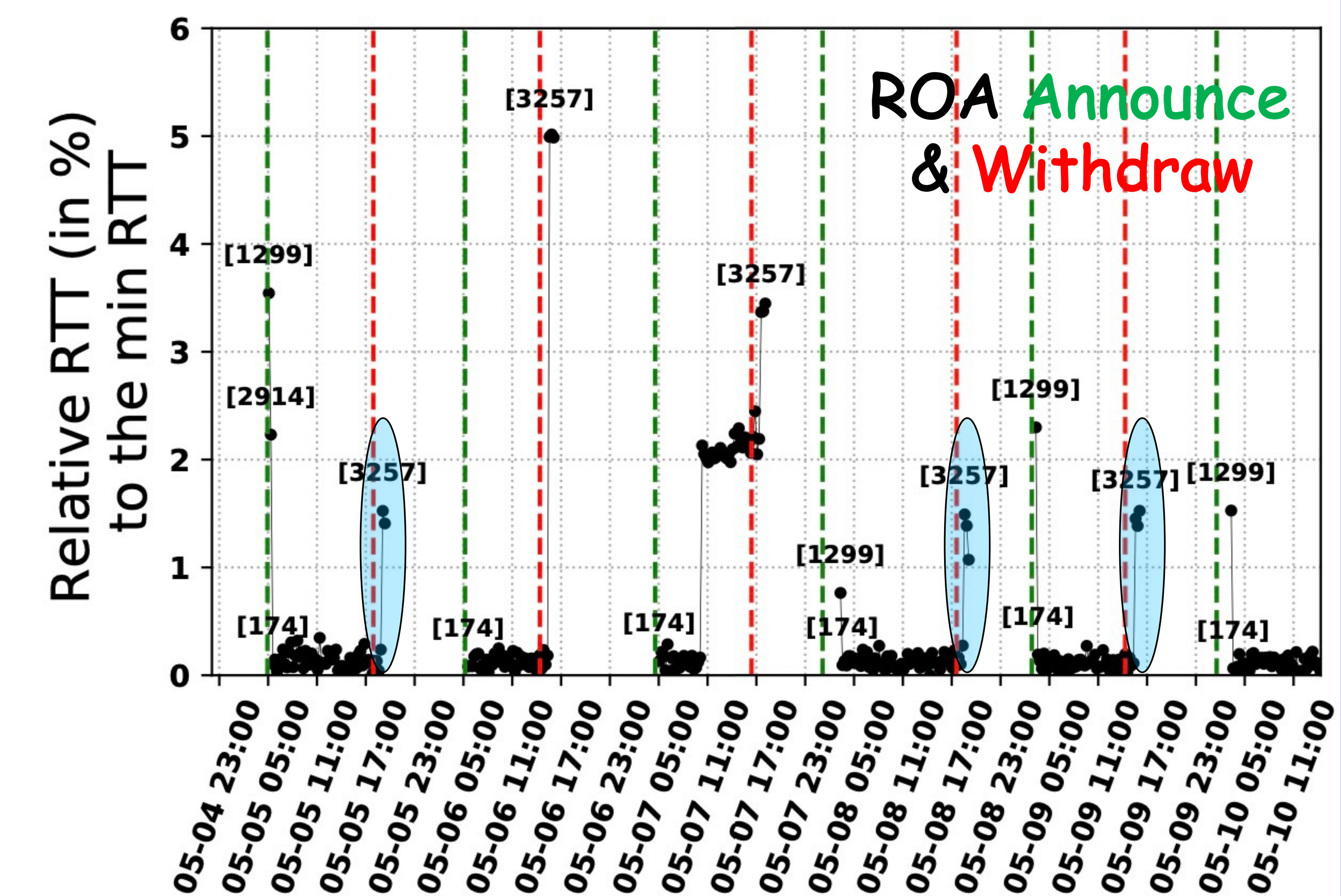


RP Stuck ROA
Bug Again

Data Plane Measurement

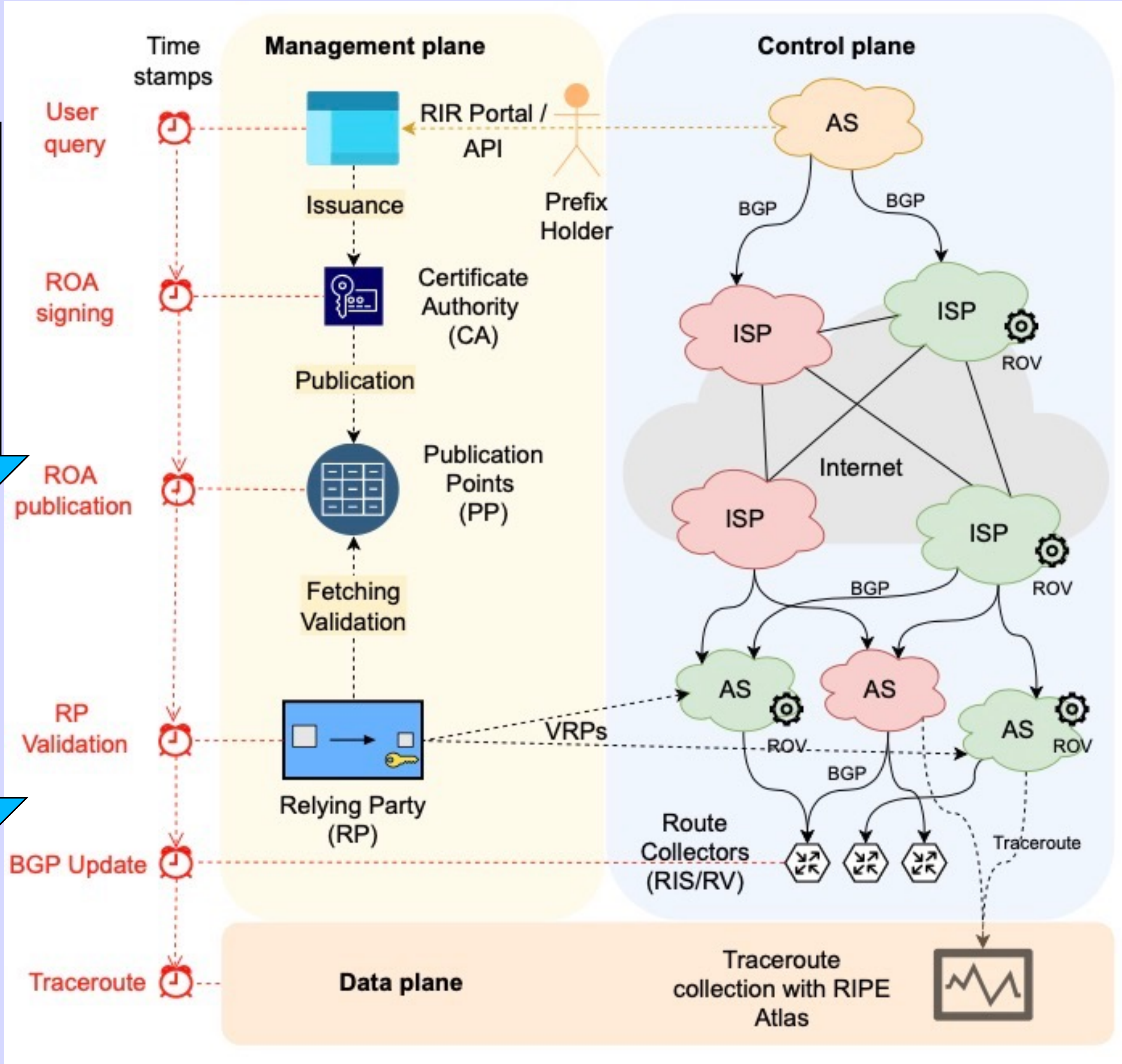
- Ran *traceroute* from Atlas Probes
- To the Test prefixes
- Every 15 minutes
- Result similar to BGP at RIPE/RIS, but
- Path hunting after a Withdraw is beautifully obvious

Data Plane & Path Hunting



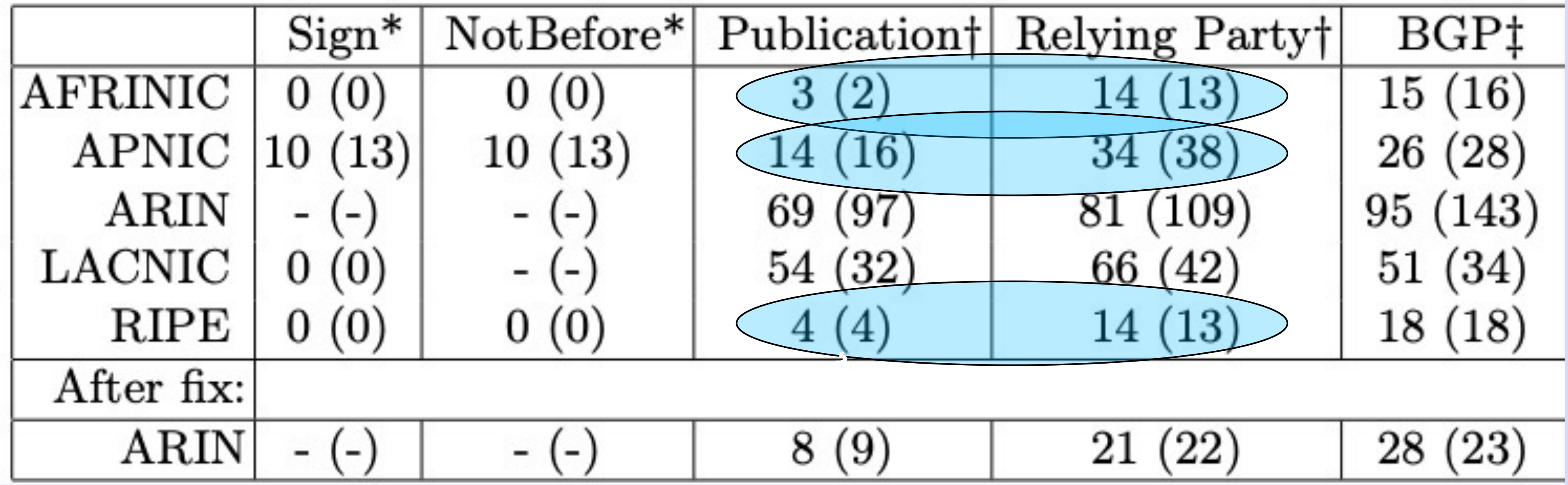
RIR Delay

ISP Delay



Some Problems and Anomalies

ISP Polling Delay Is Big



	Sign*	NotBefore*	Publication†	Relying Party†	BGP‡
AFRINIC	0 (0)	0 (0)	3 (2)	14 (13)	15 (16)
APNIC	10 (13)	10 (13)	14 (16)	34 (38)	26 (28)
ARIN	- (-)	- (-)	69 (97)	81 (109)	95 (143)
LACNIC	0 (0)	- (-)	54 (32)	66 (42)	51 (34)
RIPE	0 (0)	0 (0)	4 (4)	14 (13)	18 (18)
After fix:					
ARIN	- (-)	- (-)	8 (9)	21 (22)	28 (23)

Assume ARIN and LacNIC
TimeZone anomalies are fixed

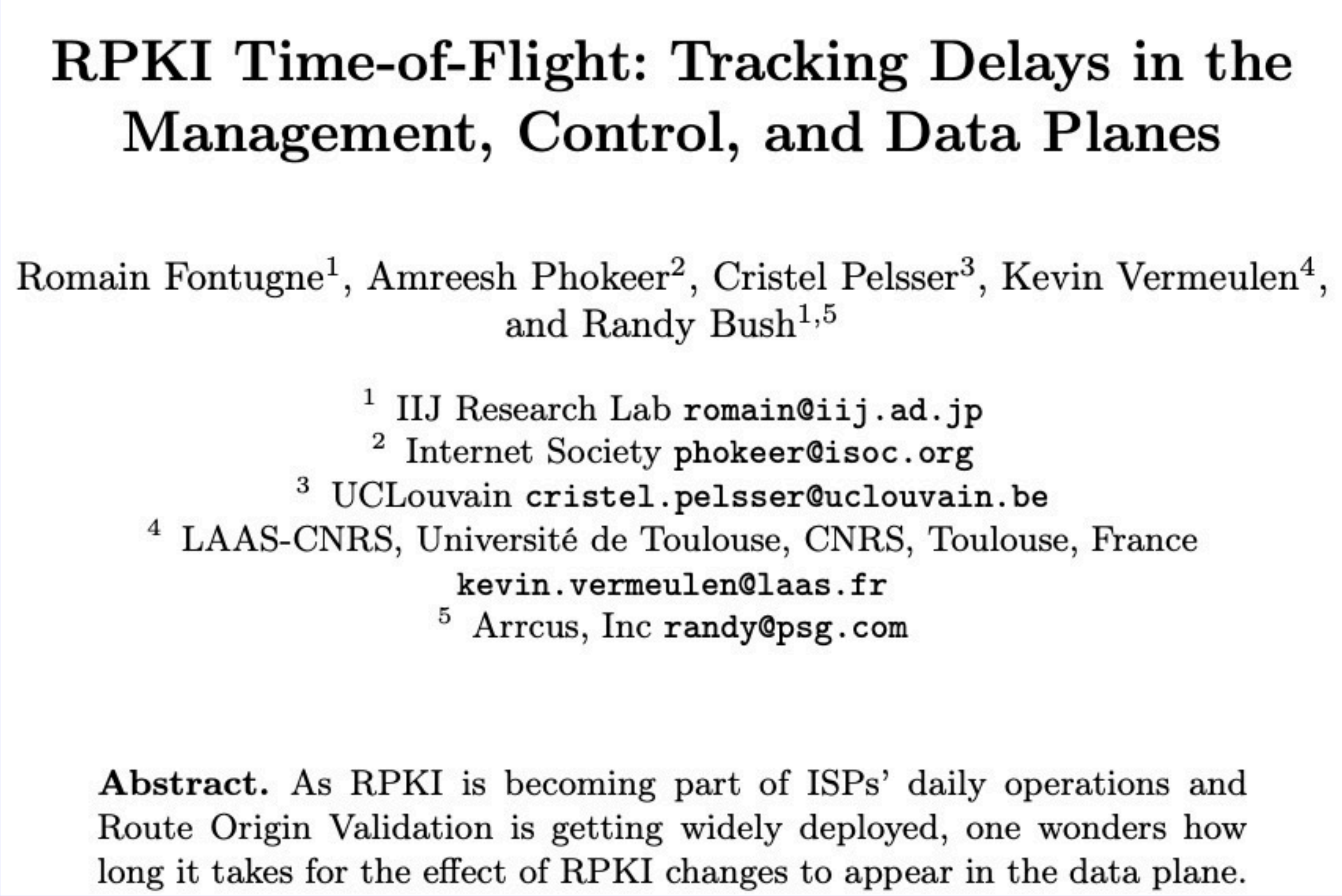
Problems

- BGP propagates in minutes. RPKI propagates in O(hour). This has business impacts, e.g.
  - Time to Repair for a bad ROA
  - Time to authorize a DDoS mitigator
- Two RIRs with HSM in GMT and CAs in Local Time Zone. Reported and 'fixed'
- Some RPs Head of Line Block on Slow PPs
- ROA Anatomy varies between RIRs

Limitations of Study

- Relying Party software:
  - Fixed fetch rate so poor resolution
  - Only one RP software package used
- Did not measure RP to Router. But that is Notify driven so *should be* fast
- Did not measure delegated CAs
- RIR API/Screen-Scrape Unreliable

From the Paper in PAM 2023



RPKI Time-of-Flight: Tracking Delays in the Management, Control, and Data Planes

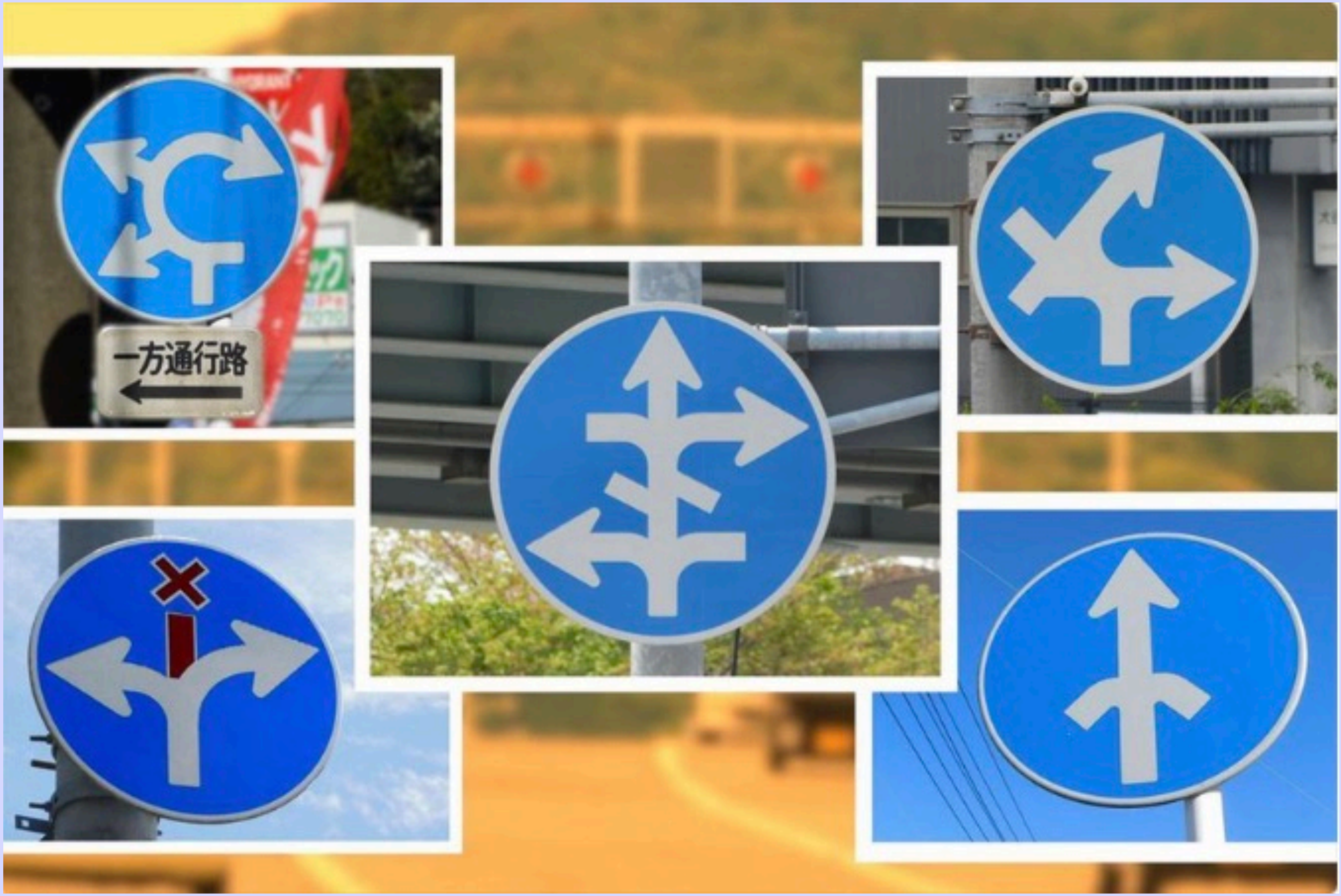
Romain Fontugne[1], Amreesh Phokeer[2], Cristel Pelsser[3], Kevin Vermeulen[4], and Randy Bush[1,5]

[1] IIJ Research Lab romain@iij.ad.jp
[2] Internet Society phokeer@isoc.org
[3] UCLouvain cristel.pelsser@uclouvain.be
[4] LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France
kevin.vermeulen@laas.fr
[5] Arrcus, Inc randy@psg.com

Abstract. As RPKI is becoming part of ISPs' daily operations and Route Origin Validation is getting widely deployed, one wonders how long it takes for the effect of RPKI changes to appear in the data plane.

https://archive.psg.com/pam2023-rov-ecosystem.pdf

What We Can Do?



Go Faster

- CAs/RIRs Publish Very Frequently
- RPs Poll 10 mins or more frequently when using RRDP
- Caching and If-Modified-Since means the load on PPs is negligible
- RPs poll frequently, but not too frequently, if rsync. Maybe 30 mins
- Yes, this discourages rsync

As Protocol Designers

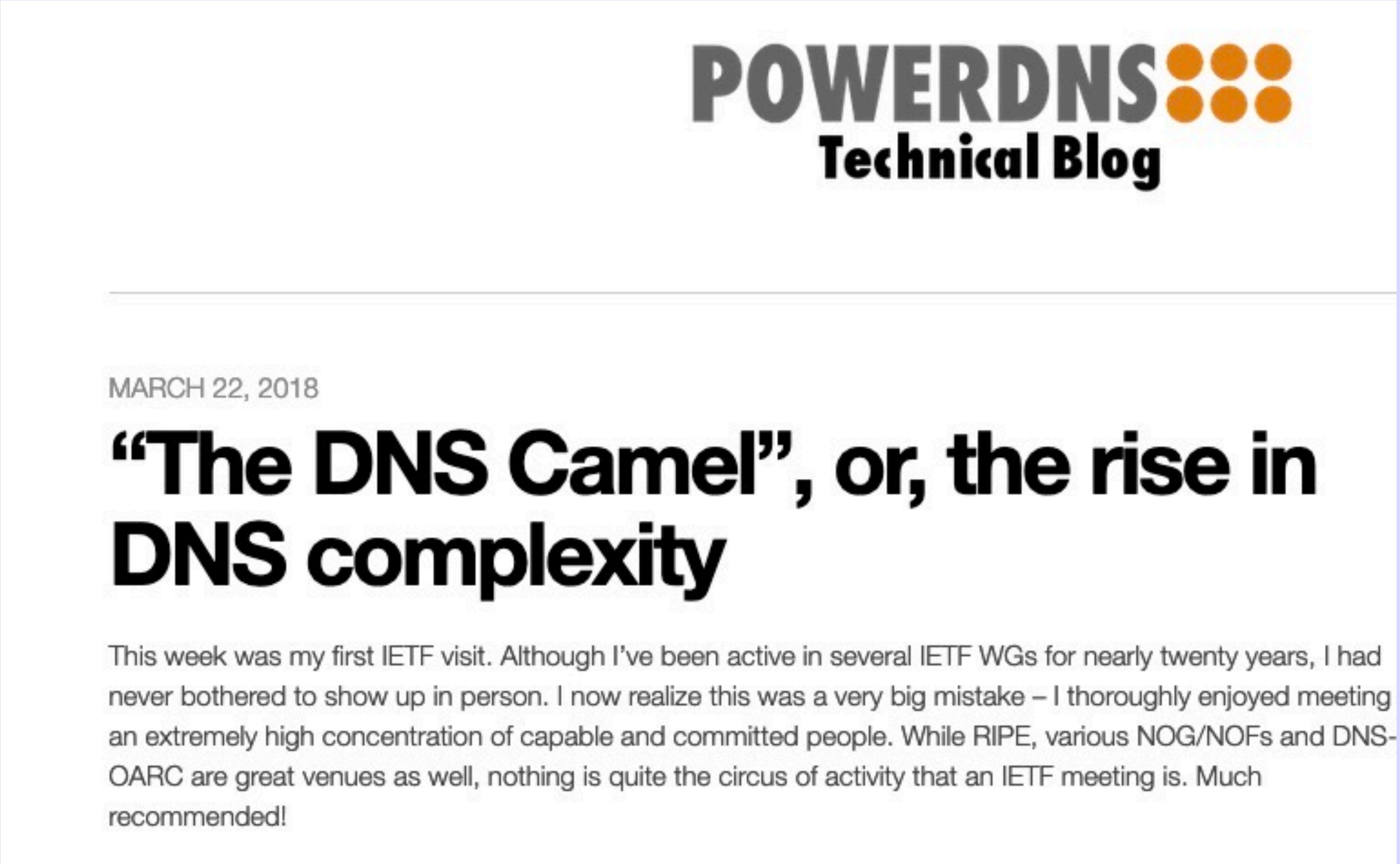
- BGP is the only large scale ‘push’ protocol we have, but
- BGP Transport is
  - Dangerously Shared Fate
  - Unordered, Reordering is Guaranteed
- DNS does not handle Make Before Break

`draft-bates-bgp4-nlri-orig-verif` (1998)

We Fantasize About a Flooding Protocol for the Inter-Provider Management Plane That Is Immune to Routing Attack

A Warning
and
a Plea for Simplicity

Bert Hubert in 2018



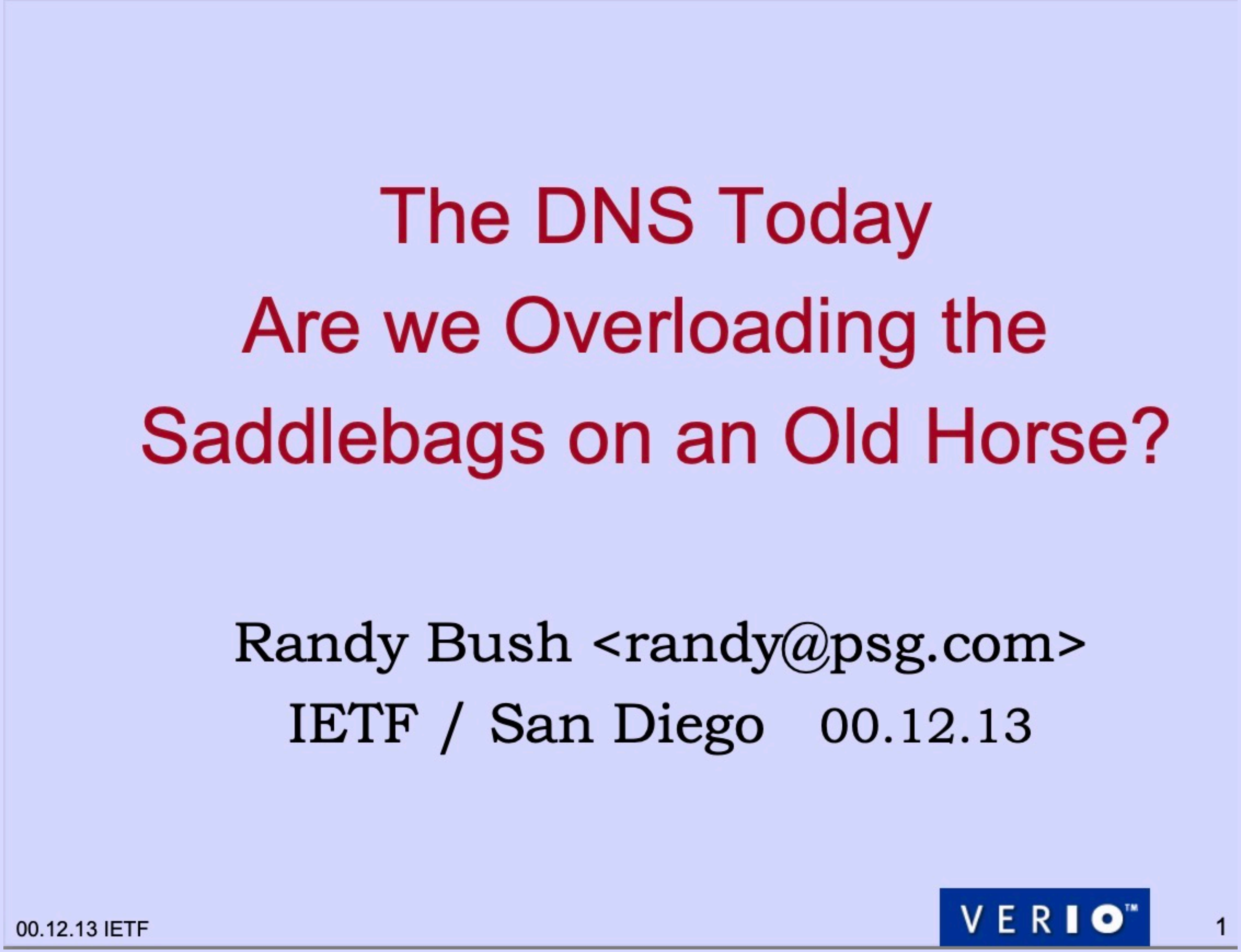
POWERDNS
Technical Blog

MARCH 22, 2018

"The DNS Camel", or, the rise in DNS complexity

This week was my first IETF visit. Although I've been active in several IETF WGs for nearly twenty years, I had never bothered to show up in person. I now realize this was a very big mistake – I thoroughly enjoyed meeting an extremely high concentration of capable and committed people. While RIPE, various NOG/NOFs and DNS-OARC are great venues as well, nothing is quite the circus of activity that an IETF meeting is. Much recommended!

But 18 Years Earlier



The DNS Today
Are we Overloading the Saddlebags on an Old Horse?

Randy Bush <randy@psg.com>

IETF / San Diego 00.12.13

00.12.13 IETF

VERIO

1

The computing scientist's main challenge is not to get confused by the complexities of his own making
-- E. W. Dijkstra

Permissionless Research Thanks To

Arrcus
Cisco
Equinix
Google
Juniper
NTT
Sprint

For Donated

- Rack Space
- Bandwidth
- Routers
- Switches
- Servers
- Etc. Etc.

Questions?

And Position Statements Pretending to be Questions ☺