The RPKI & Origin Validation

APRICOT / Hong Kong

2011.02.19

Randy Bush <randy@psg.com>
Rob Austein <sra@hactrn.net>
Michael Elkins <me@sigpipe.org>
Steve Bellovin <smb@cs.columbia.edu>

And a cast of thousands! Well, dozens :)

Routing Mistakes

- Routing errors are significant and have very high customer impact
- We need to fix this before we are crucified in the Wall Street Journal
- 99% of mis-announcements are accidental originations of someone else's prefix -- YouTube Incident

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

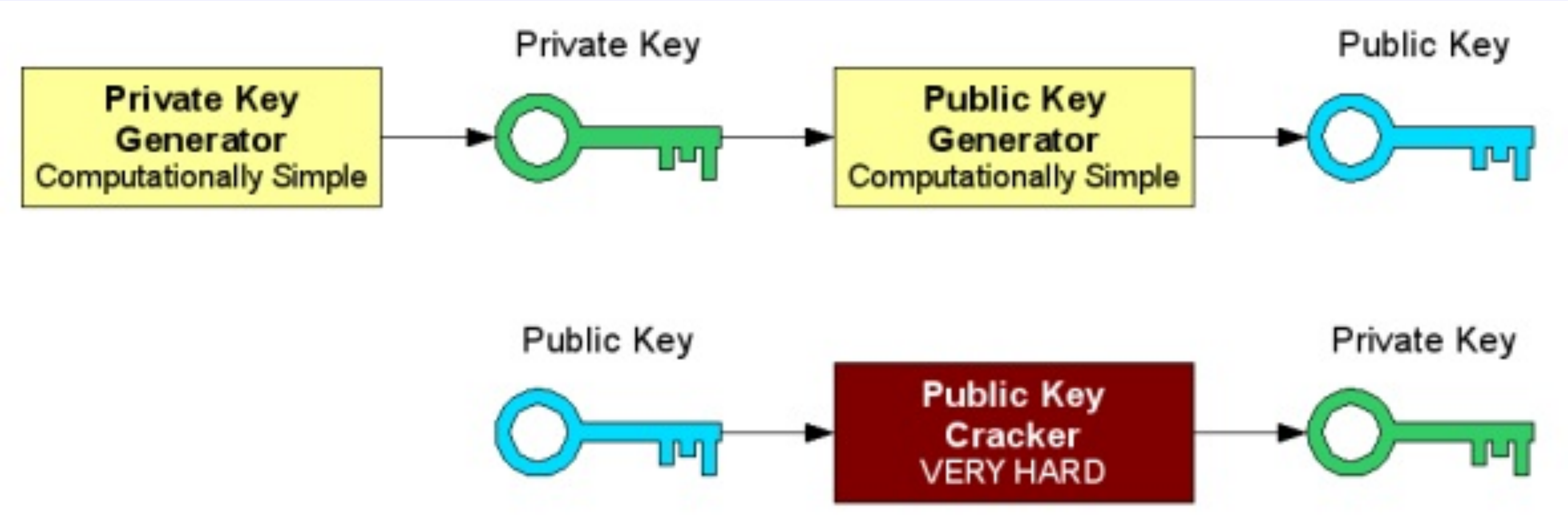
This is Not New

- 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
- 2009 – ISOC discovers problem

First We Need to Formally Know What Prefixes Belong to Whom

Resource Public Key Infrastructure (RPKI)

Key Generation

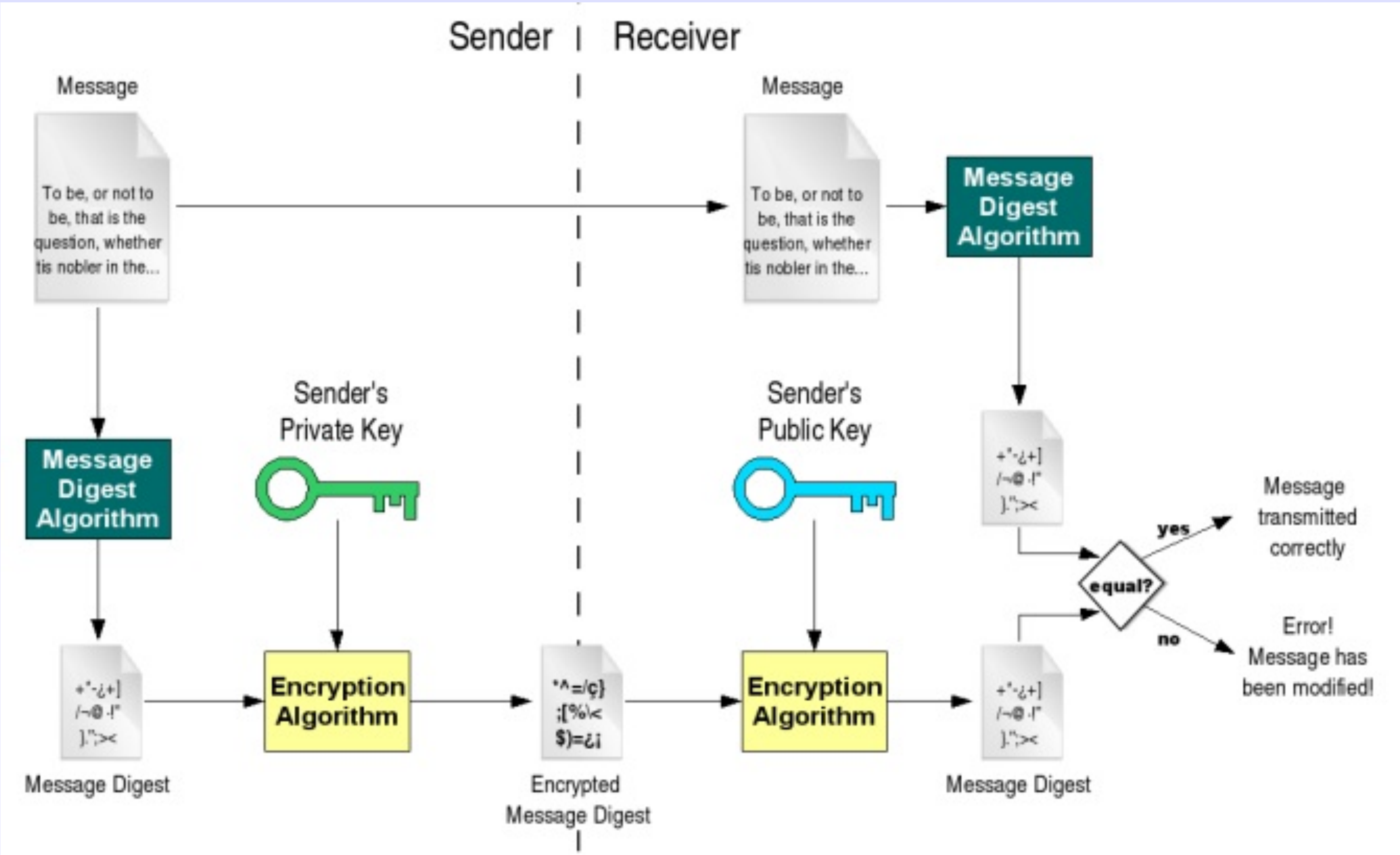

Private Key
Public Key
Private Key Generator Computationally Simple
Public Key Generator Computationally Simple
Public Key
Private Key
Public Key Cracker VERY HARD

From- http://gdp.globus.org/gt4-tutorial/multiplehtml/ch09s03.html

Public-Key Cryptography

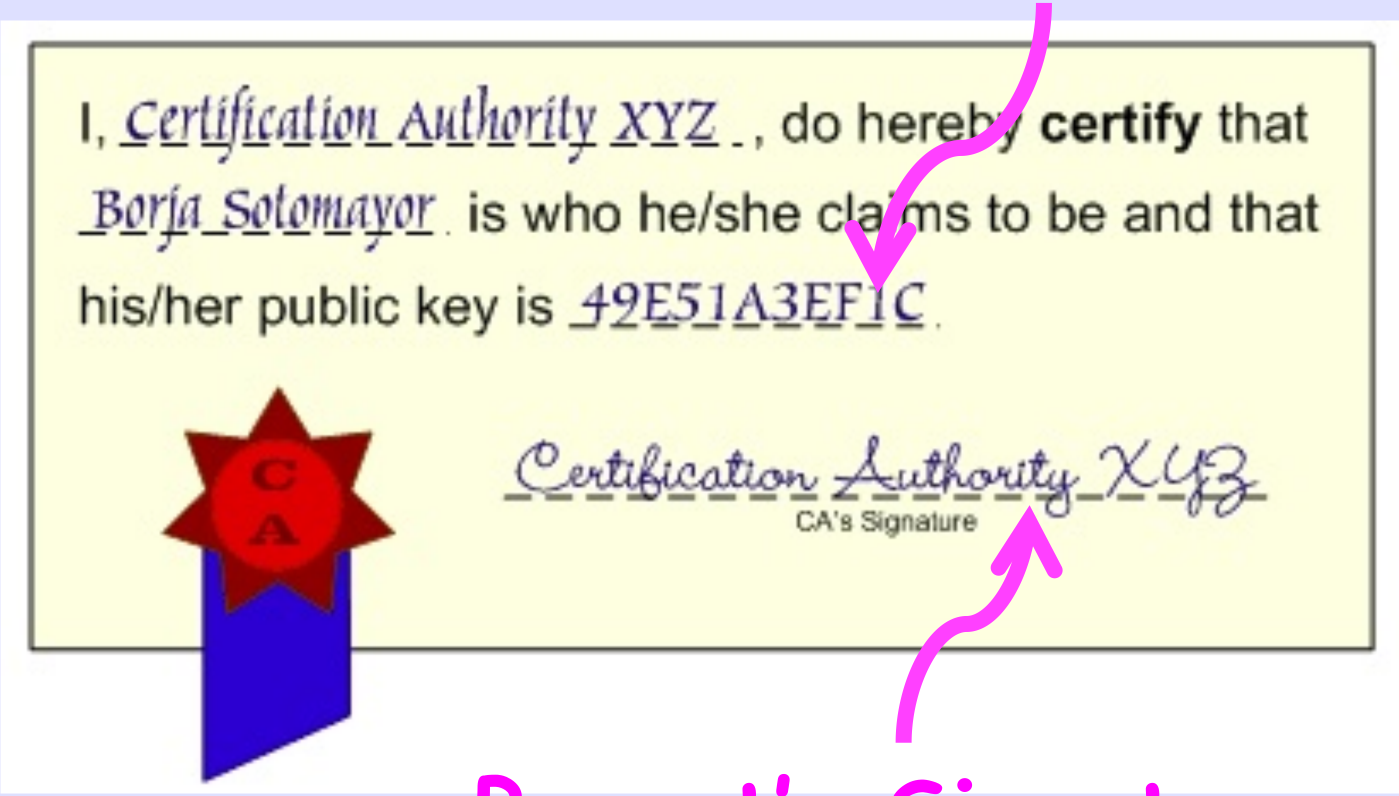
- **Private key**: Must be known *only* by its owner.
- **Public key**: Known to everyone (it is *public*)
- **Relation between two keys**: What one key encrypts, the other decrypts, and vice versa. If you encrypt something with my public key (which you know, because it's public :-), I would need my private key to decrypt it.

Digital Signature



Certificates

Over My Public Key

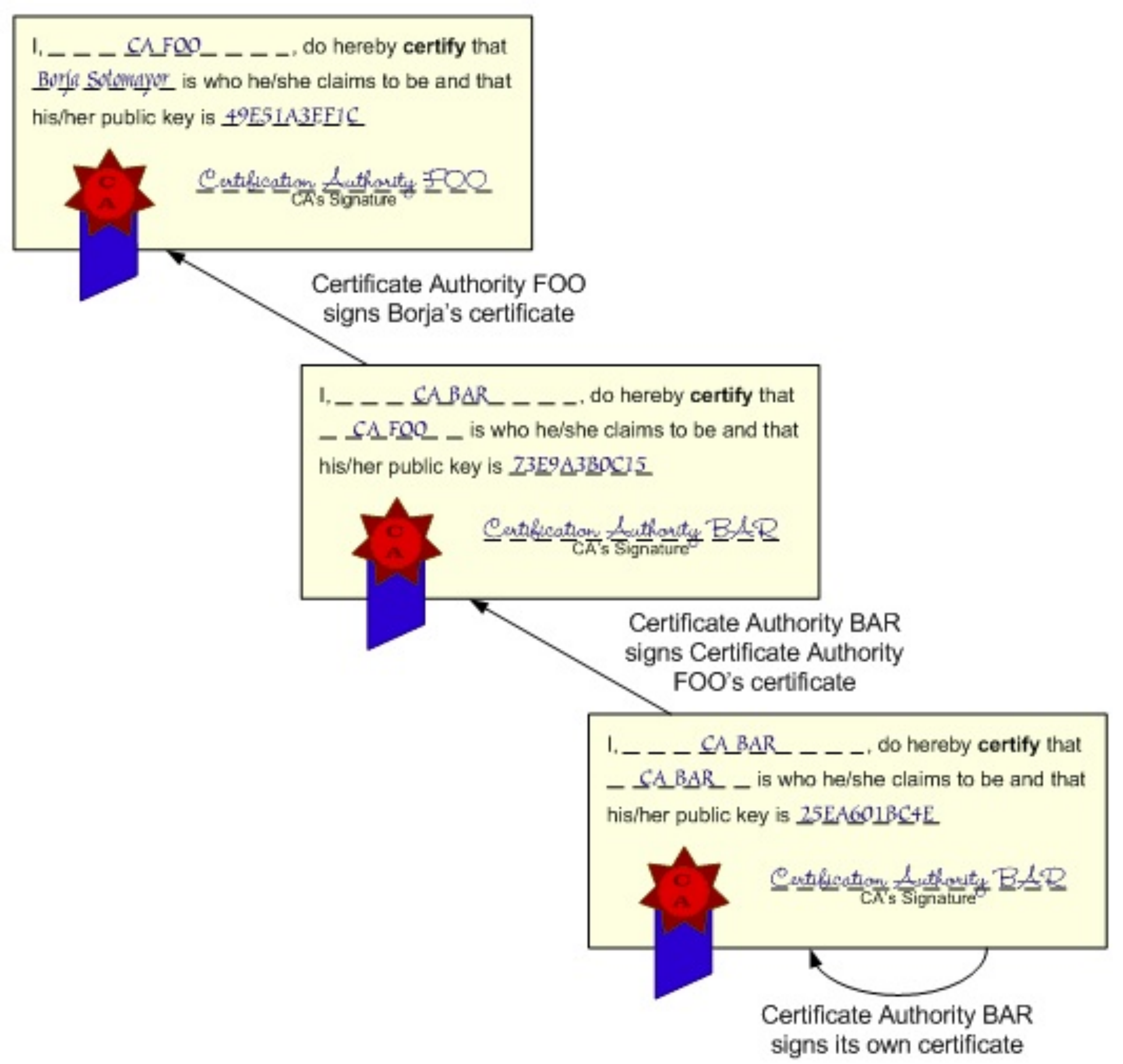


I, Certification Authority XYZ, do hereby **certify** that Borja Sotomayor is who he/she claims to be and that his/her public key is 49E51A3EF1C.

Certification Authority XYZ
CA's Signature

Parent's Signature

Certificate Hierarchy



I, _ _ _ _ CA FOO _ _ _ _, do hereby **certify** that Borja Sotomayor is who he/she claims to be and that his/her public key is 49E51A3EF1C

Certification Authority FOO
CA's Signature

Certificate Authority FOO signs Borja's certificate

I, _ _ _ _ CA BAR _ _ _ _, do hereby **certify** that _ CA FOO _ is who he/she claims to be and that his/her public key is 73E9A3B0C15

Certification Authority BAR
CA's Signature

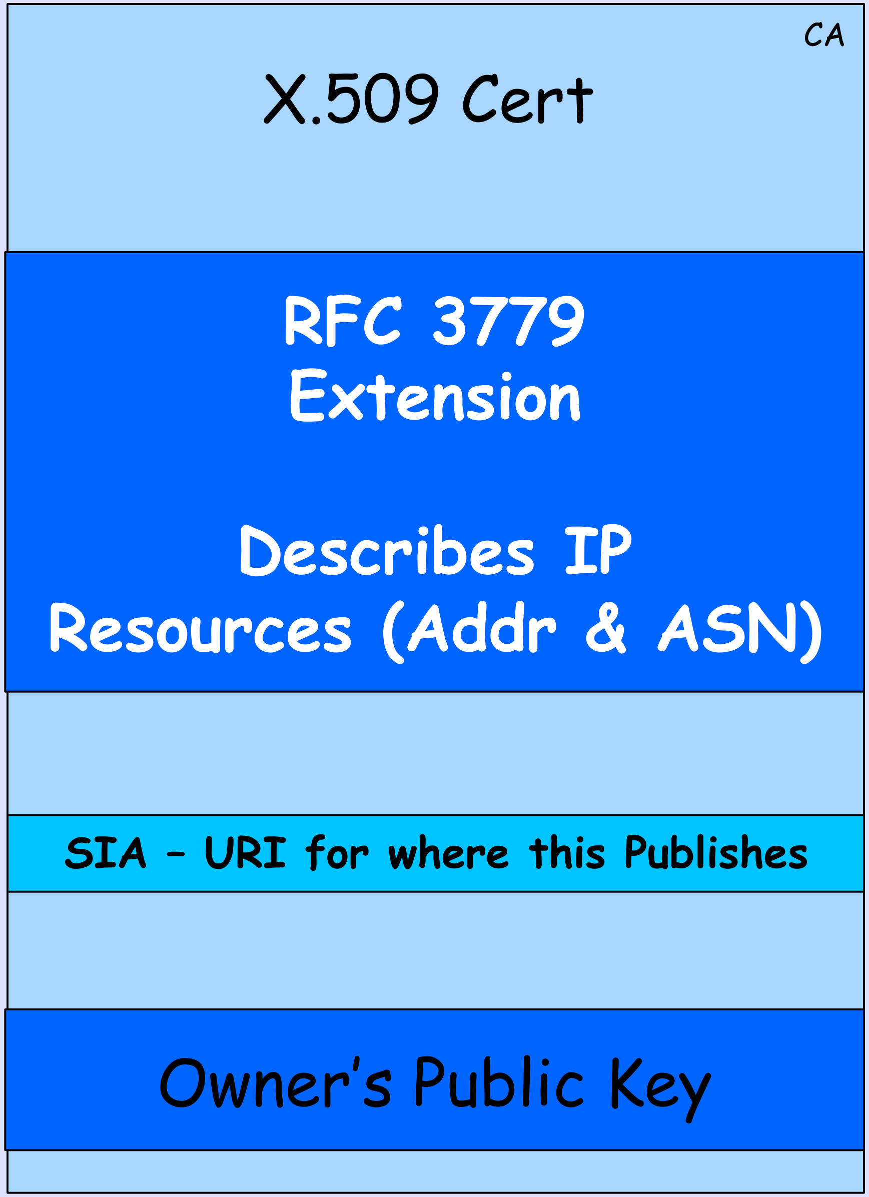
Certificate Authority BAR signs Certificate Authority FOO's certificate

I, _ _ _ _ CA BAR _ _ _ _, do hereby **certify** that _ CA BAR _ is who he/she claims to be and that his/her public key is 25EA601BC4E

Certification Authority BAR
CA's Signature

Certificate Authority BAR signs its own certificate

X.509 Certificate w/ 3779 Ext



CA

X.509 Cert

RFC 3779
Extension

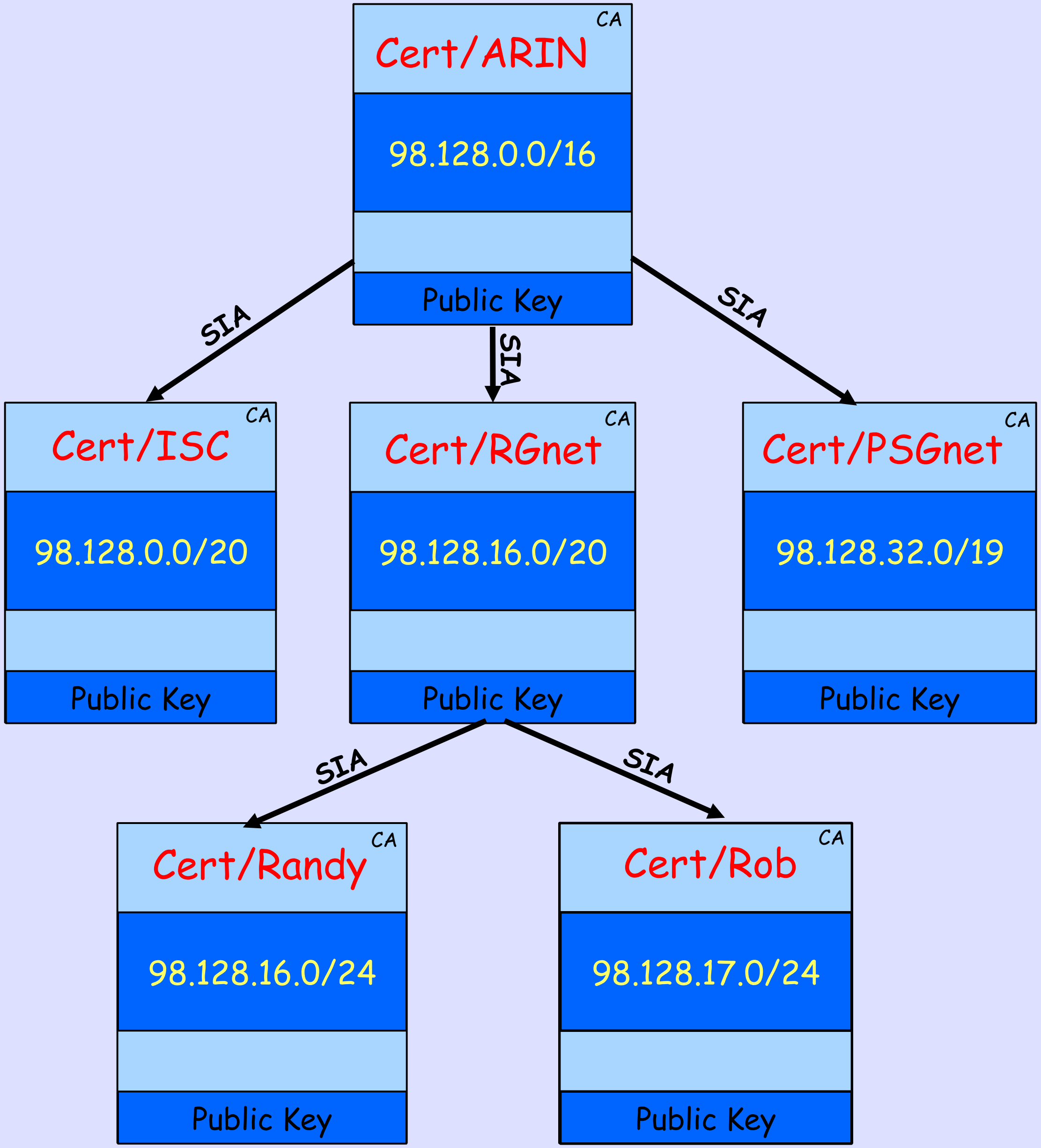
Describes IP
Resources (Addr & ASN)

SIA – URI for where this Publishes

Owner's Public Key

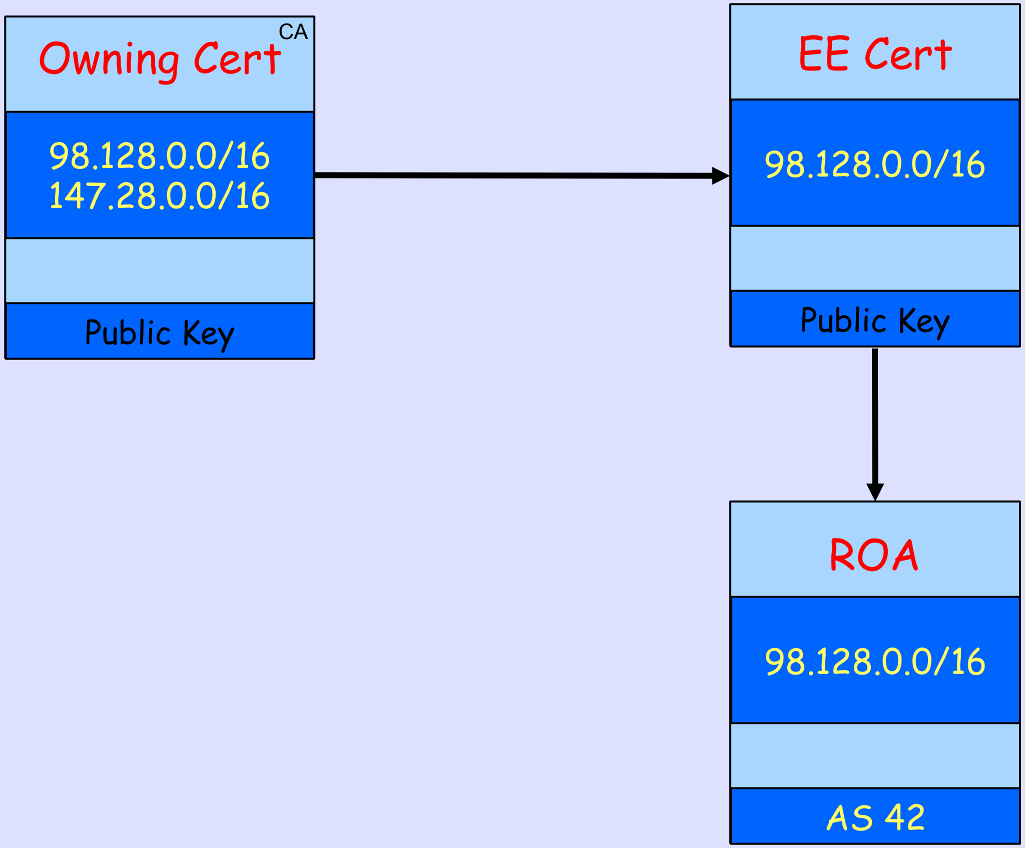
Being
Developed & Deployed
by
RIRs and Operators

Certificate Hierarchy follows Allocation Hierarchy



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

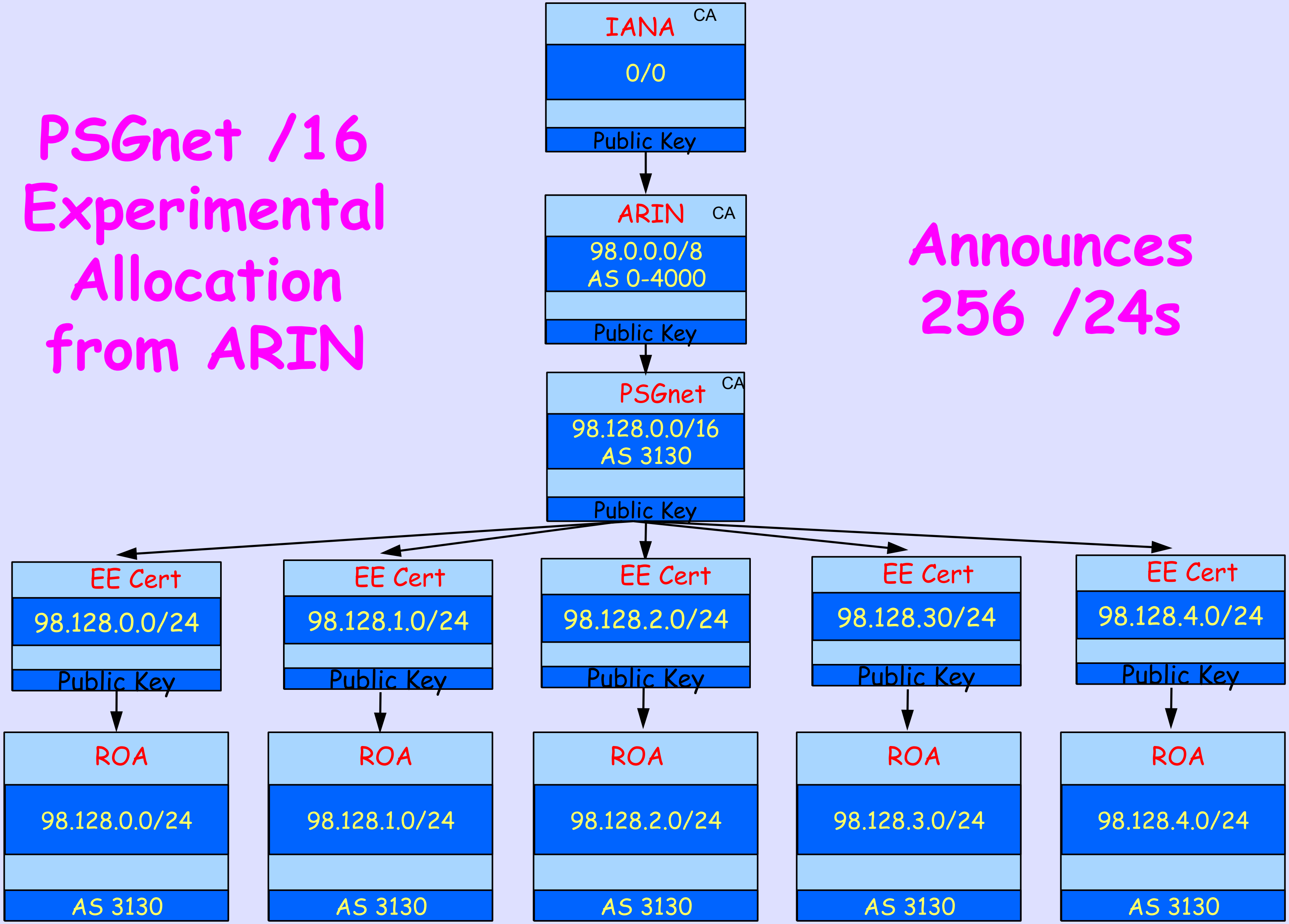
Route Origin Authorization (ROA)



End Entity Cert
can not sign certs.
can sign other things
e.g. ROAs

This is not a Cert
It is a signed blob

PSGnet /16 Experimental Allocation from ARIN



IANA CA
0/0
Public Key

ARIN CA
98.0.0.0/8
AS 0-4000
Public Key

PSGnet CA
98.128.0.0/16
AS 3130
Public Key

EE Cert
98.128.0.0/24
Public Key

EE Cert
98.128.1.0/24
Public Key

EE Cert
98.128.2.0/24
Public Key

EE Cert
98.128.30/24
Public Key

EE Cert
98.128.4.0/24
Public Key

ROA
98.128.0.0/24
AS 3130

ROA
98.128.1.0/24
AS 3130

ROA
98.128.2.0/24
AS 3130

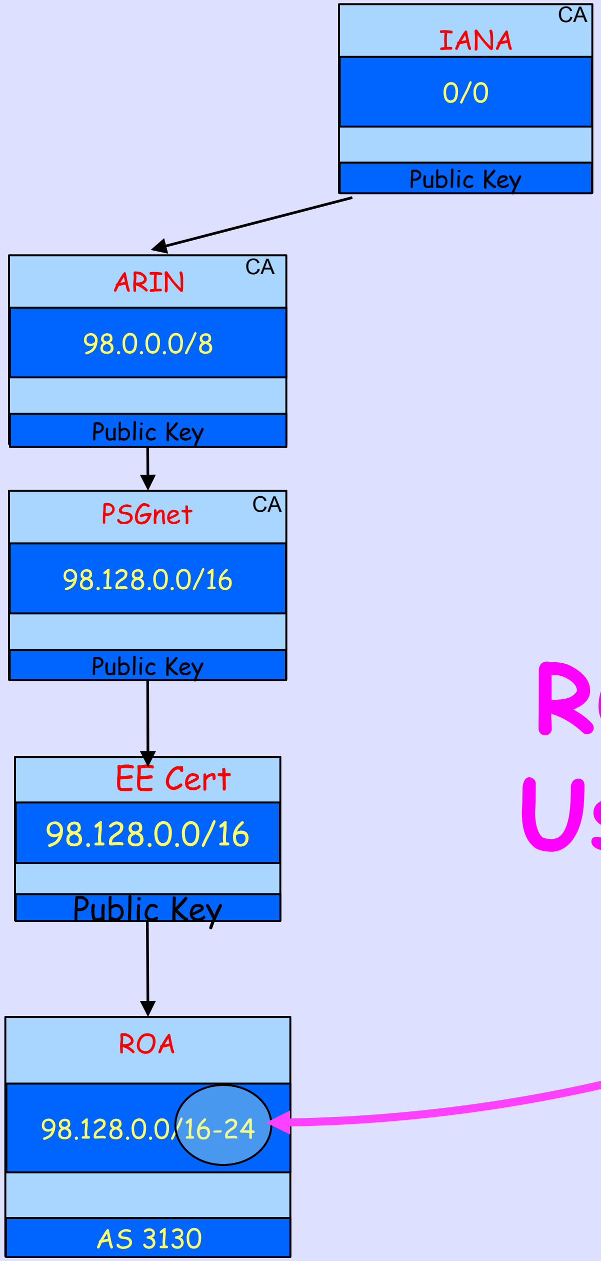
ROA
98.128.3.0/24
AS 3130

ROA
98.128.4.0/24
AS 3130

Announces 256 /24s

Too Many EE Certs and ROAs, Yucchhy!

ROA Aggregation Using Max Length



CA

IANA

0/0

Public Key

CA

ARIN

98.0.0.0/8

Public Key

CA

PSGnet

98.128.0.0/16

Public Key

EE Cert

98.128.0.0/16

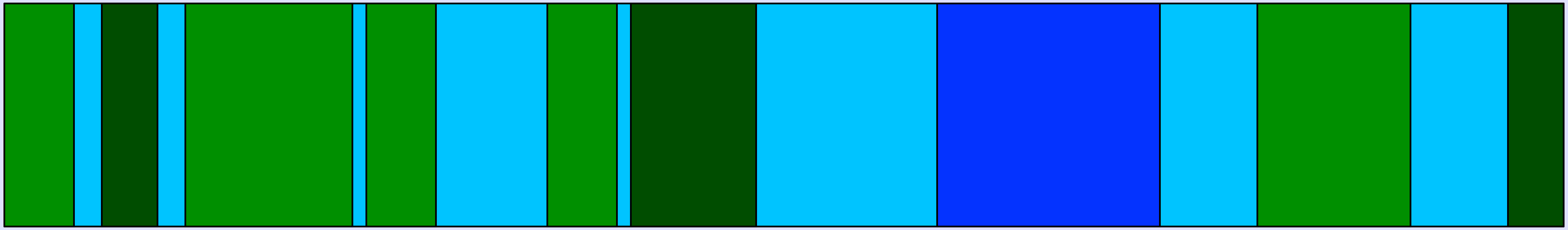
Public Key

ROA

98.128.0.0/16-24

AS 3130

Allocation in Reality



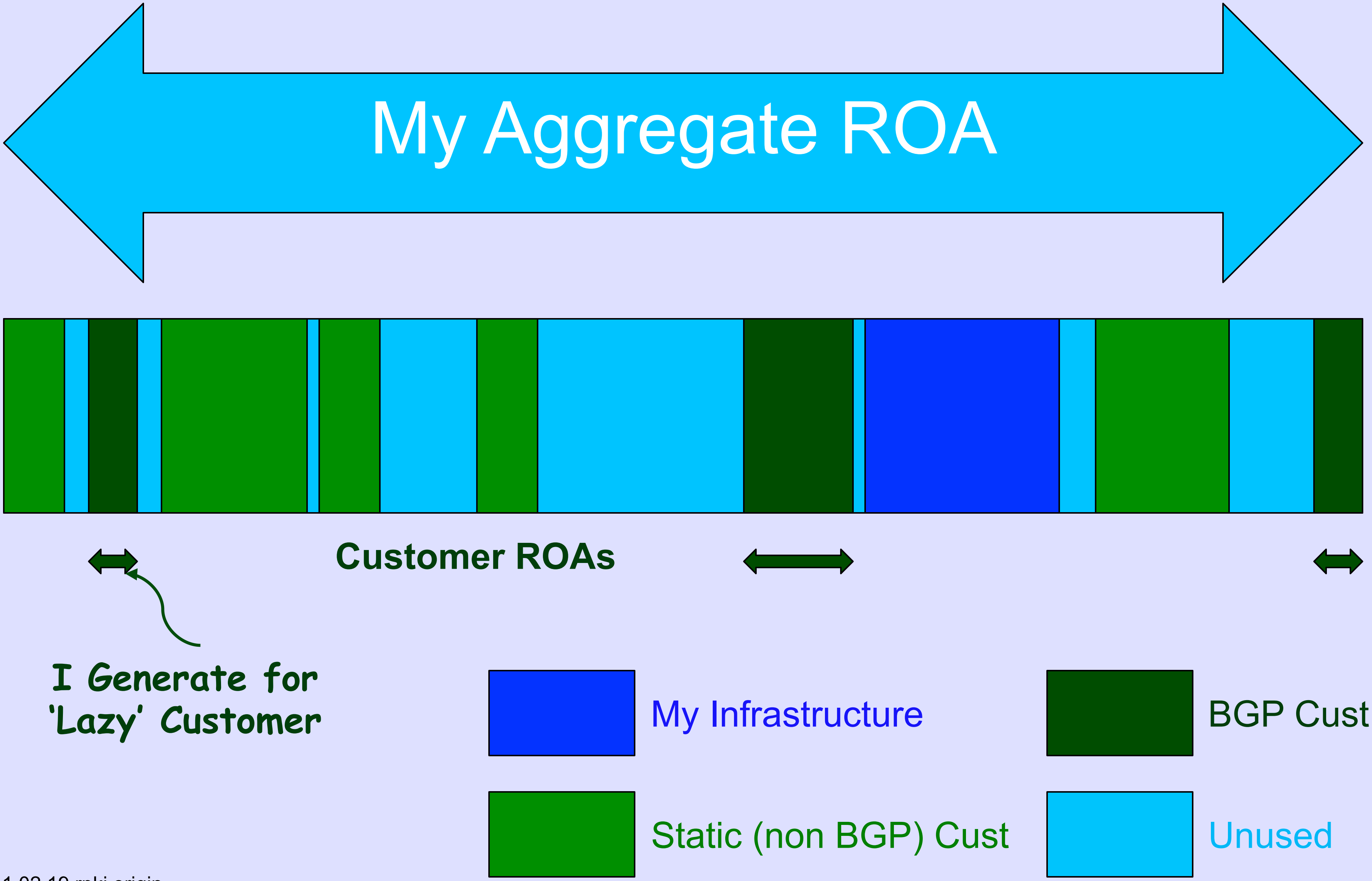
🟦 My Infrastructure

🟩 BGP Cust

🟩 Static (non BGP) Cust

🟦 Unused

ROA Use



My Aggregate ROA

Customer ROAs

I Generate for 'Lazy' Customer

🟦 My Infrastructure

🟩 BGP Cust

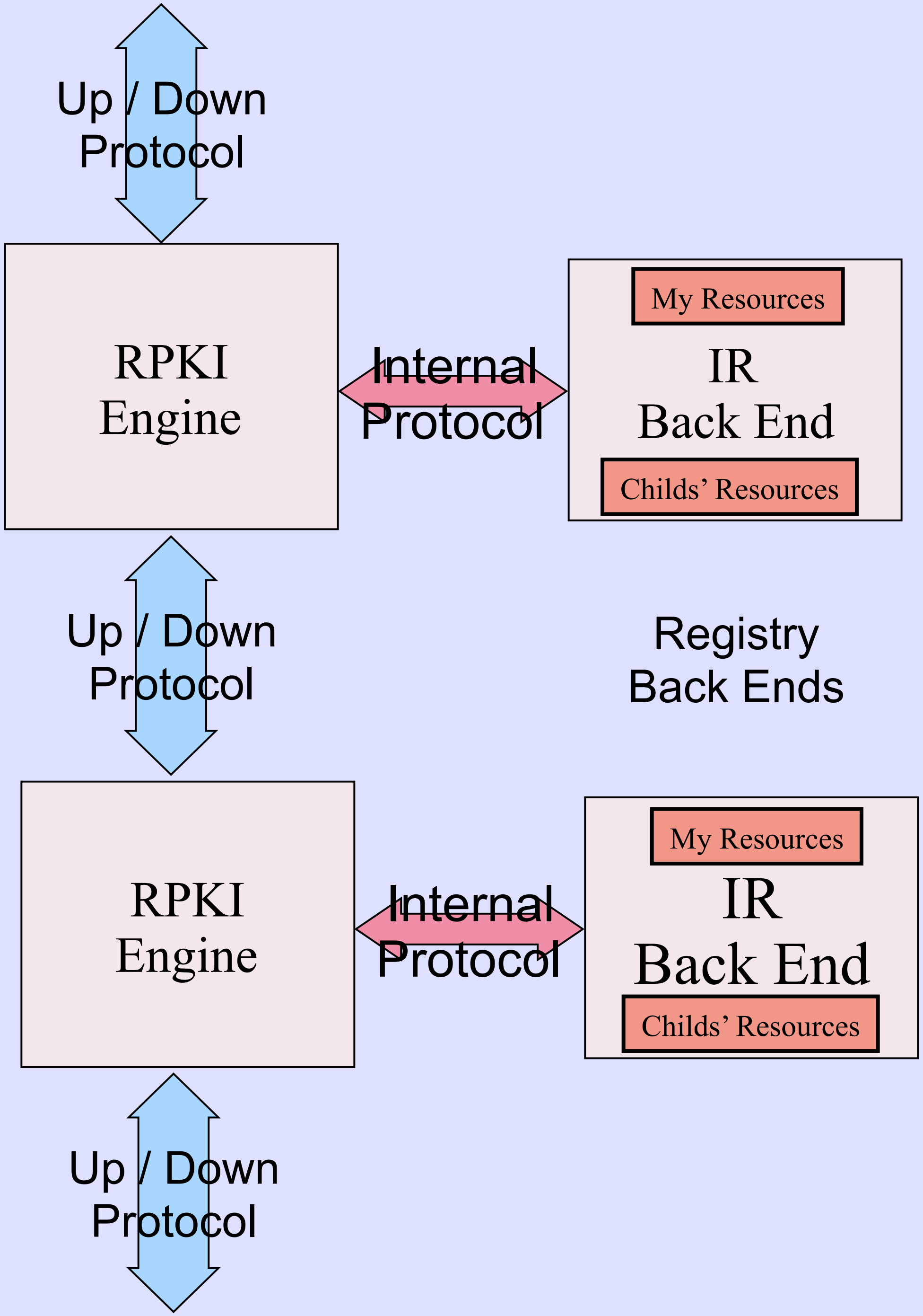
🟩 Static (non BGP) Cust

🟦 Unused

Running Code

And the Three RPKI Protocols

Simple Parent and Simple Child



Up / Down Protocol

RPKI Engine

Internal Protocol

My Resources

IR Back End

Childs' Resources

Up / Down Protocol

Registry Back Ends

RPKI Engine

Internal Protocol

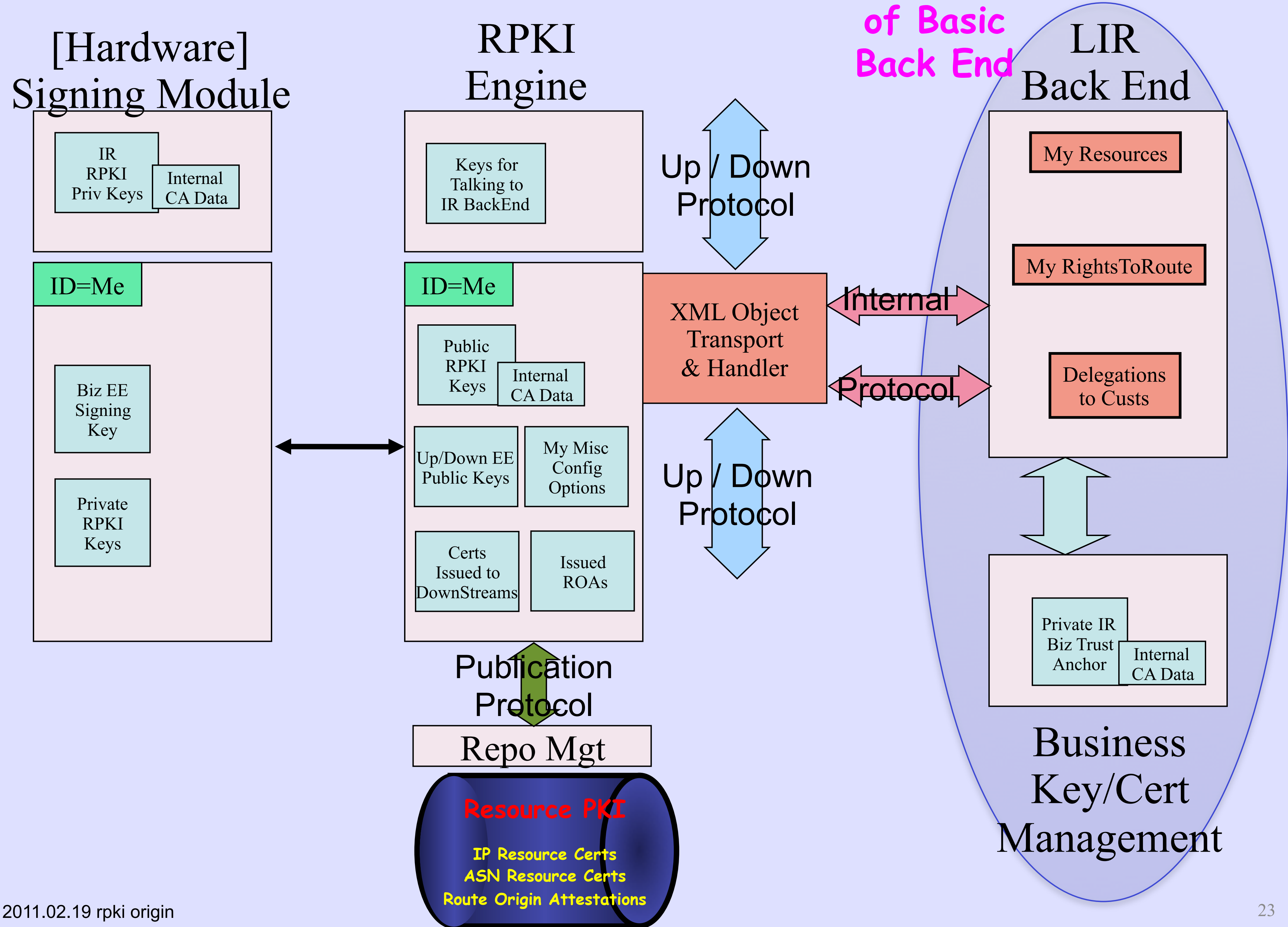
My Resources

IR Back End

Childs' Resources

Up / Down Protocol

Prototype of Basic Back End



[Hardware] Signing Module

IR RPKI Priv Keys

Internal CA Data

ID=Me

Biz EE Signing Key

Private RPKI Keys

RPKI Engine

Keys for Talking to IR BackEnd

ID=Me

Public RPKI Keys

Internal CA Data

Up/Down EE Public Keys

My Misc Config Options

Certs Issued to DownStreams

Issued ROAs

Up / Down Protocol

XML Object Transport & Handler

Up / Down Protocol

Internal Protocol

Publication Protocol

Repo Mgt

Resource PKI

IP Resource Certs

ASN Resource Certs

Route Origin Attestations

LIR Back End

My Resources

My RightsToRoute

Delegations to Custs

Private IR Biz Trust Anchor

Internal CA Data

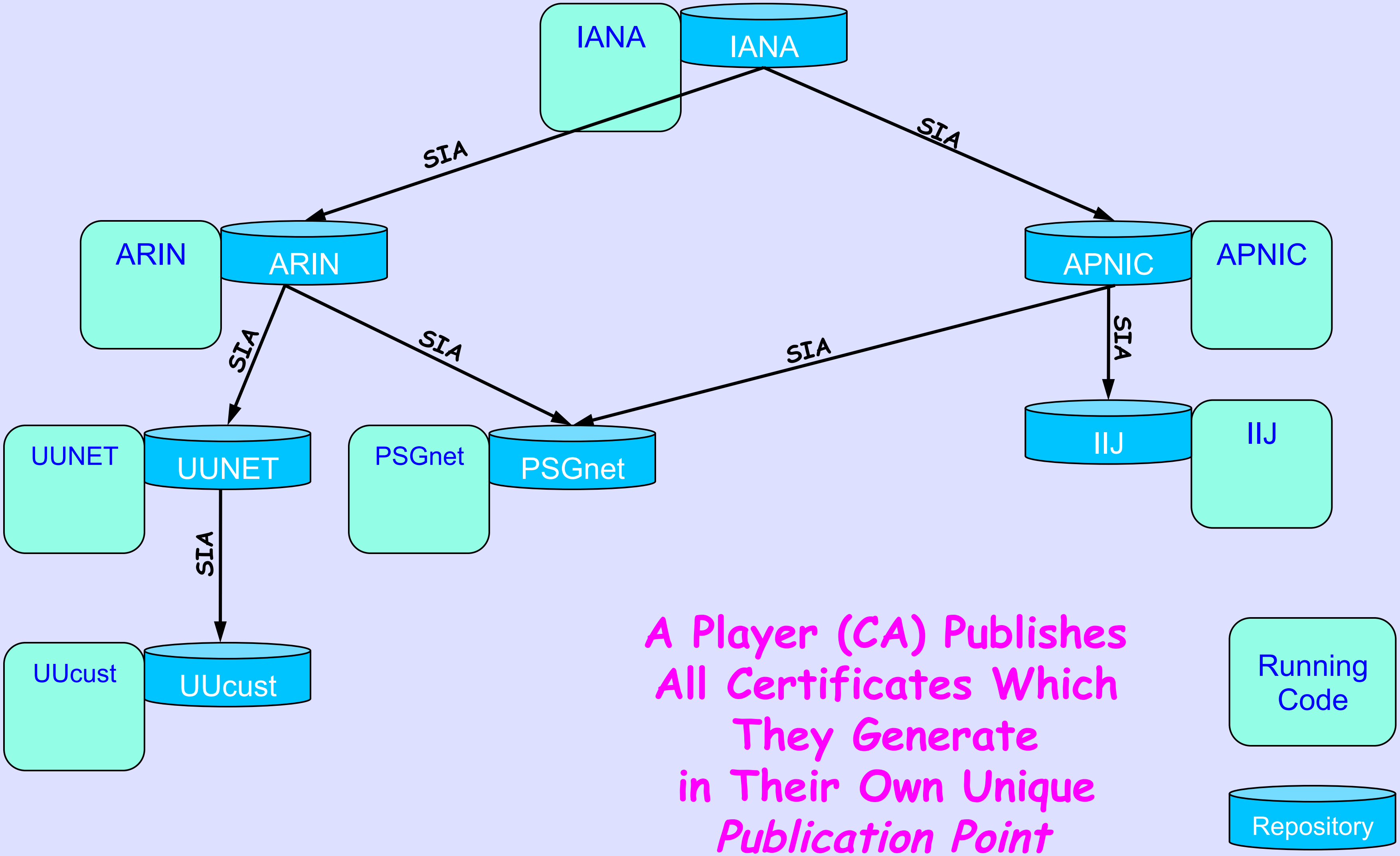
Business Key/Cert Management

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

RPKI DataBase

IP Resource Certs
ASN Resource Certs
Route Origin Attestations

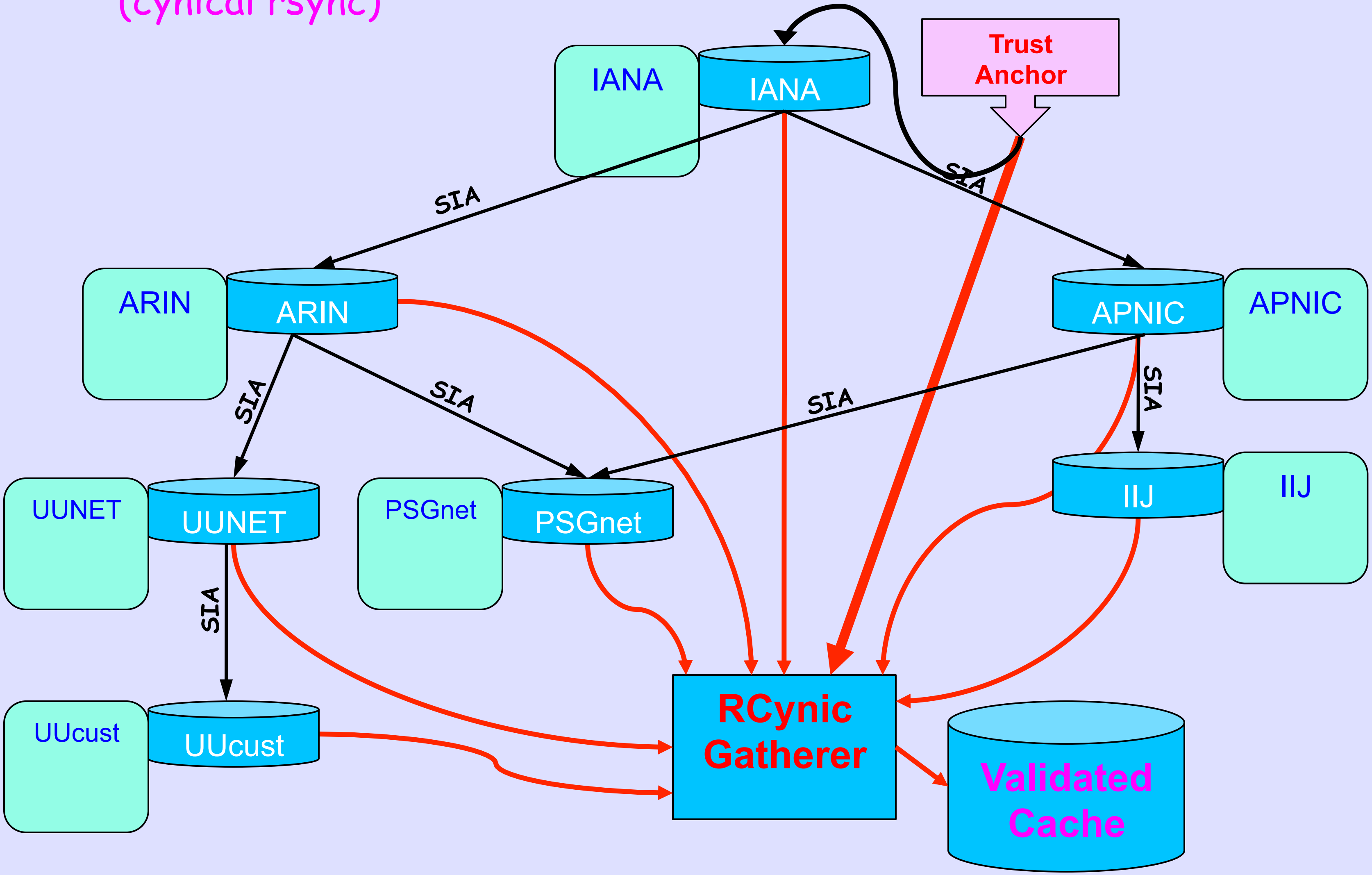
Distributed RPKI DataBase



A Player (CA) Publishes All Certificates Which They Generate in Their Own Unique *Publication Point*

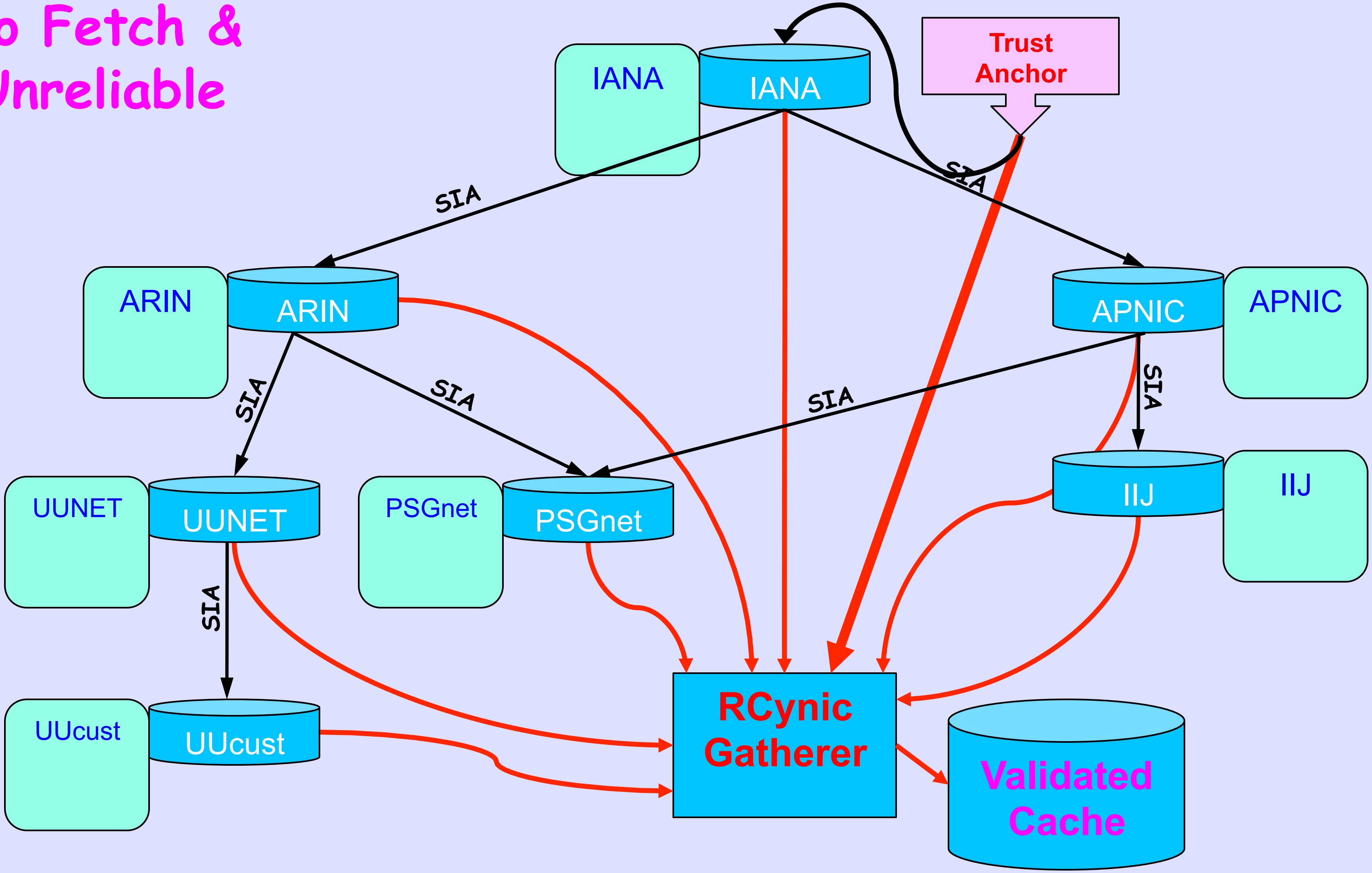
RCynic Cache Gatherer

(cynical rsync)

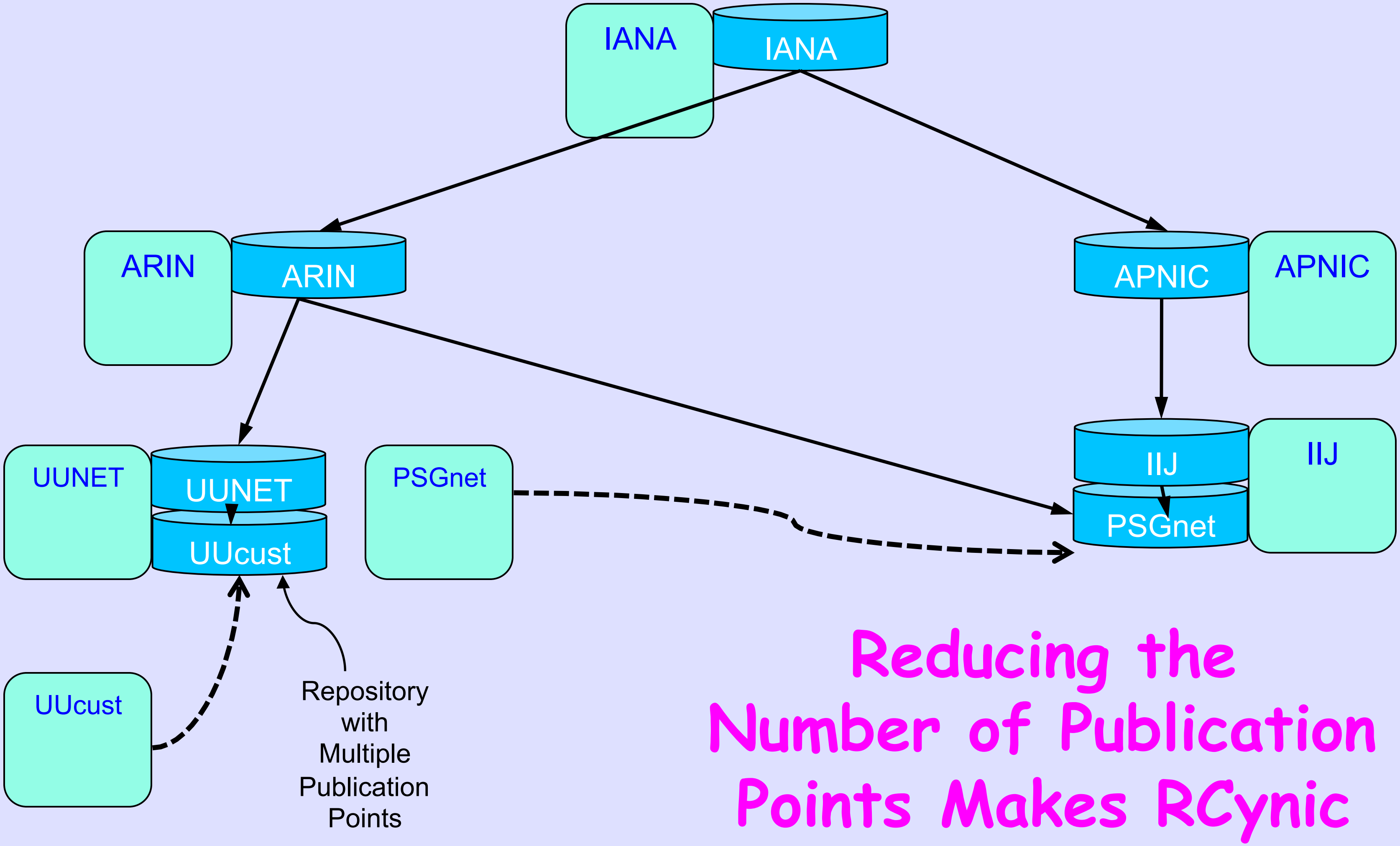


Reliability Issue

Expensive
To Fetch &
Unreliable



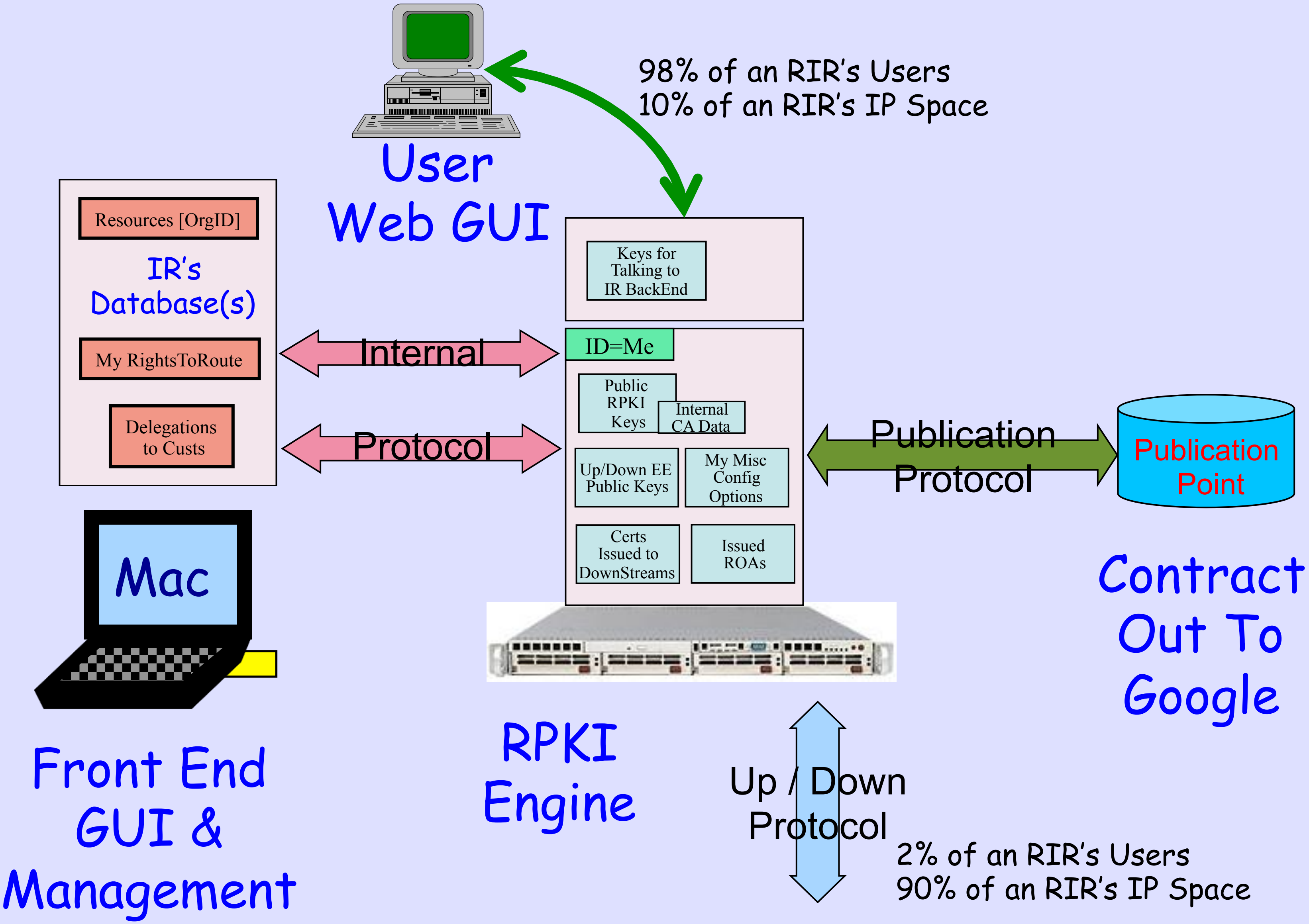
Reliability Via Hosted Publication



Repository with Multiple Publication Points

Reducing the Number of Publication Points Makes RCynic More Efficient

A Usage Scenario



98% of an RIR's Users
10% of an RIR's IP Space

User
Web GUI

Resources [OrgID]

IR's
Database(s)

My RightsToRoute

Delegations
to Custs

Internal

Protocol

Keys for
Talking to
IR BackEnd

ID=Me

Public
RPKI
Keys

Internal
CA Data

Up/Down EE
Public Keys

My Misc
Config
Options

Certs
Issued to
DownStreams

Issued
ROAs

Publication
Protocol

Publication
Point

Contract
Out To
Google

Mac

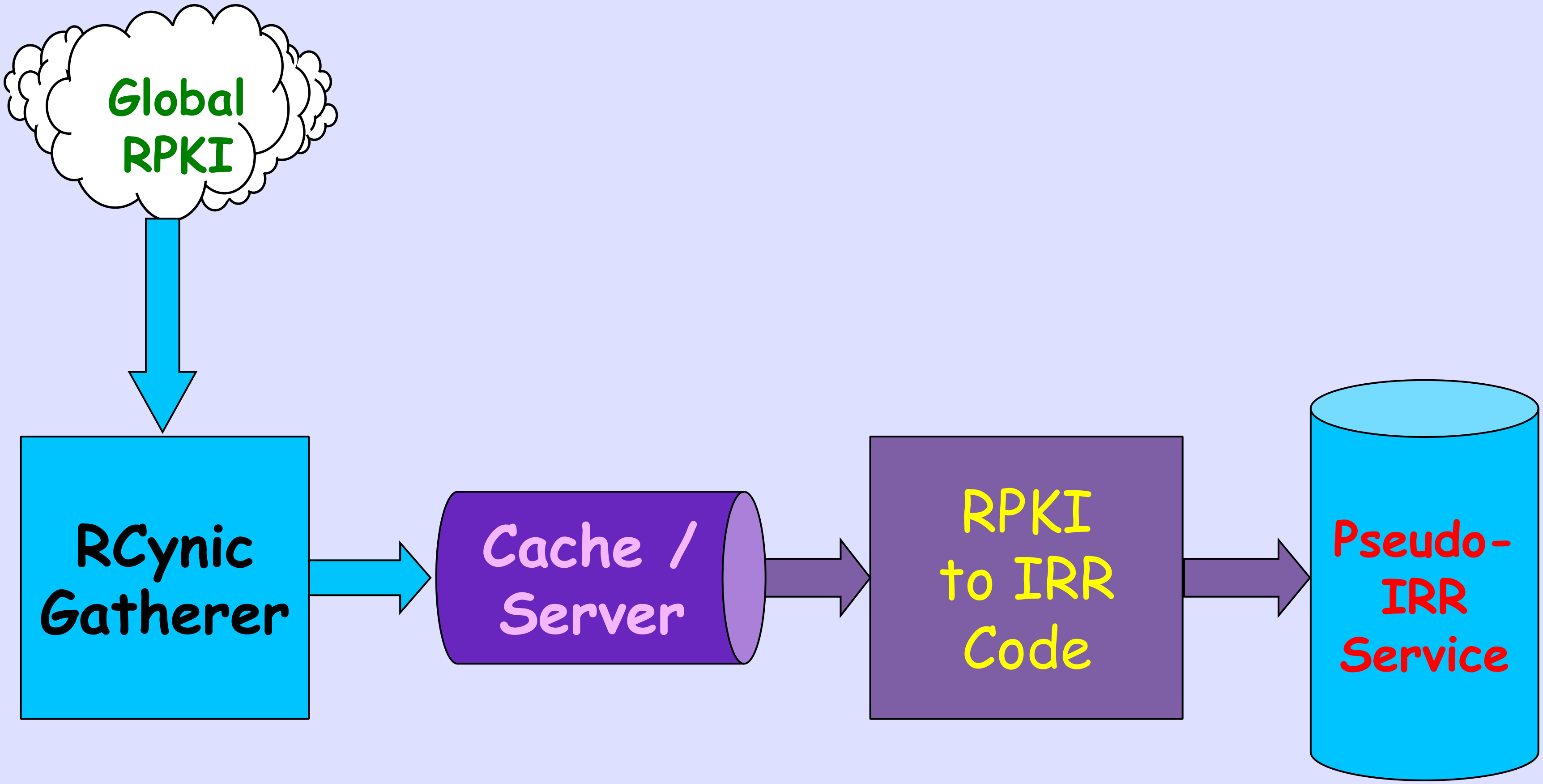
Front End
GUI &
Management

RPKI
Engine

Up / Down
Protocol

2% of an RIR's Users
90% of an RIR's IP Space

RPKI -> pseudo-IRR



Pseudo-IRR Data

```
% whois -h whois.rpki.net 147.28.0.0
route:          147.28.0.0/16
descr:          147.28.0.0/16-16
origin:         AS3130
notify:         irr-hack@rpki.net
mnt-by:         MAINT-RPKI
changed:        irr-hack@rpki.net 20101029
source:         RPKI
```

Pseudo-IRR Data

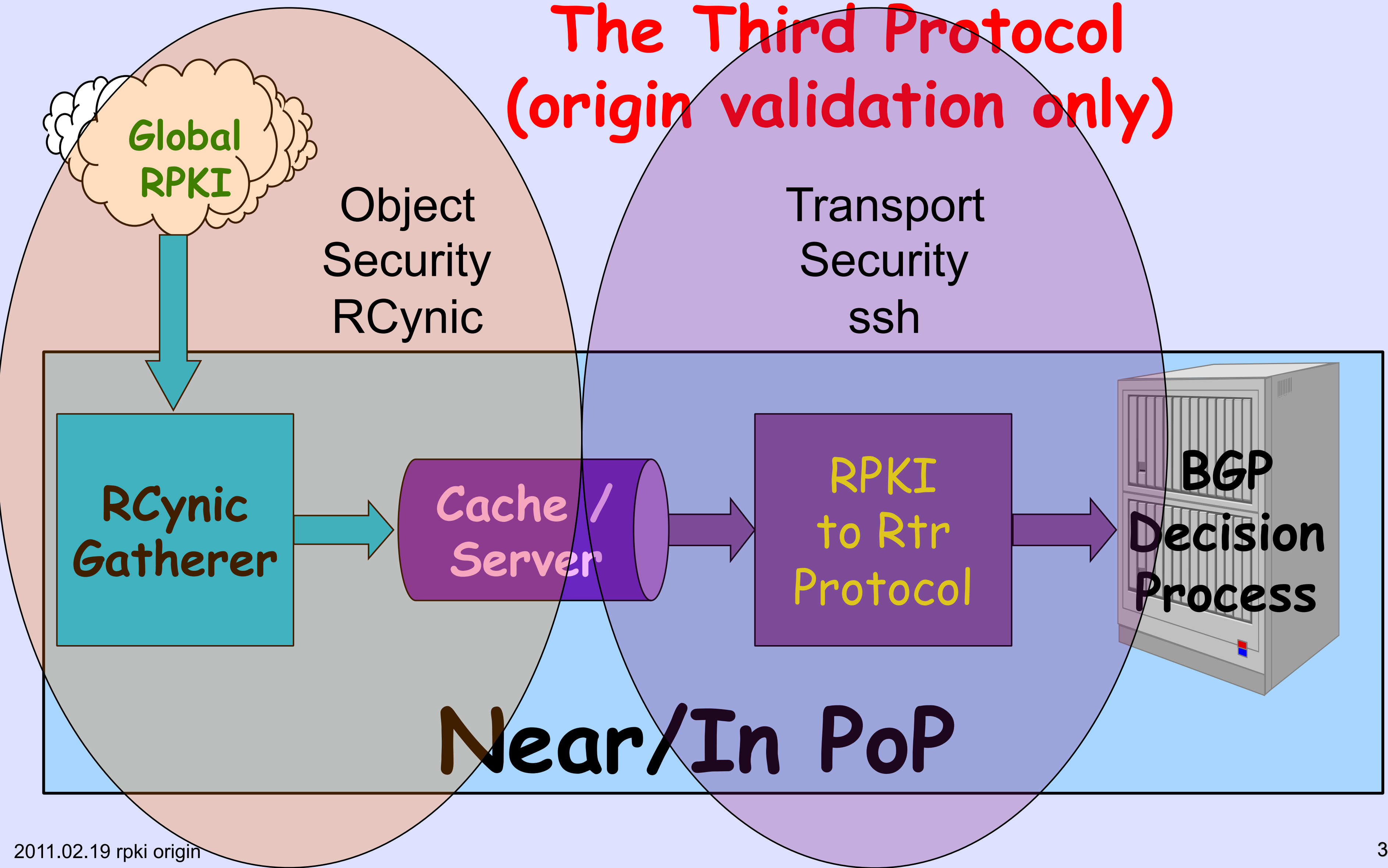
- Formally validated data
- Can be used by NOC/Admin to check when a customer or peer calls
- Can be used to generate router ACL filters for BGP peers

Origin 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!**
- Expect other vendor soon

RPKI -> Router

The Third 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
    ~                             ~
```

Reset Query

```
0          8          16         24        31
.-------------------------------------------.
| Protocol |   PDU    |                     |
| Version  |   Type   |    reserved = zero  |
|    0     |    2     |                     |
+-------------------------------------------+
|                                           |
|                 Length=8                  |
|                                           |
`-------------------------------------------'
```

Cache Response

```
0         8         16        24       31
.---------------------------------------.
| Protocol |  PDU    |                   |
| Version  |  Type   |  reserved = zero  |
|   0      |   3     |                   |
+---------------------------------------+
|                                       |
|               Length=8                |
|                                       |
`---------------------------------------'
```

IPv4 Prefix

```
0          8          16         24        31
.-------------------------------------------.
| Protocol |   PDU    |                     |
| Version  |   Type   |  reserved = zero    |
|    0     |    4     |                     |
+-------------------------------------------+
|                                           |
|                 Length=20                 |
|                                           |
+-------------------------------------------+
|          |  Prefix  |   Max    |          |
|  Flags   |  Length  |  Length  |   zero   |
|          |   0..32  |   0..32  |          |
+-------------------------------------------+
|                                           |
|                IPv4 prefix                |
|                                           |
+-------------------------------------------+
|                                           |
|         Autonomous System Number          |
|                                           |
`-------------------------------------------'
```

IPv6 Prefix

```
0          8          16         24        31
  .-------------------------------------------.
  | Protocol |   PDU    |                     |
  | Version  |   Type   |   reserved = zero   |
  |    0     |    6     |                     |
  +-------------------------------------------+
  |                                           |
  |                 Length=40                 |
  |                                           |
  +-------------------------------------------+
  |          |  Prefix  |   Max    |          |
  |  Flags   |  Length  |  Length  |   zero   |
  |          |  0..128  |  0..128  |          |
  +-------------------------------------------+
  |                                           |
  +---                                     ---+
  |                                           |
  +---            IPv6 prefix              ---+
  |                                           |
  +---                                     ---+
  |                                           |
  +-------------------------------------------+
  |                                           |
  |         Autonomous System Number          |
  |                                           |
  `-------------------------------------------'
```

End of Data

```
0          8          16          24        31
.-------------------------------------------.
| Protocol |    PDU     |                     |
| Version  |    Type    |   reserved = zero   |
|    0     |     7      |                     |
+-------------------------------------------+
|                                           |
|                 Length=12                 |
|                                           |
+-------------------------------------------+
|                                           |
|               Serial Number               |
|                                           |
`-------------------------------------------'
```

Notify (Think DNS)

```
0          8          16         24        31
.--------------------------------------------.
| Protocol |   PDU    |                      |
| Version  |   Type   |    reserved = zero   |
|    0     |    0     |                      |
+--------------------------------------------+
|                                            |
|                Length=12                   |
|                                            |
+--------------------------------------------+
|                                            |
|               Serial Number                |
|                                            |
`--------------------------------------------'
```

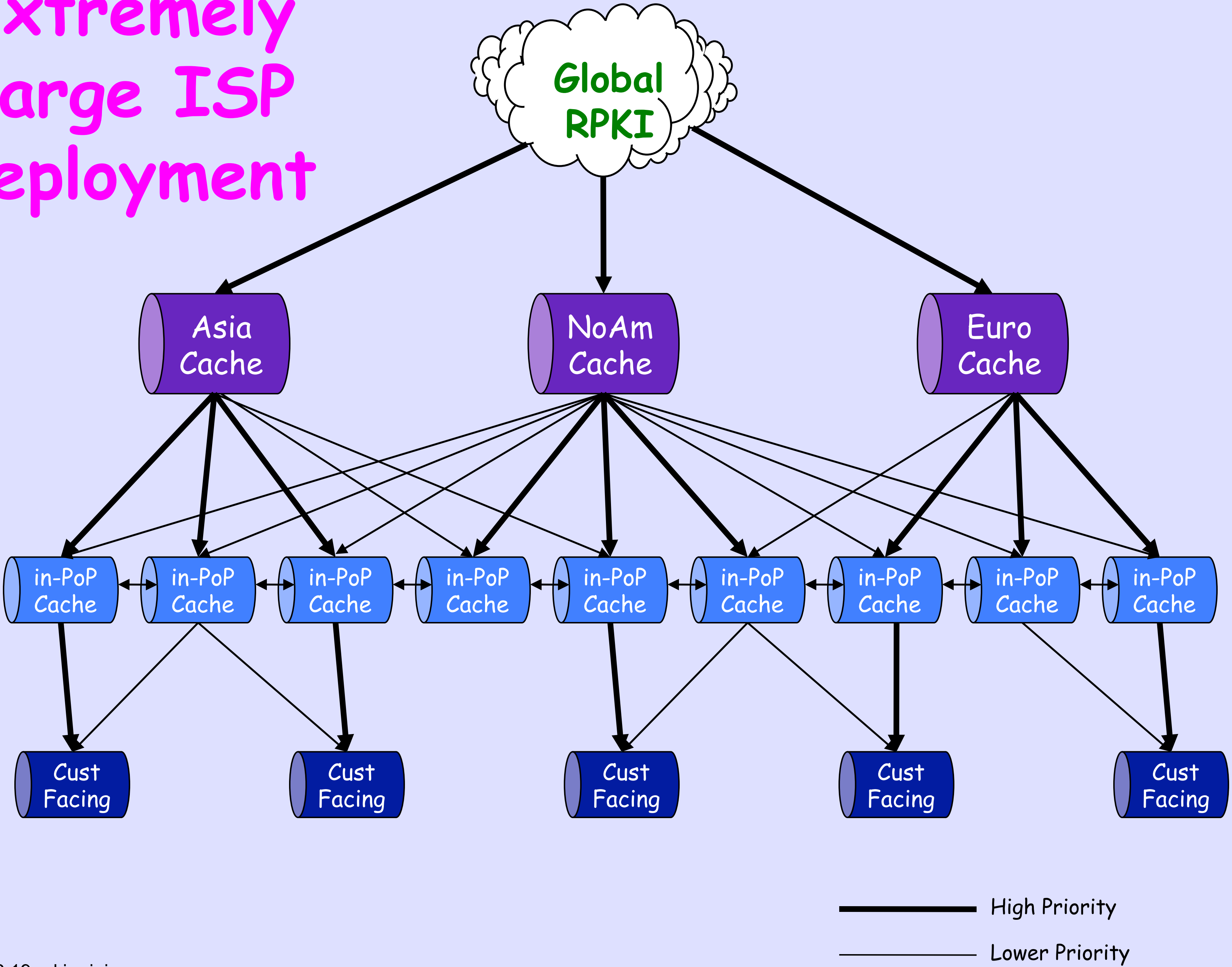
Serial Query

```
 0          8          16         24        31
   .-------------------------------------------.
   | Protocol |   PDU    |                     |
   | Version  |   Type   |    reserved = zero  |
   |    0     |    1     |                     |
   +-------------------------------------------+
   |                                           |
   |                Length=12                  |
   |                                           |
   +-------------------------------------------+
   |                                           |
   |               Serial Number               |
   |                                           |
   `-------------------------------------------'
```

Error Response

```
0          8          16         24        31
.-------------------------------------------.
| Protocol |   PDU    |                     |
| Version  |   Type   |    Error Number     |
|    0     |    10    |                     |
+-------------------------------------------+
|                                           |
|                  Length                   |
|                                           |
+-------------------------------------------+
|                                           |
|       Length of Encapsulated PDU          |
|                                           |
+-------------------------------------------+
|                                           |
~           Copy of Erroneous PDU           ~
|                                           |
+-------------------------------------------+
|                                           |
|           Length of Error Text            |
|                                           |
+-------------------------------------------+
|                                           |
|              Arbitrary Text               |
|                    of                     |
~          Error Diagnostic Message         ~
|                                           |
`-------------------------------------------'
```

Extremely Large ISP Deployment

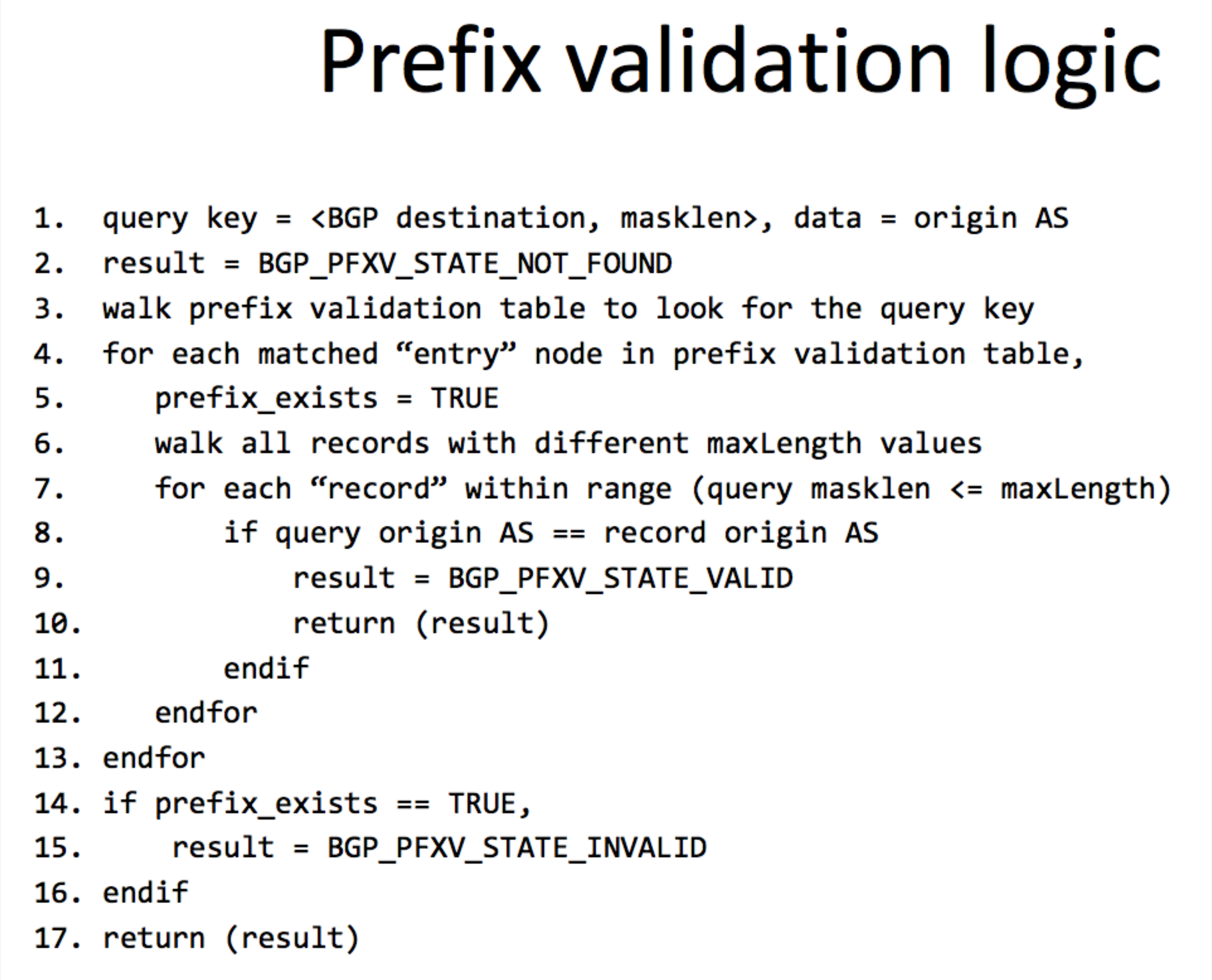


Configure

```
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
- **Unknown** – No matching or covering ROA was found



Prefix validation logic

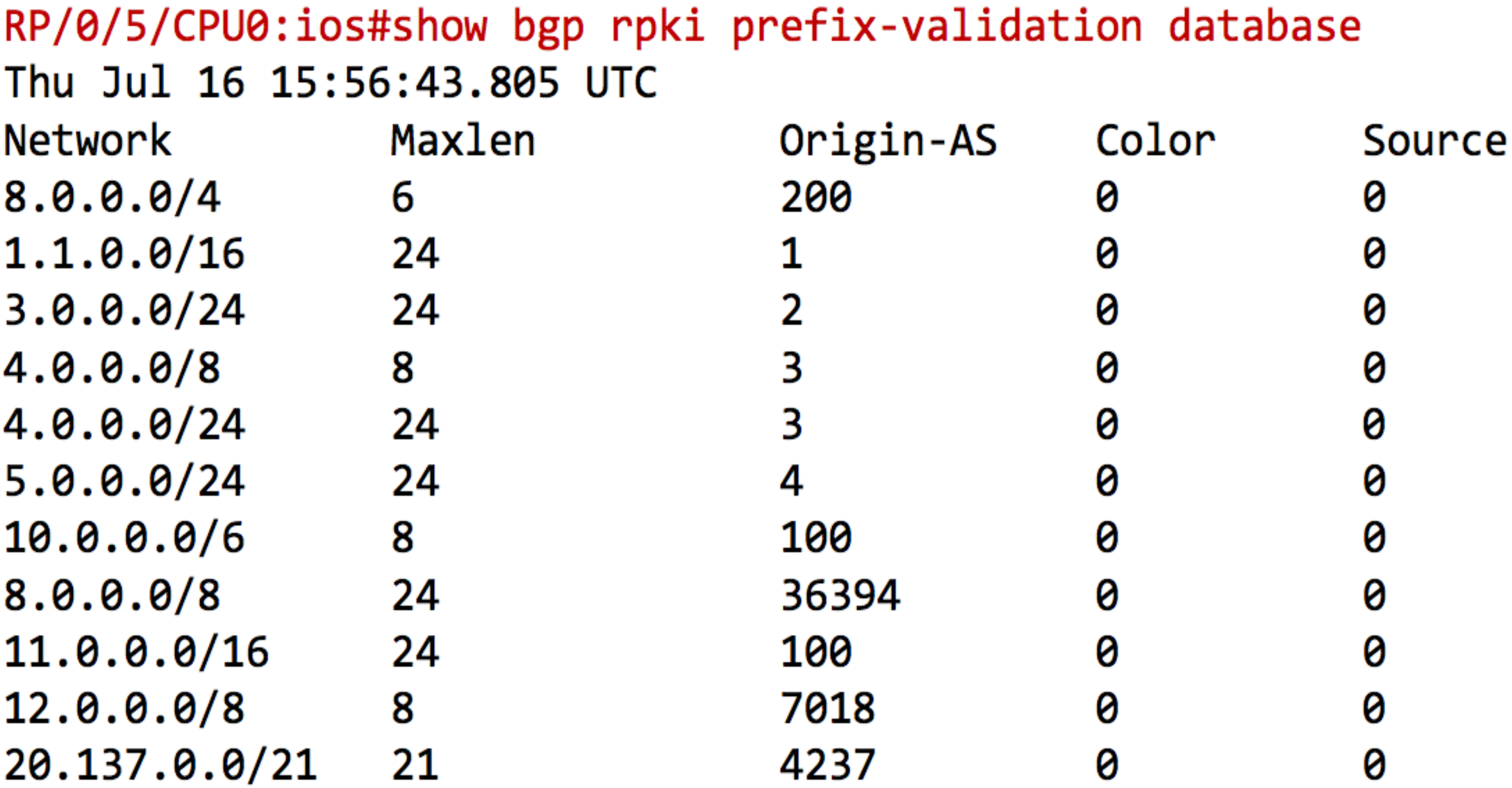
```
query key = <BGP destination, masklen>, data = origin AS
result = BGP_PFXV_STATE_NOT_FOUND
walk prefix validation table to look for the query key
for each matched “entry” node in prefix validation table,
    prefix_exists = TRUE
    walk all records with different maxLength values
    for each “record” within range (query masklen <= maxLength)
        if query origin AS == record origin AS
            result = BGP_PFXV_STATE_VALID
            return (result)
        endif
    endfor
endfor
if prefix_exists == TRUE,
     result = BGP_PFXV_STATE_INVALID
endif
return (result)
```

Policy Override Knobs

- Disable Validity Check Completely
- Disable Validity Check for a Peer
- Disable Validity Check for Prefixes

When check is disabled, the result is "Not Found," i.e. as if there was no ROA

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
```

Defaults

- Origin Validation is Enabled if you have configured a cache server peering
- RPKI Poll Interval is 30 Minutes
- No Effect on Policy unless you have configured it

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
...
```

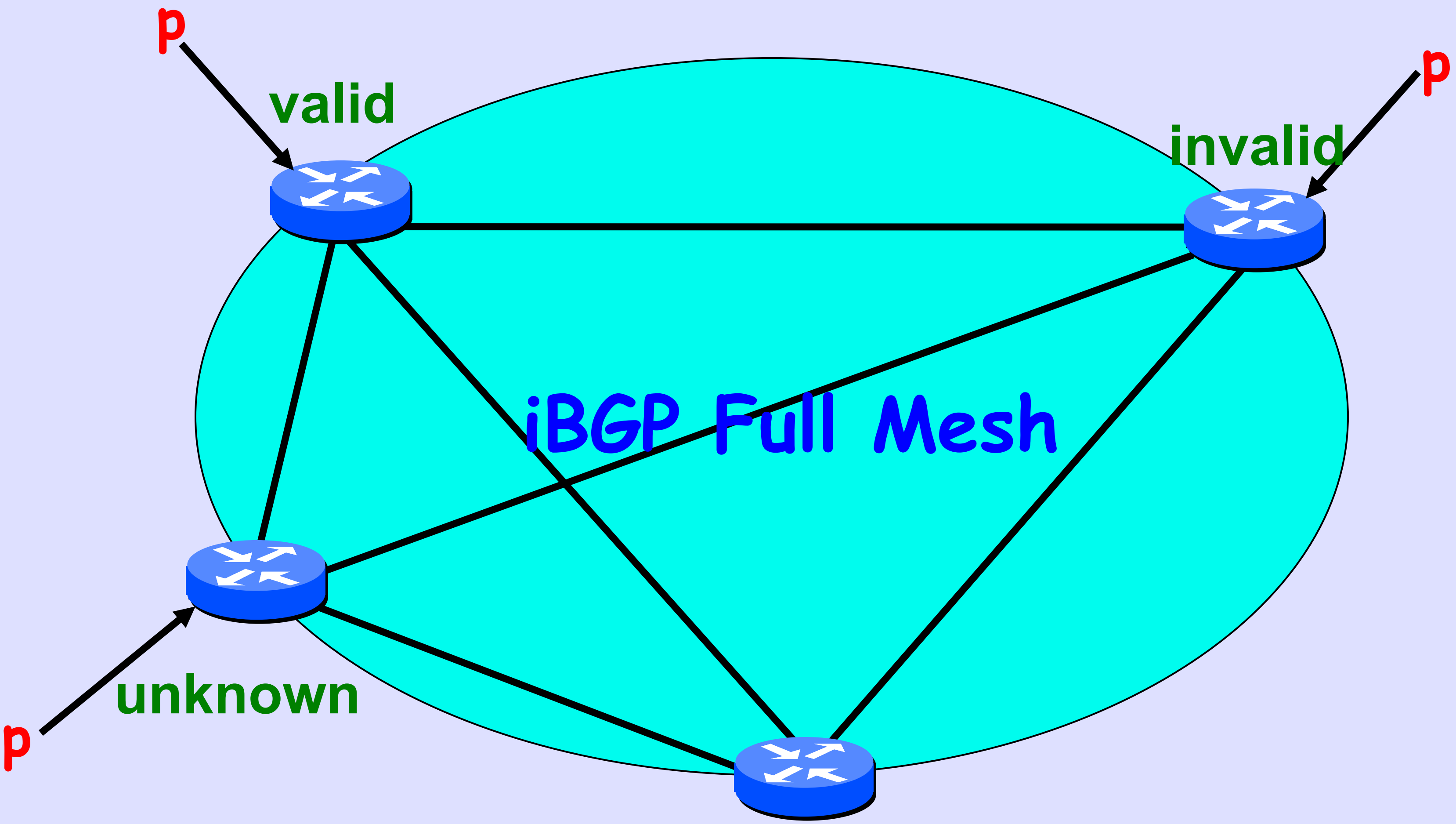
Bad Dog!

```
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



which do i choose?
why do i choose it?

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
```

The Solution is to Allow Operator to Test and then Set Local Policy

Fairly Secure

```
route-map validity-0
  match rpki-invalid
  drop
route-map validity-1
  match rpki-not-found
  set localpref 50
// valid defaults to 100
```

Paranoid

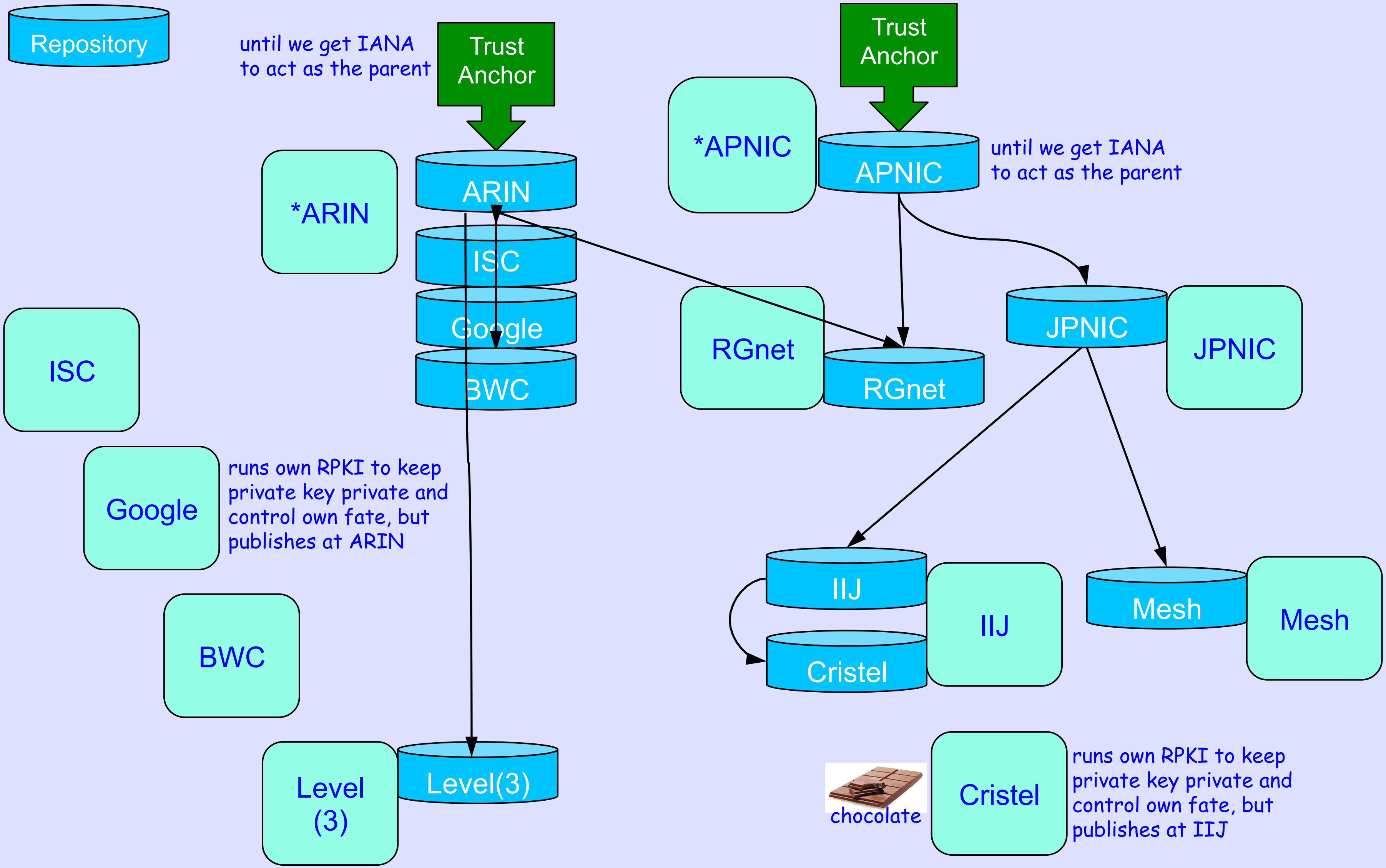
```
route-map validity-0
  match rpki-valid
  set localpref 110
route-map validity-1
  drop
```

After AS-Path

```
route-map validity-0
 match rpki-unknown
   set metric 50
route-map validity-1
   match rpki-invalid
   set metric 25
route-map validity-2
   set metric 100
```

Running Code

The Open TestBed

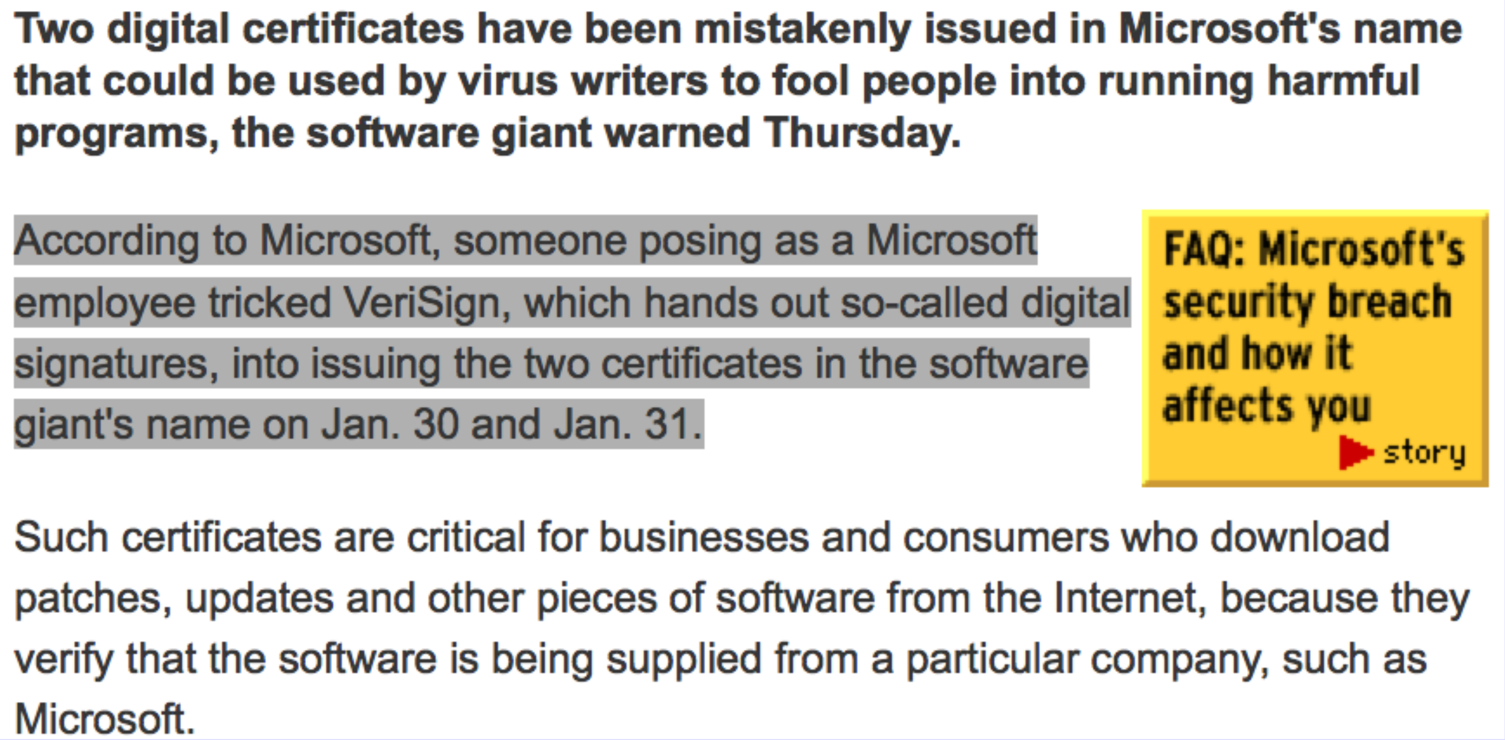

Repository
until we get IANA to act as the parent
Trust Anchor
*ARIN
ARIN
ISC
Google
BWC
Trust Anchor
*APNIC
APNIC
until we get IANA to act as the parent
RGnet
RGnet
JPNIC
JPNIC
ISC
Google
runs own RPKI to keep private key private and control own fate, but publishes at ARIN
BWC
IIJ
Cristel
IIJ
Mesh
Mesh
Level (3)
Level(3)
chocolate
Cristel
runs own RPKI to keep private key private and control own fate, but publishes at IIJ

* APNIC and ARIN are simulations constructed from public data

The Big Speedbump

"We will never control Internet routing"
-- The RIRs
Certificate
Expiration
Revocation

But Who Do We Trust?



Two digital certificates have been mistakenly issued in Microsoft's name that could be used by virus writers to fool people into running harmful programs, the software giant warned Thursday.

According to Microsoft, someone posing as a Microsoft employee tricked VeriSign, which hands out so-called digital signatures, into issuing the two certificates in the software giant's name on Jan. 30 and Jan. 31.

FAQ: Microsoft's security breach and how it affects you

Such certificates are critical for businesses and consumers who download patches, updates and other pieces of software from the Internet, because they verify that the software is being supplied from a particular company, such as Microsoft.

http://news.cnet.com/2100-1001-254586.html

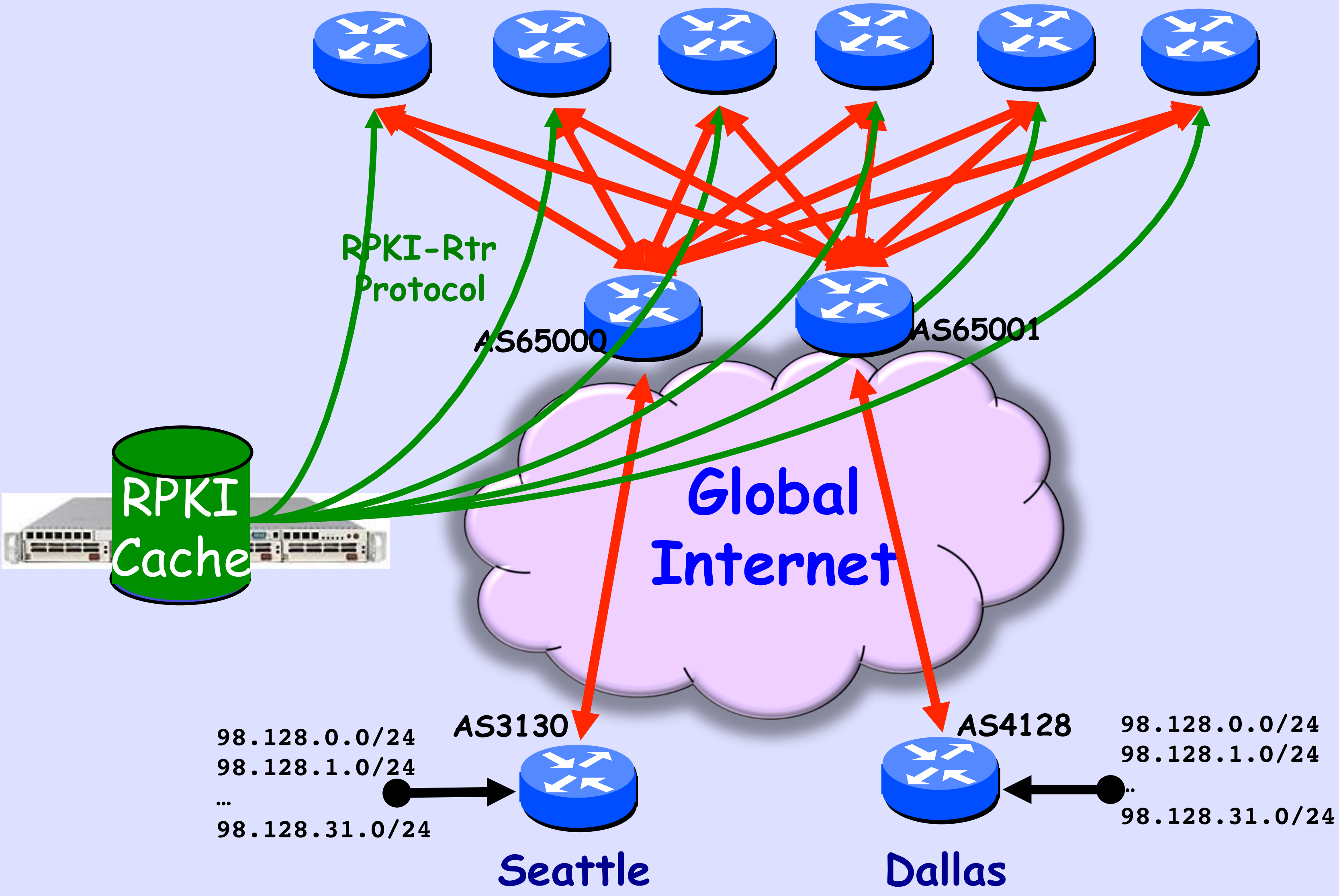
Open Source (BSD Lisc) Running Code

https://www.rpki.net/

Test Code in Routers

Talk to Ed Kern

Lab Topology



Work Supported By

- US Government

  THIS PROJECT IS SPONSORED BY THE DEPARTMENT OF HOMELAND SECURITY UNDER AN INTERAGENCY AGREEMENT WITH THE AIR FORCE RESEARCH LABORATORY (AFRL).

- ARIN
- Internet Initiative Japan
- Cisco, Google, NTT, Equinix