

The RPKI & Origin Validation

APRICOT / Kuala Lumpur

2010.03.03

Rob Austein <sra@isc.org>

Steve Bellovin <smb@cs.columbia.edu>

Randy Bush <randy@psg.com>

And a cast of thousands! Well, dozens :)

Routing is Very Fragile

- Mis-originations happen weekly, some very notable, YouTube incident
- How long can we survive on "The Web as Random Acts of Kindness," TED Talk by Jonathan Zittrain?
- We need to fix this before we are crucified in the WSJ like Toyota

Routing Mistakes

- Routing errors are significant and have very high customer impact
- February, someone mis-announced an IIJ prefix, much customer damage
- 99% of mis-announcements are accidental originations of someone else's prefix -- Goog, UU, IIJ, ...

Why Origin Validation?

- Prevent YouTube accident
- Prevent 7007 accident, UU/Sprint 2 days!
- Prevents most accidental announcements
- Does not prevent malicious path attacks such as the Kapela/Pilosov DefCon attack
- That requires "Path Validation" and locking the data plane to the control plane, the next steps, by my children

The Goal

- Keep the Internet working!!!
- Seriously reduce routing damage from mis-configuration, mis-origination

Non-Goal

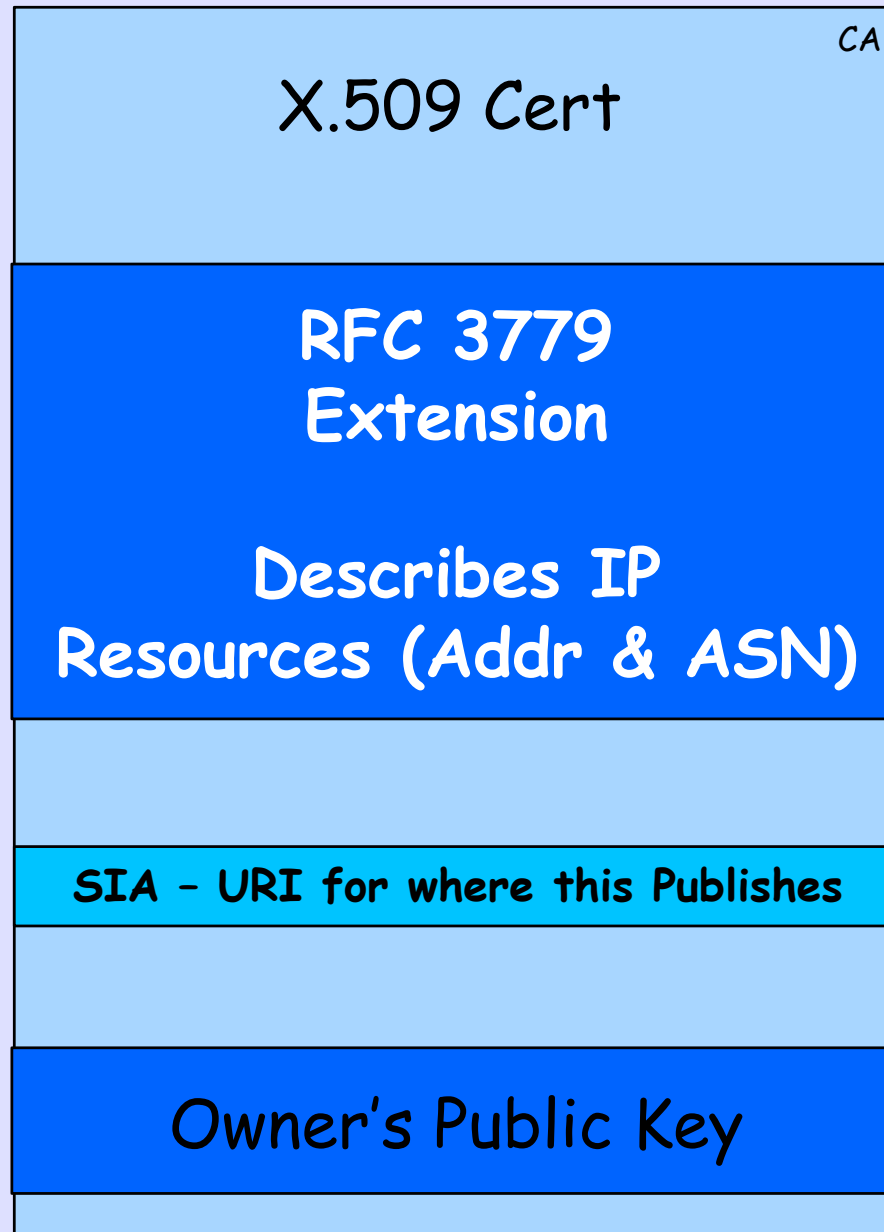
- Keep RIRs in business by selling X.509 Certificates
- This is extortion = RIRs control routing

This is Not New

- 1986 - Bellovin identifies vulnerability
- 2000 - S-BGP - X.509 PKI to support Secure BGP - Kent, Lynn, et al.
- 2003 - NANOG S-BGP Workshop
- 2006 - ARIN & APNIC start work on RPKI. RIPE starts in 2008.
- 2009 - RPKI Open Testbed and running code in test routers

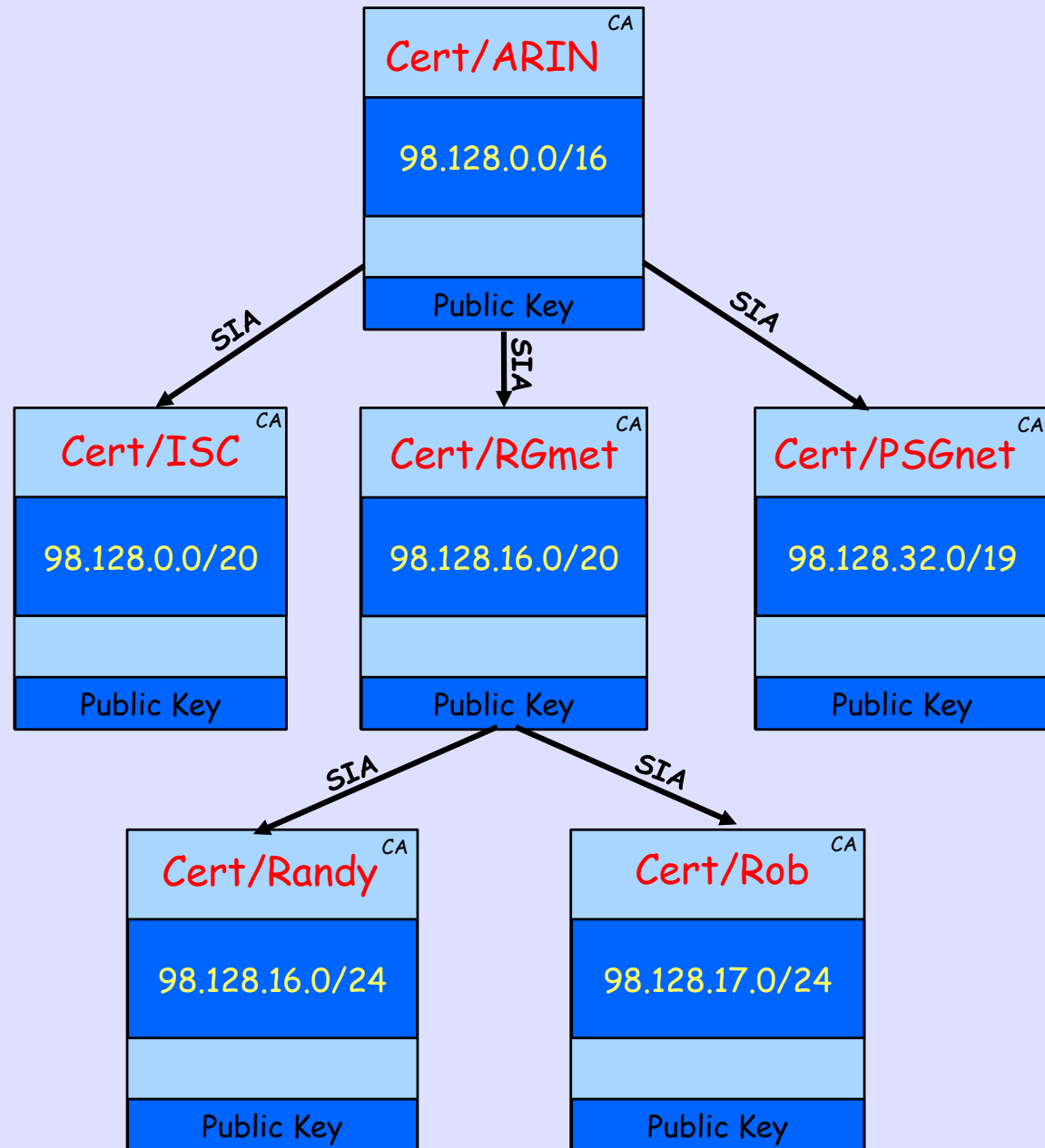
Resource
Public
Key
Infrastructure
(RPKI)

X.509 Certificate w/ 3779 Ext



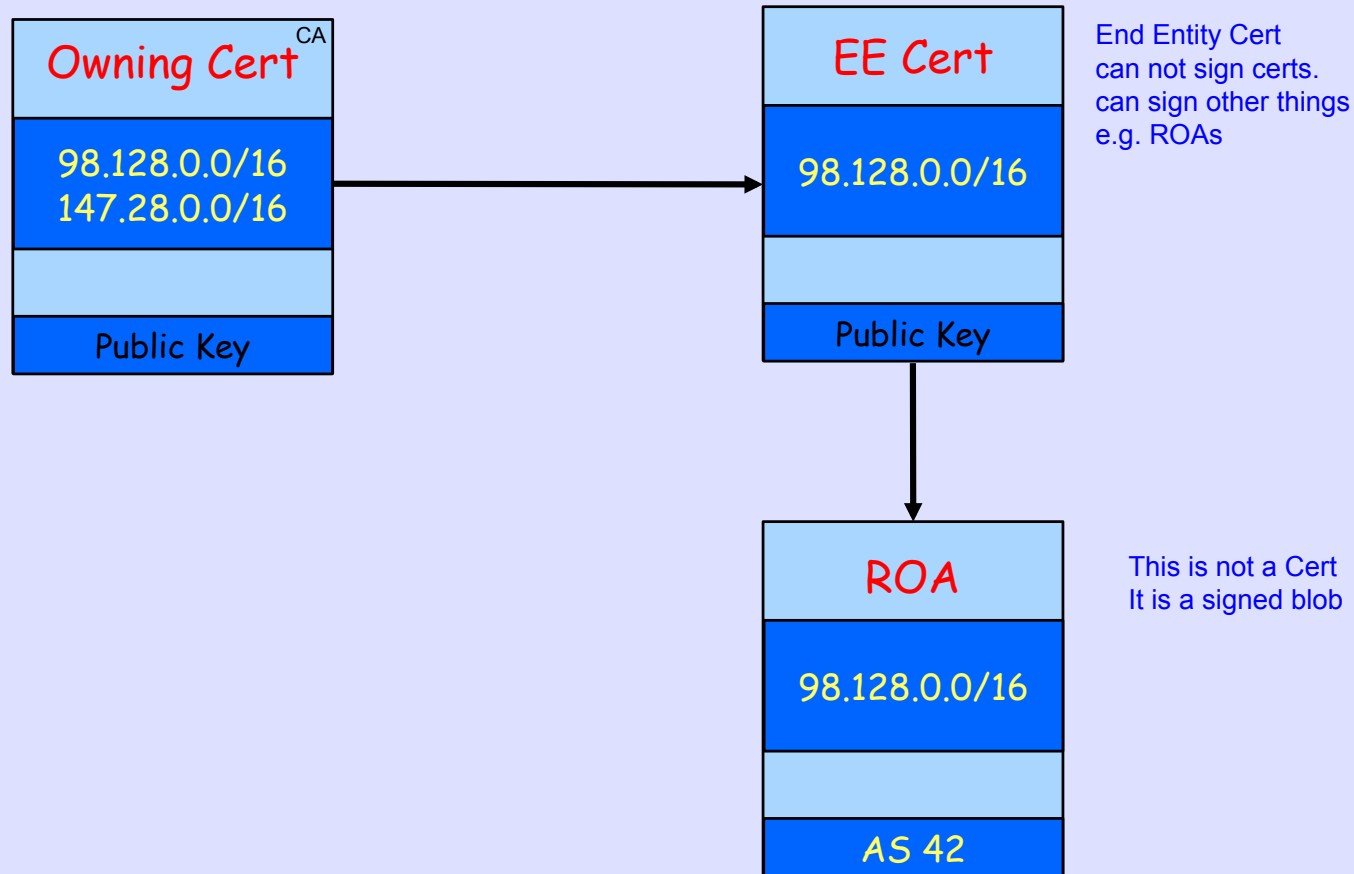
Being
Developed & Deployed
by
RIRs and Operators

Certificate Hierarchy follows Allocation Hierarchy



That's Who Owns It
but
Who May Route It?

Route Origin Authorization (ROA)

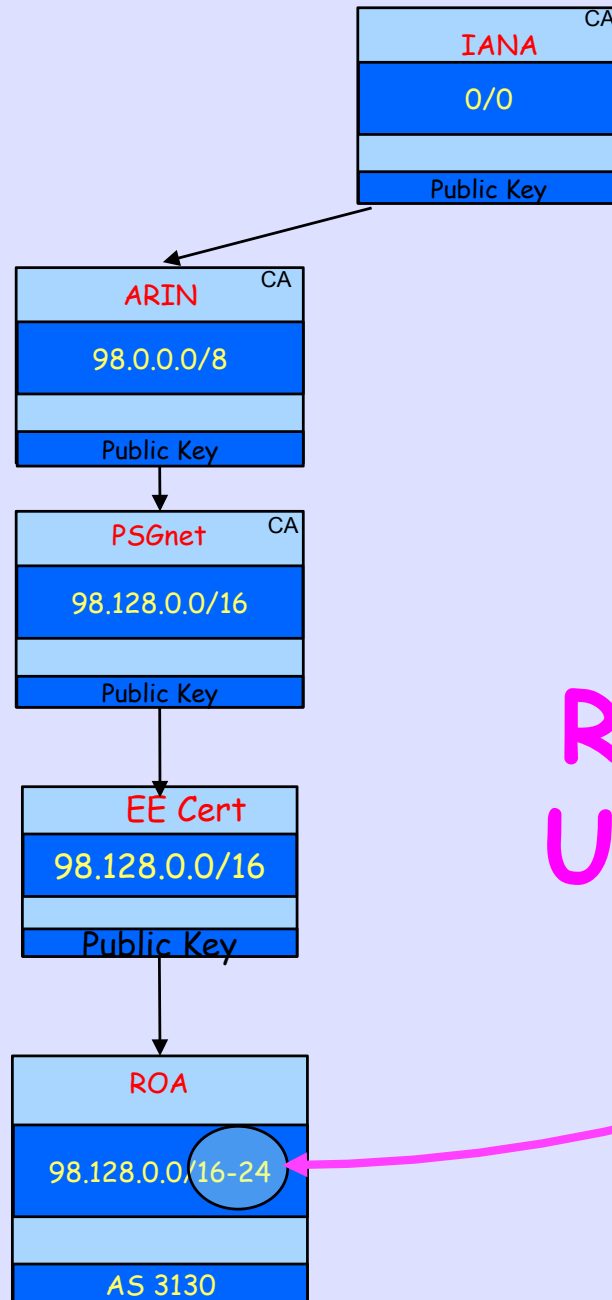


PSGnet /16
Experimental
Allocation
from ARIN

Announces
256 /24s



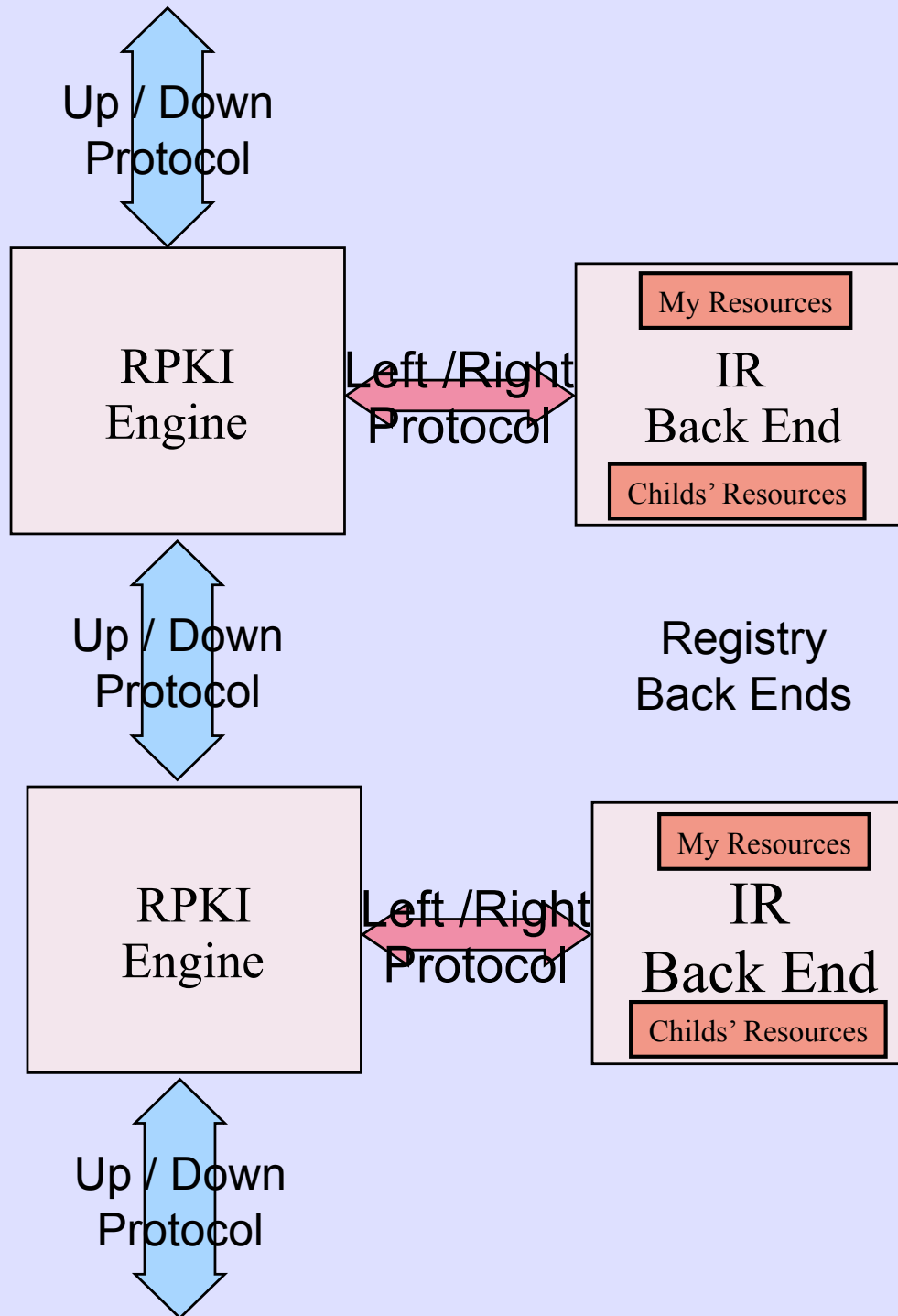
Too Many EE Certs and ROAs, Yucchhy!



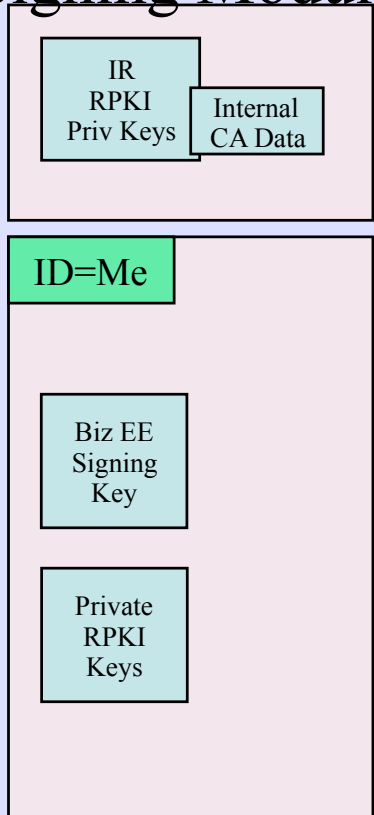
ROA Aggregation Using Max Length

Running Code And the Three RPKI Protocols

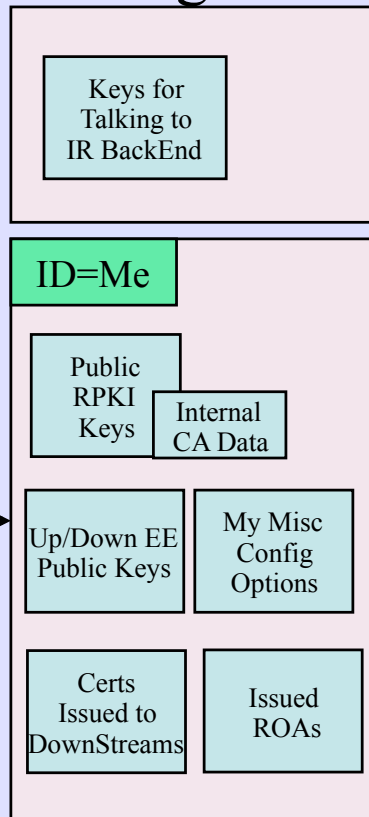
Simple Parent and Simple Child



[Hardware]
Signing Module

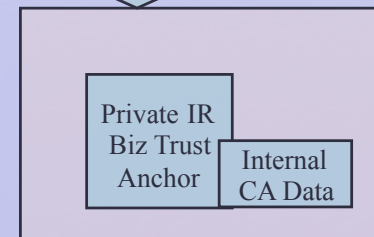
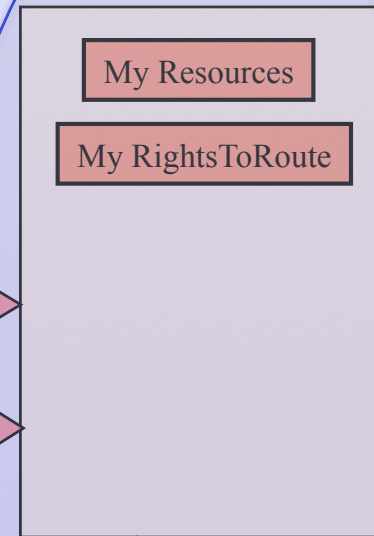


RPKI Engine

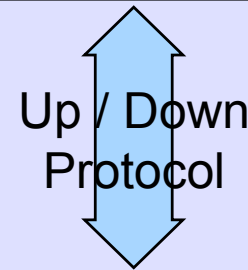
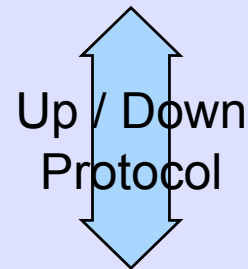


Prototype
of Basic
Back End

IR
Back End



Business
Key/Cert
Management

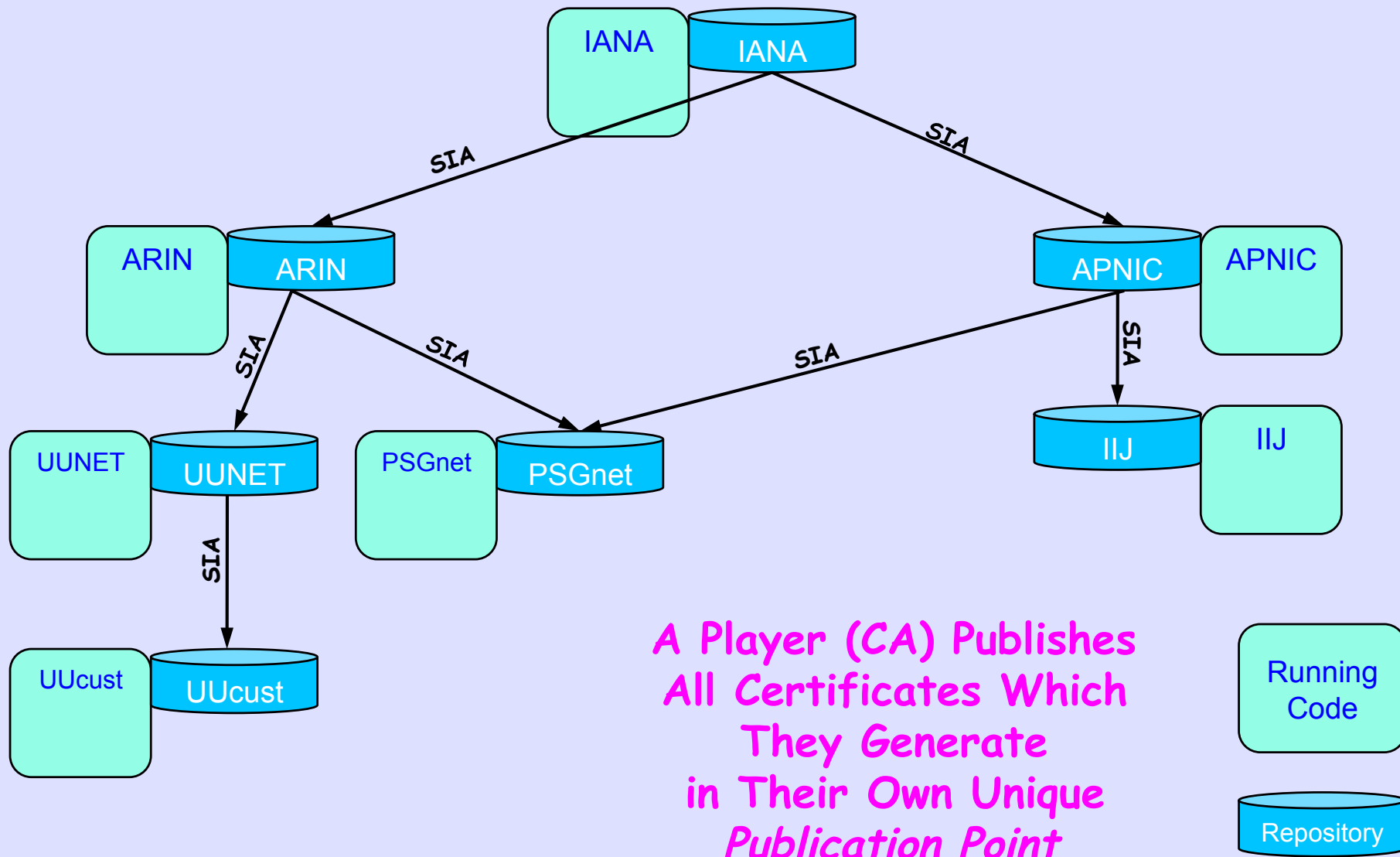


Big, Centralized, & Scary We Don't Do This

RPKI DataBase

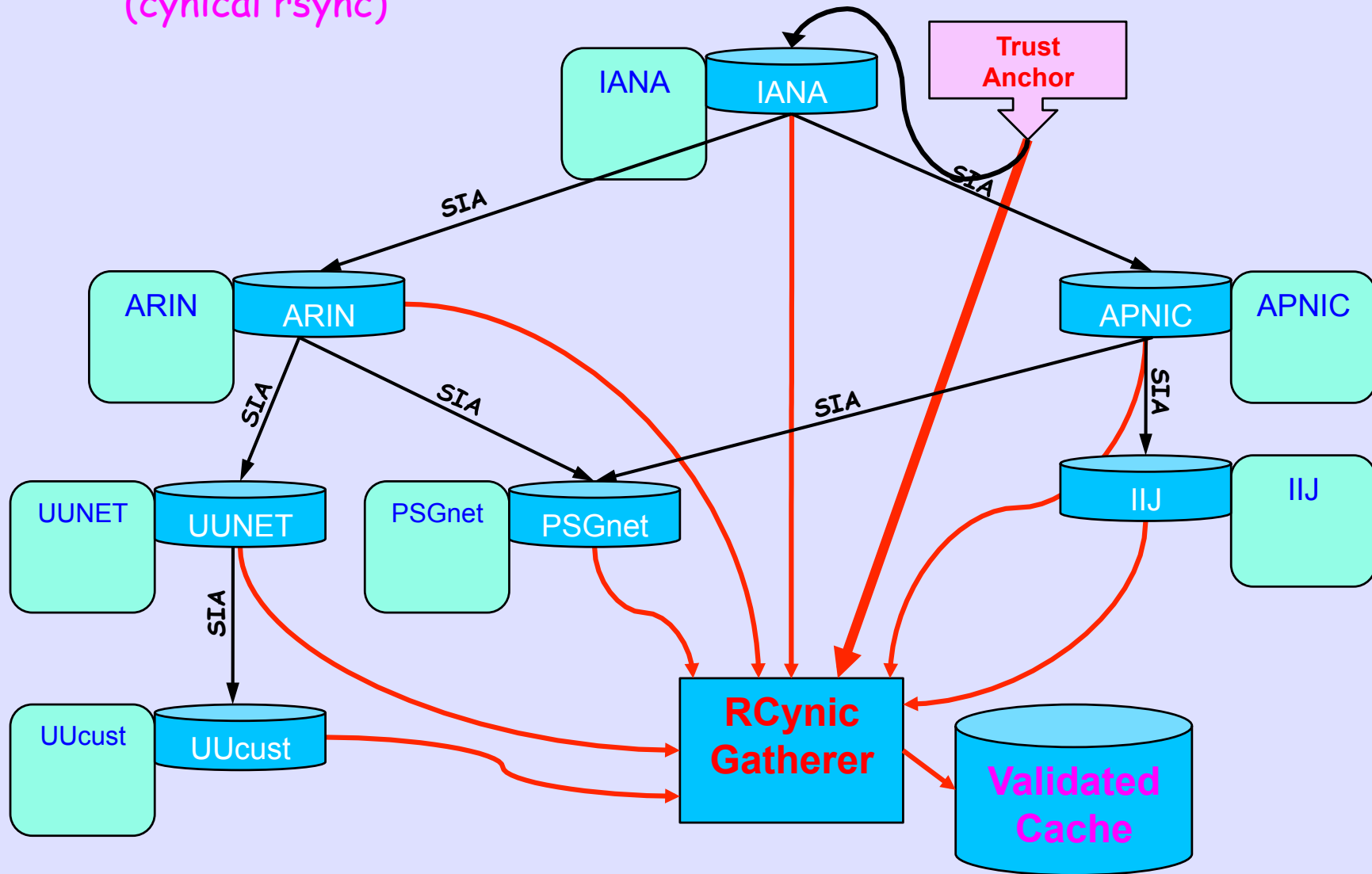
**IP Resource Certs
ASN Resource Certs
Route Origin Attestations**

Distributed RPKI DataBase



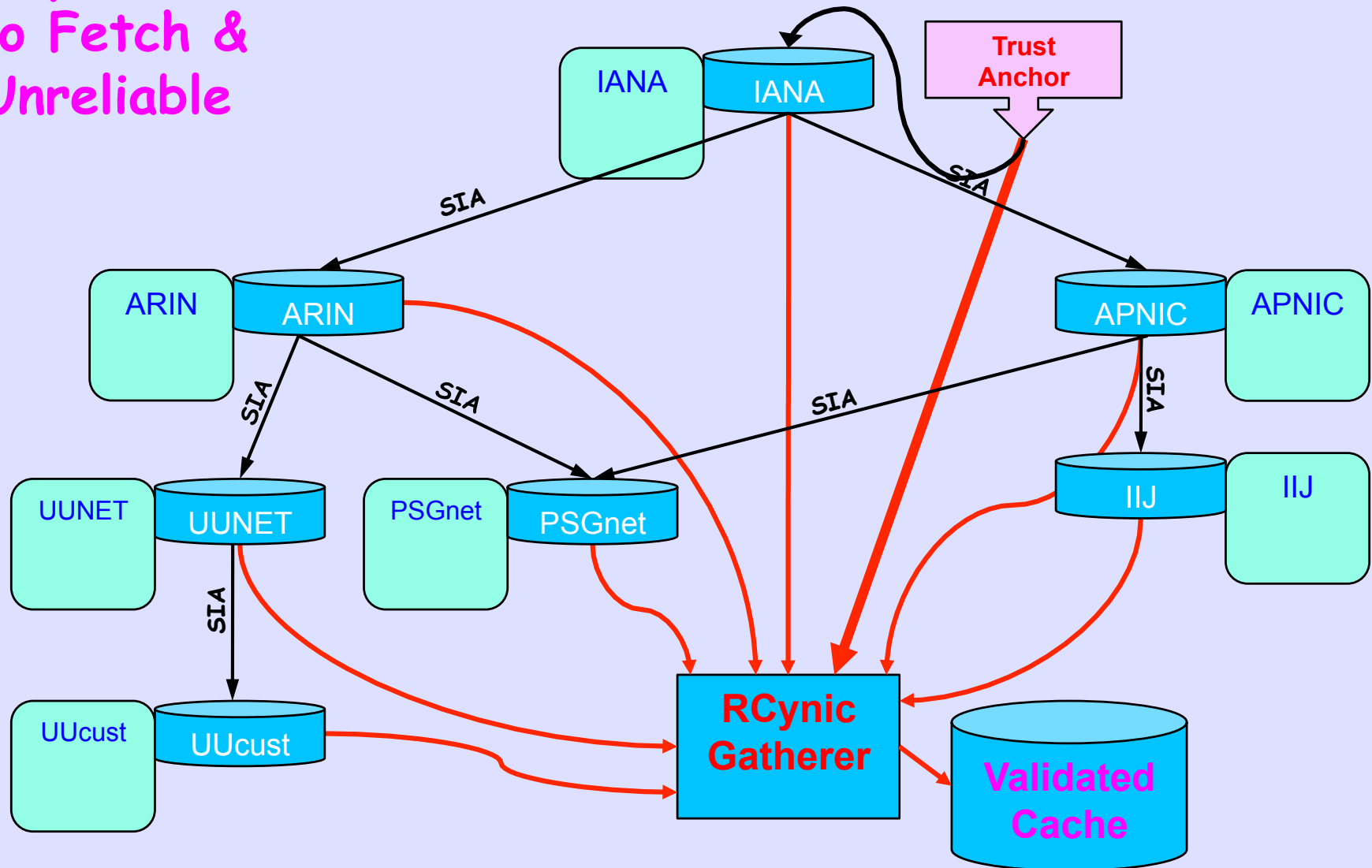
RCynic Cache Gatherer

(cynical rsync)

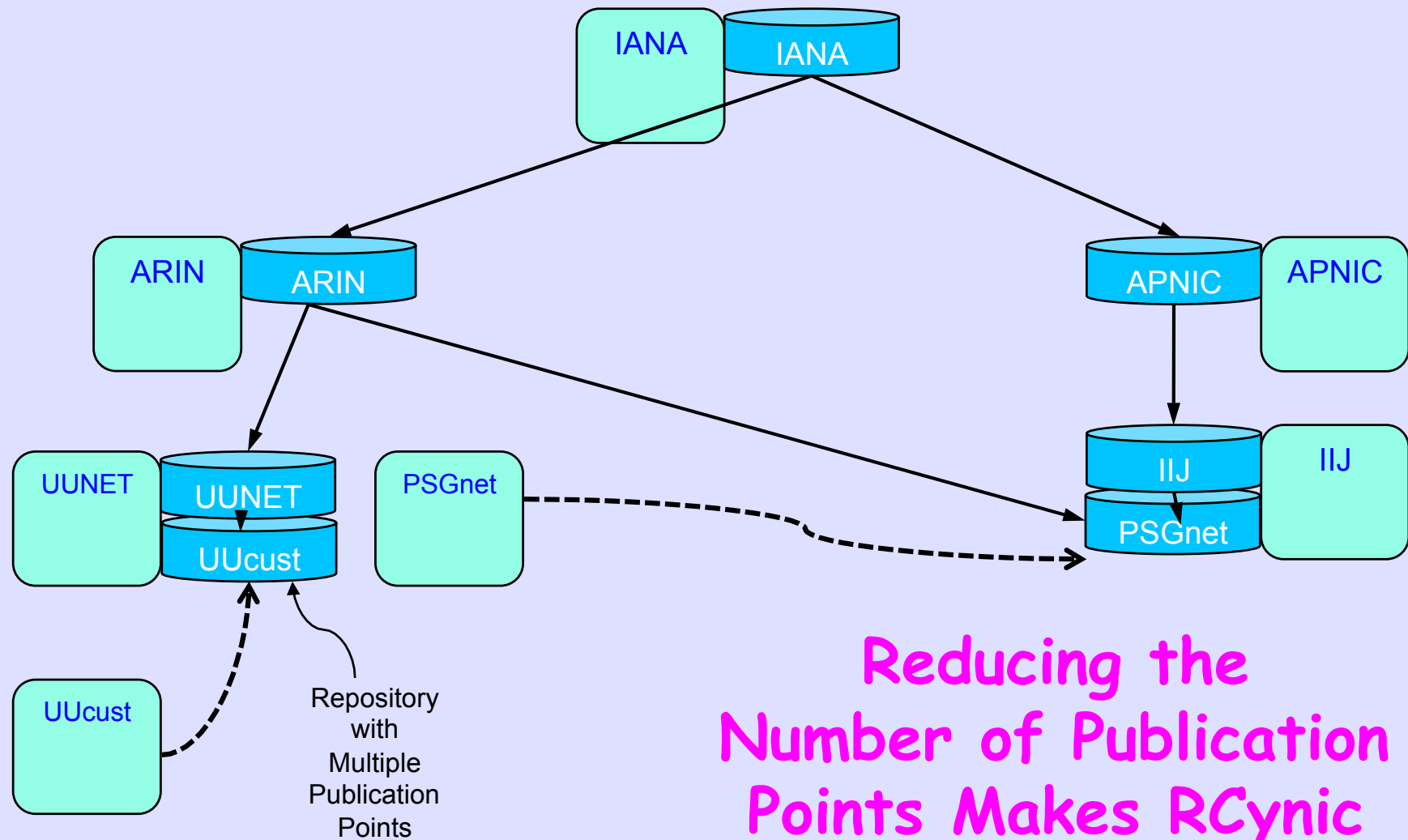


Reliability Issue

Expensive
To Fetch &
Unreliable



Reliability Via Hosted Publication



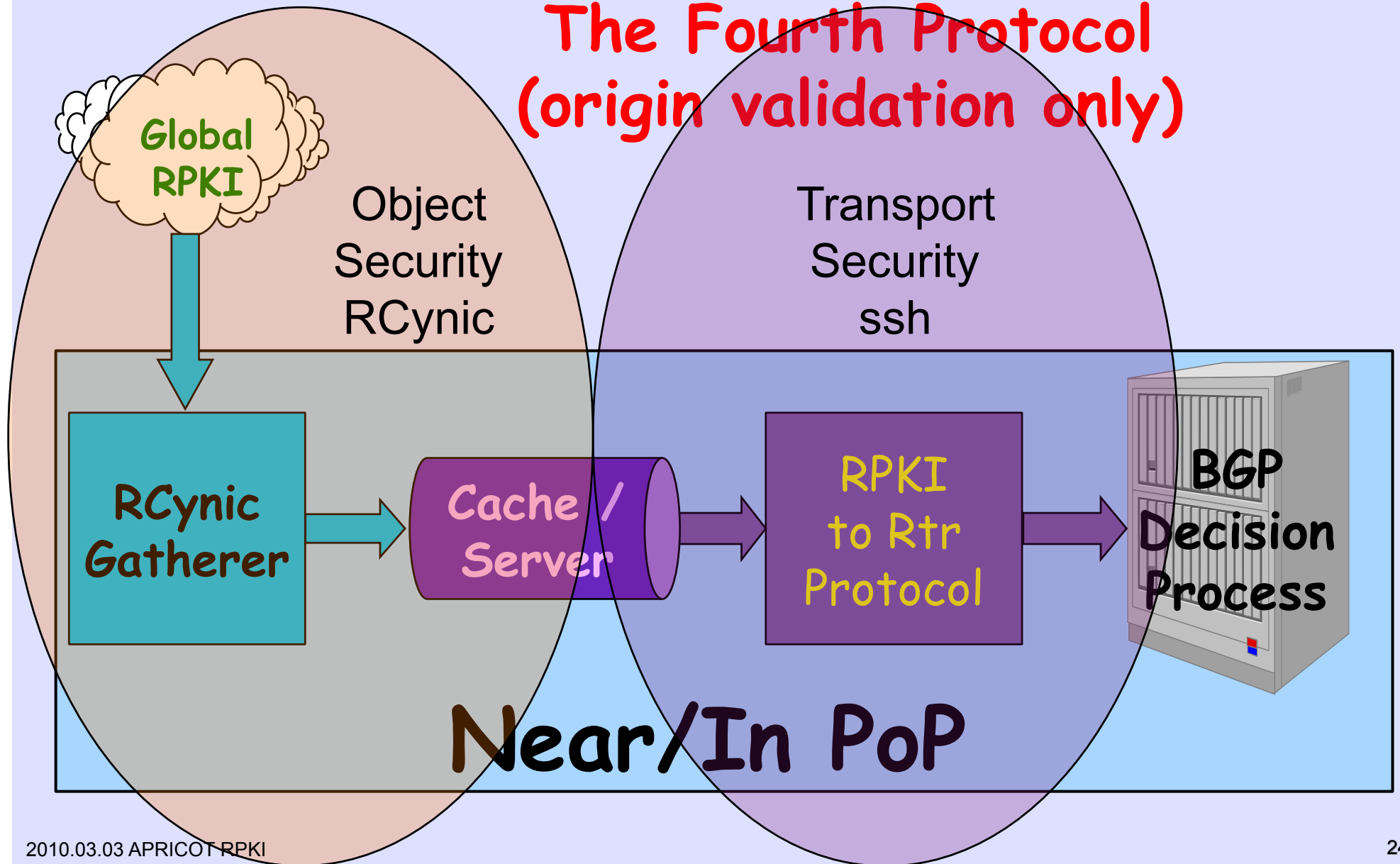
Reducing the Number of Publication Points Makes RSync Much More Efficient

Origin Only Validation

- Cisco IOS and IOS-XR test code have Origin Validation now
- Work continues daily in test routers
- Compute load much less than ACLs from IRR data, **10μsec per update!**
- Need to get all vendors on board

RPKI -> Router

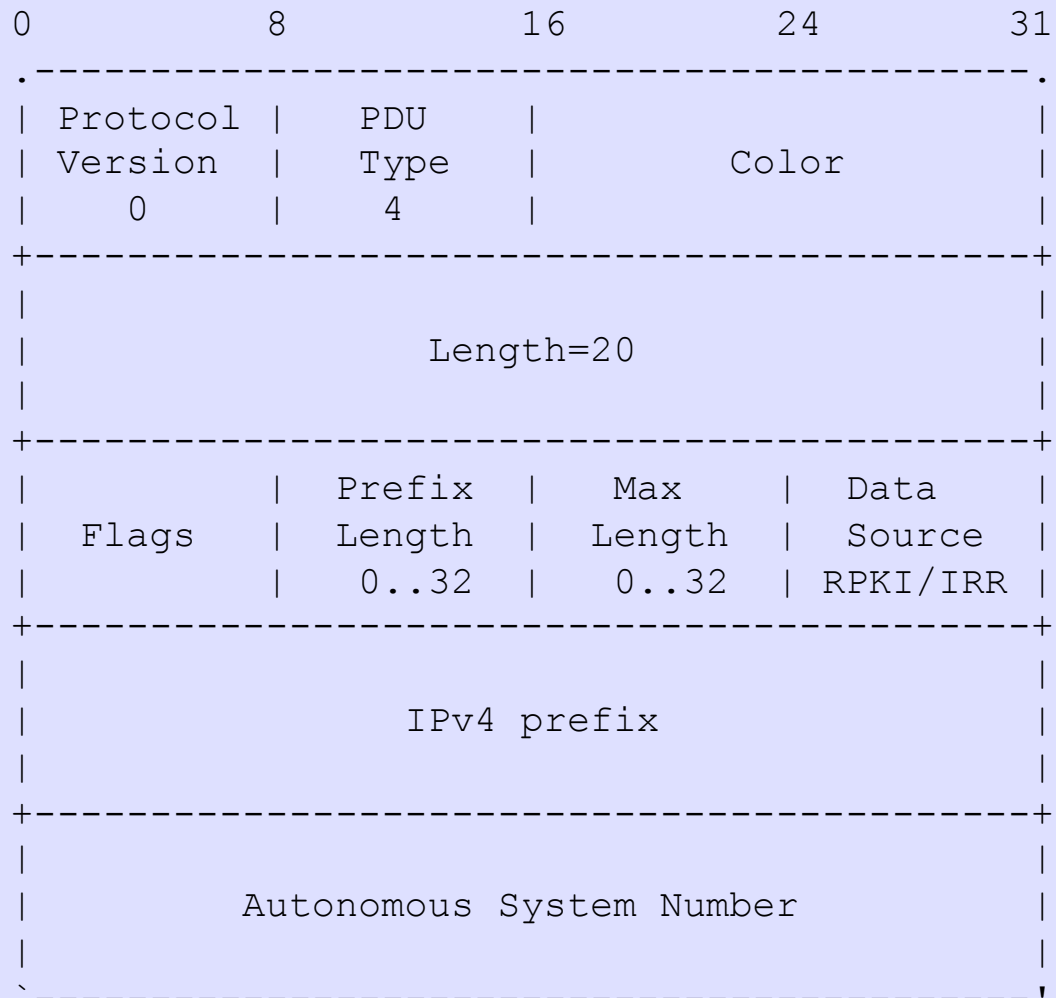
The Fourth Protocol
(origin validation only)



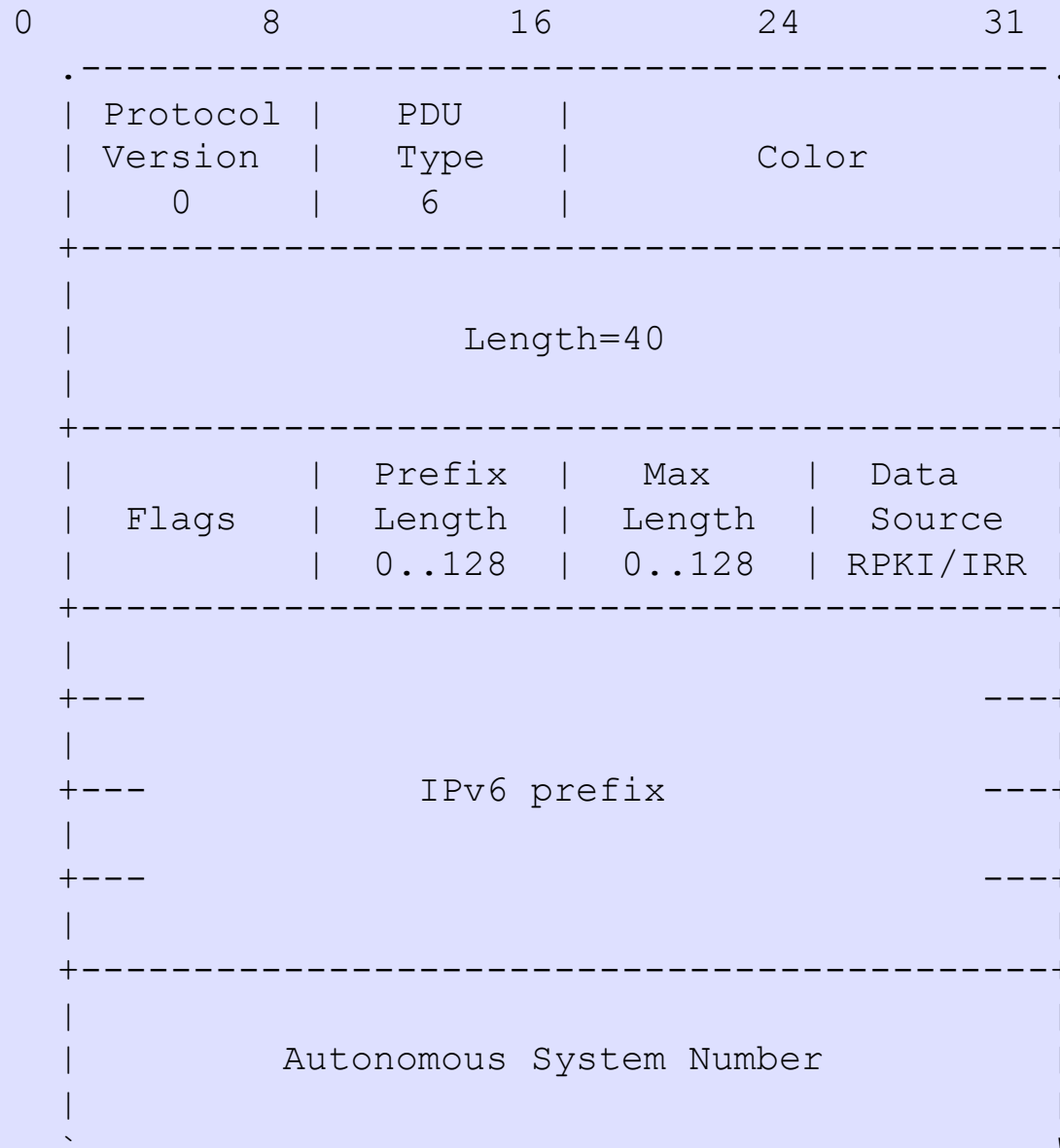
Typical Exchange

```
Cache                                     Router
| <----- Reset Query -----> | R requests data
|                                     |
| ----- Cache Response -----> | C confirms request
| ----- IPvX Prefix -----> | C sends zero or more
| ----- IPvX Prefix -----> | IPv4 and IPv6 Prefix
| ----- IPvX Prefix -----> | Payload PDUs
| ----- End of Data -----> | C sends End of Data
|                                     | and sends new serial
~                                     ~
| ----- Notify -----> | (optional)
|                                     |
| <----- Serial Query -----> | R requests data
|                                     |
| ----- Cache Response -----> | C confirms request
| ----- IPvX Prefix -----> | C sends zero or more
| ----- IPvX Prefix -----> | IPv4 and IPv6 Prefix
| ----- IPvX Prefix -----> | Payload PDUs
| ----- End of Data -----> | C sends End of Data
|                                     | and sends new serial
~                                     ~
```

IPv4 Prefix



IPv6 Prefix



Configure

```
router bgp 4128 bgp router-id 198.180.152.251
```

```
bgp rpki cache 198.180.150.1 42420 refresh-time 600
```

```
address-family ipv4 unicast
```

```
bgp dampening collect-statistics ebgp
```

```
redistribute static route-policy vb-ebgp-out
```

```
...
```

Result of Check

- Valid - A matching/covering ROA was found with a matching AS number
- Invalid - A matching or covering ROA was found, but AS number did not match, and there was no valid one
- Not Found - No matching or covering ROA was found

Prefix validation logic

```
1. query key = <BGP destination, masklen>, data = origin AS
2. result = BGP_PFXV_STATE_NOT_FOUND
3. walk prefix validation table to look for the query key
4. for each matched "entry" node in prefix validation table,
5.     prefix_exists = TRUE
6.     walk all records with different maxLength values
7.     for each "record" within range (query masklen <= maxLength)
8.         if query origin AS == record origin AS
9.             result = BGP_PFXV_STATE_VALID
10.            return (result)
11.        endif
12.    endfor
13. endfor
14. if prefix_exists == TRUE,
15.     result = BGP_PFXV_STATE_INVALID
16. endif
17. return (result)
```

Policy Overrides

- Disable prefix validation globally
- Disable prefix validation per EBGP peer
- Disable prefix validation for a set of prefixes

[When disabled, the "state" of such EBGP learnt routes will be set to "not-found"]

- Allow "invalid" routes for bestpath selection
- Disallow "not-found" routes for bestpath selection
- Set arbitrary communities based on "validity state" on neighbor outbound for debugging purposes

Show commands

```
RP/0/5/CPU0:ios#show bgp rpki prefix-validation database
```

```
Thu Jul 16 15:56:43.805 UTC
```

| Network | Maxlen | Origin-AS | Color | Source |
|---------------|--------|-----------|-------|--------|
| 8.0.0.0/4 | 6 | 200 | 0 | 0 |
| 1.1.0.0/16 | 24 | 1 | 0 | 0 |
| 3.0.0.0/24 | 24 | 2 | 0 | 0 |
| 4.0.0.0/8 | 8 | 3 | 0 | 0 |
| 4.0.0.0/24 | 24 | 3 | 0 | 0 |
| 5.0.0.0/24 | 24 | 4 | 0 | 0 |
| 10.0.0.0/6 | 8 | 100 | 0 | 0 |
| 8.0.0.0/8 | 24 | 36394 | 0 | 0 |
| 11.0.0.0/16 | 24 | 100 | 0 | 0 |
| 12.0.0.0/8 | 8 | 7018 | 0 | 0 |
| 20.137.0.0/21 | 21 | 4237 | 0 | 0 |

Good Dog!

```
RP/0/1/CPU0:r0.dfw#show bgp 192.158.248.0/24
```

```
BGP routing table entry for 192.158.248.0/24
```

```
Versions:
```

| Process | bRIB/RIB | SendTblVer |
|---------|----------|------------|
| Speaker | 132327 | 132327 |

```
Last Modified: Oct 2 01:06:47.630 for 13:33:12
```

```
Paths: (6 available, best #3)
```

```
Advertised to peers (in unique update groups):
```

```
204.69.200.26
```

```
Path #1: Received by speaker 0
```

```
2914 1299 6939 6939 27318
```

```
157.238.224.149 from 157.238.224.149 (129.250.0.85)
```

```
Origin IGP, metric 0, localpref 100, valid, external, \
```

```
origin validity state: valid
```

```
Community: 2914:420 2914:2000 2914:3000 4128:380
```

```
Path #2: Received by speaker 0
```

```
...
```

Bad Dog!

```
RP/0/1/CPU0:r0.dfw#sh bgp 64.9.224.0
```

```
BGP routing table entry for 64.9.224.0/20
```

```
Versions:
```

| Process | bRIB/RIB | SendTblVer |
|---------|----------|------------|
| Speaker | 0 | 0 |

```
Last Modified: Oct 2 17:38:27.630 for 4d22h
```

```
Paths: (6 available, no best path)
```

```
Not advertised to any peer
```

```
Path #1: Received by speaker 0
```

```
2914 3356 36492
```

```
157.238.224.149 from 157.238.224.149 (129.250.0.85)
```

```
Origin IGP, metric 2, localpref 100, valid, external,  
origin validity state: invalid
```

```
Community: 2914:420 2914:2000 2914:3000 4128:380
```

Strange Dog!

```
RP/0/1/CPU0:r0.dfw#sh bgp 147.28.0.0
```

```
BGP routing table entry for 147.28.0.0/16
```

```
Versions:
```

| Process | bRIB/RIB | SendTblVer |
|---------|----------|------------|
| Speaker | 337691 | 337691 |

```
Last Modified: Oct 2 17:40:16.630 for 4d22h
```

```
Paths: (6 available, best #1)
```

```
Advertised to peers (in unique update groups):
```

```
204.69.200.26
```

```
Path #1: Received by speaker 0
```

```
2914 3130
```

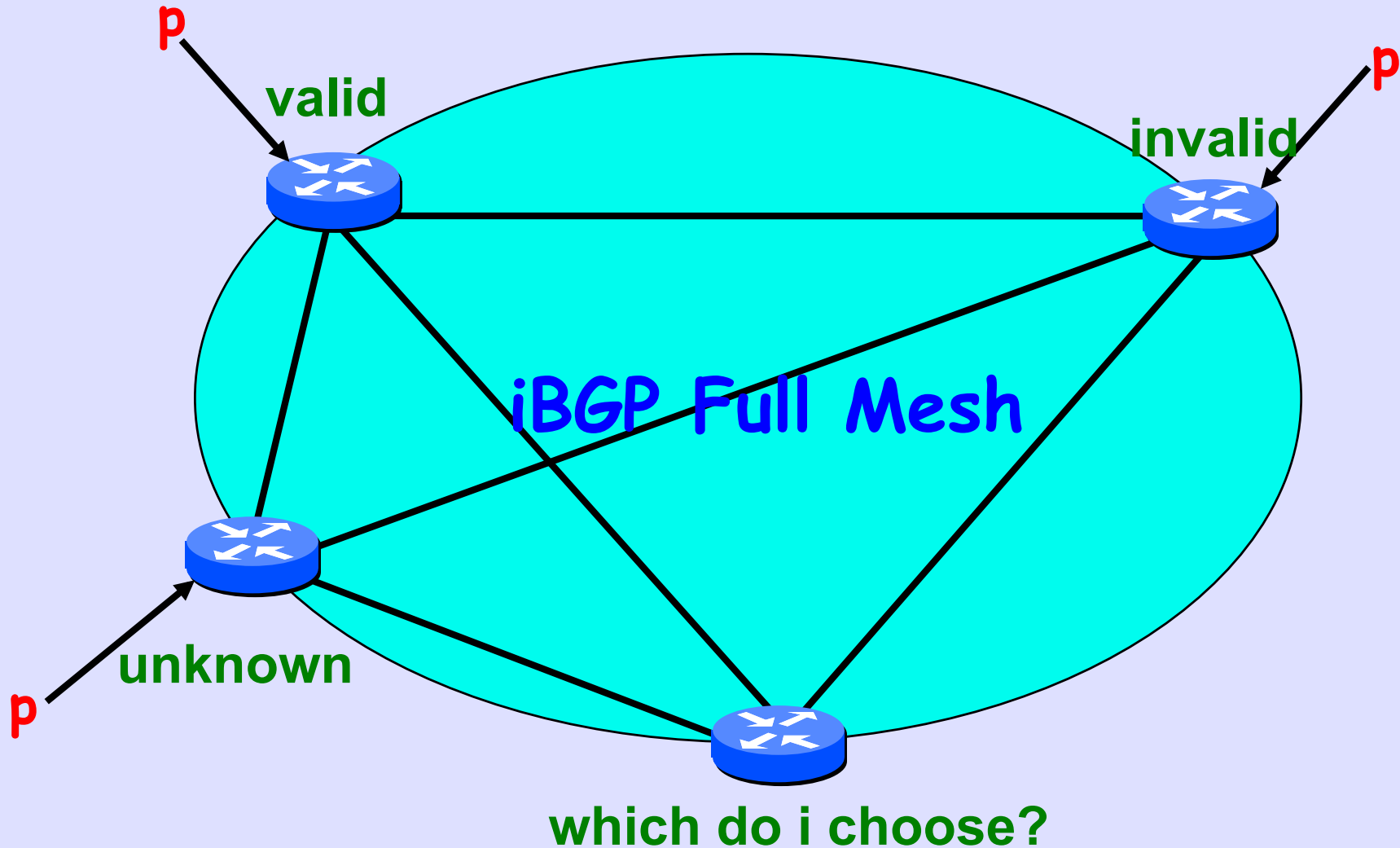
```
157.238.224.149 from 157.238.224.149 (129.250.0.85)
```

```
Origin IGP, metric 68, localpref 100, valid, external, \
```

```
origin validity state: not found
```

```
Community: 2914:410 2914:2000 2914:3000 4128:380
```

iBGP Hides Validity State



Unknown Beat Valid!

```
r1.iad#sh ip bg 198.180.152.0
```

```
BGP routing table entry for 198.180.152.0/24, version 324176
```

```
Paths: (2 available, best #1, table default)
```

```
Not advertised to any peer
```

```
2914 4128
```

```
129.250.10.157 (metric 1) from 198.180.150.253  
(198.180.150.253)
```

```
Origin IGP, metric 51, localpref 100, valid, internal, best
```

```
Community: 2914:410 2914:2000 2914:3000 3927:380
```

```
1239 2914 4128
```

```
144.232.18.81 from 144.232.18.81 (144.228.241.254)
```

```
Origin IGP, metric 0, localpref 100, valid, external
```

```
Community: 3927:380
```

```
Sovc state valid
```

MED Beat Valid

```
r1.iad#sh ip bg 147.28.0.0
```

```
BGP routing table entry for 147.28.0.0/16, version 142233
```

```
Paths: (2 available, best #1, table default)
```

```
Not advertised to any peer
```

```
2914 3130
```

```
129.250.10.157 (metric 1) from 198.180.150.253  
(198.180.150.253)
```

```
Origin IGP, metric 105, localpref 100, valid, internal, best
```

```
Community: 2914:410 2914:2000 2914:3000 3927:380
```

```
1239 3130
```

```
144.232.18.81 from 144.232.18.81 (144.228.241.254)
```

```
Origin IGP, metric 653, localpref 100, valid, external
```

```
Community: 3927:380
```

```
Sovc state valid
```

Highly Secure Use

```
route-map validity-0  
  match rpki-valid  
  set localpref 100  
route-map validity-1  
  set localpref 50
```

After AS-Path

```
route-map validity-0
```

```
  match rpki-valid
```

```
  set metric 100
```

```
route-map validity-1
```

```
  match rpki-unknown
```

```
  set metric 50
```

```
route-map validity-2
```

```
  set metric 25
```

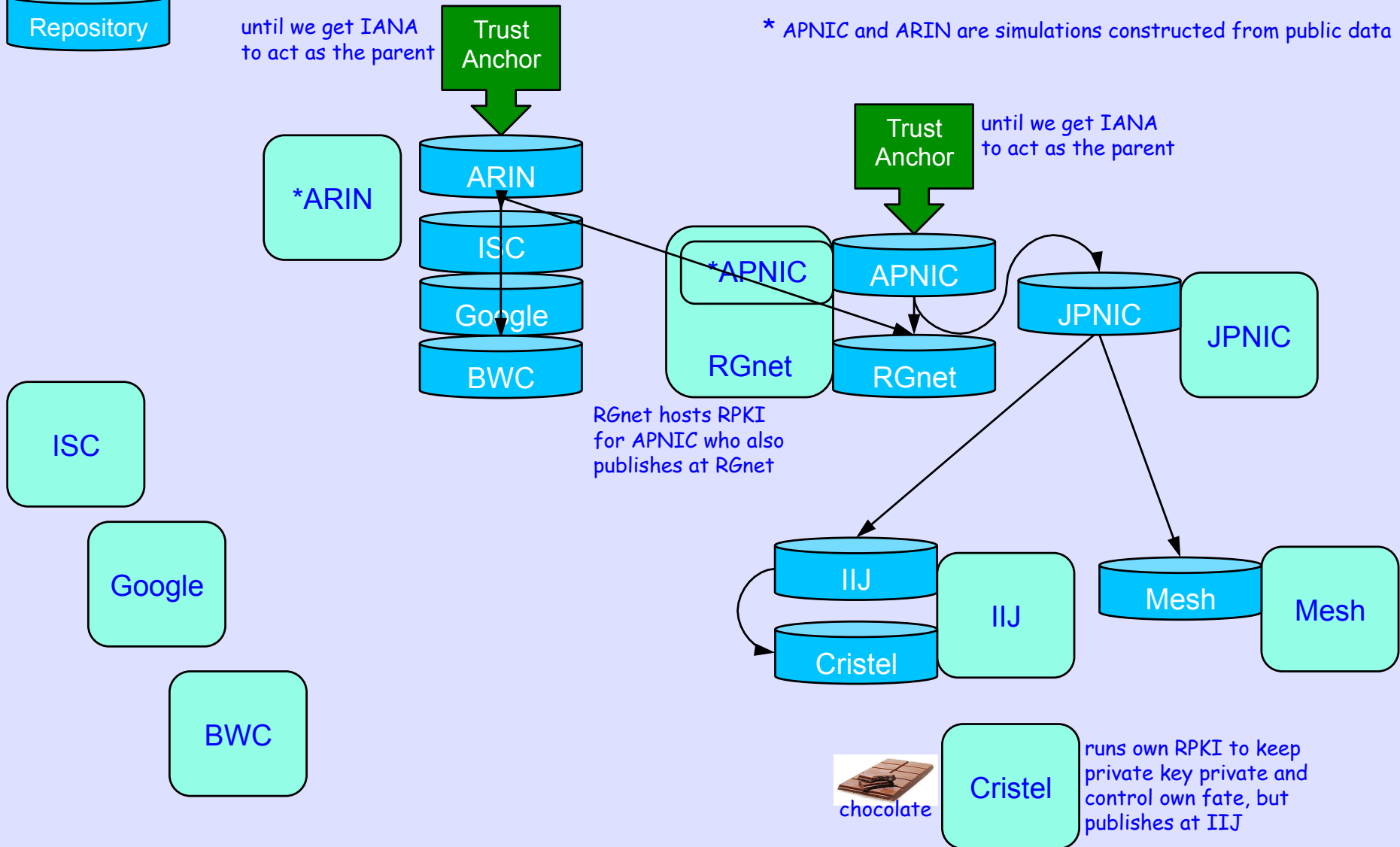

The Open TestBed

Running Code

Repository

until we get IANA to act as the parent

* APNIC and ARIN are simulations constructed from public data



RPKI is Open Source

<https://subvert-rpki.hactrn.net/>

Work Supported By

- **US Government**

SUPPORTED BY, OR IN PART BY, THE U. S. ARMY RESEARCH LABORATORY AND THE U. S. ARMY RESEARCH OFFICE UNDER CONTRACT/GRANT NUMBER W911NF-05-C-0113, THROUGH FUNDING PROVIDED BY DEPARTMENT OF HOMELAND SECURITY DIRECTORATE FOR SCIENCE AND TECHNOLOGY

- **ARIN**

- **Cisco**

SUPPORTED BY, OR IN PART BY A CISCO UNIVERSITY RESEARCH PROJECT GIFT VIA KEIO UNIVERSITY

- **Internet Initiative Japan**

- **Google, NTT, Equinix**