BGP Origin Validation

APNIC / Phnom Penh

2012.08.27

Randy Bush <randy@psg.com>
Rob Austein <sra@isc.org>
Steve Bellovin <smb@cs.columbia.edu>

And a cast of thousands! Well, dozens :)

Agenda

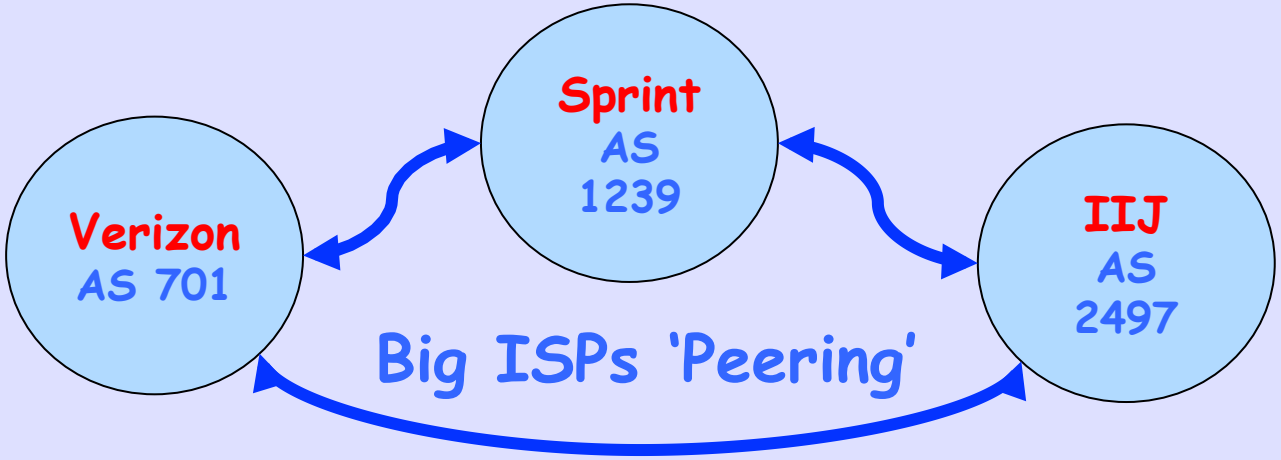
- This Presentation
  - Some Technical Background
  - Mis-Origination - YouTube Incident
  - The RPKI – Needed Infrastructure
  - BGP Origin Validation
  - BGP Path Validation (briefly)

This is Not New

- 1986 – Bellovin & Perlman identify the vulnerability
- 1999 - National Academies study called it out
- 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

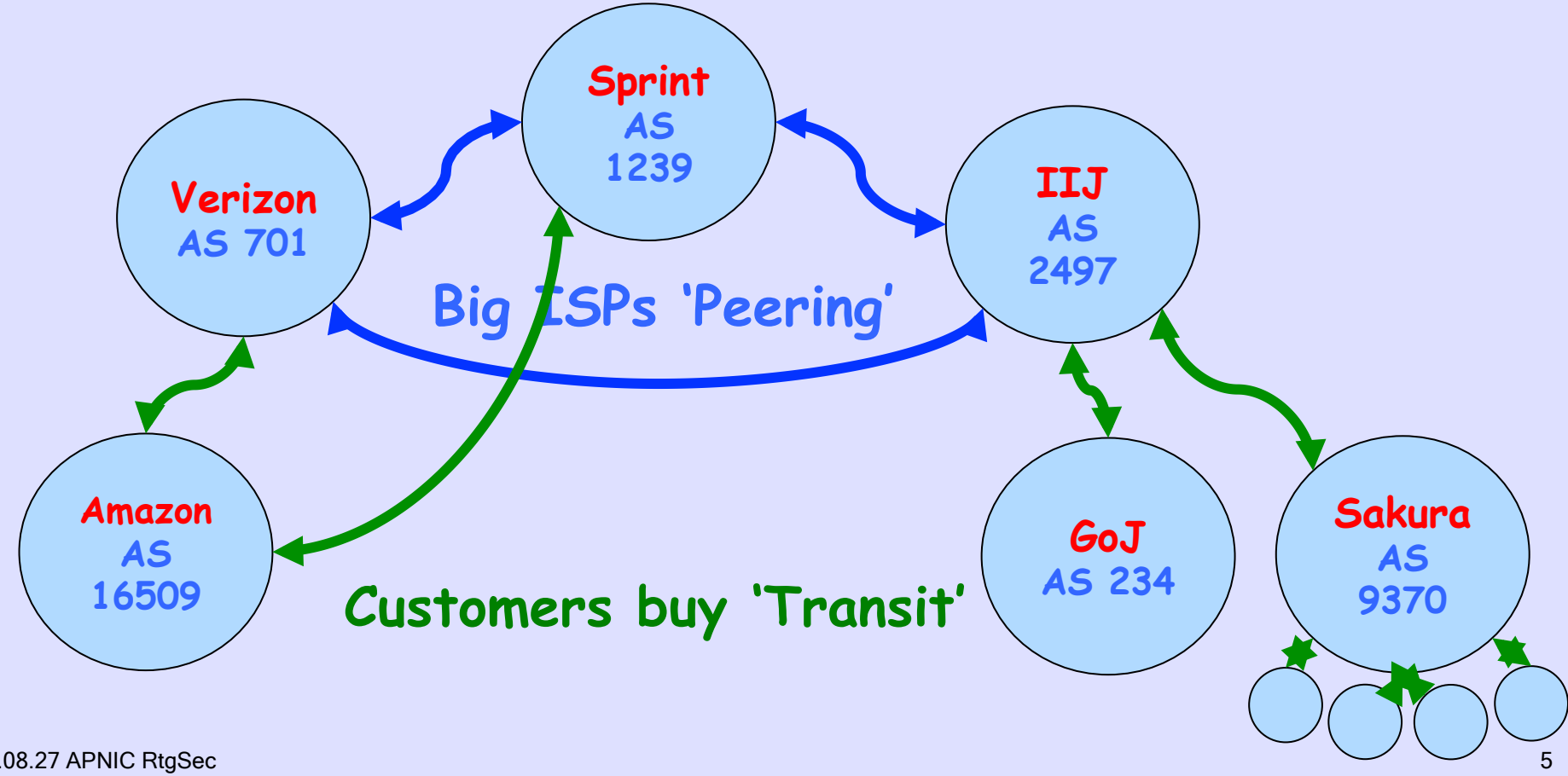
What is an AS?

An ISP or End Site

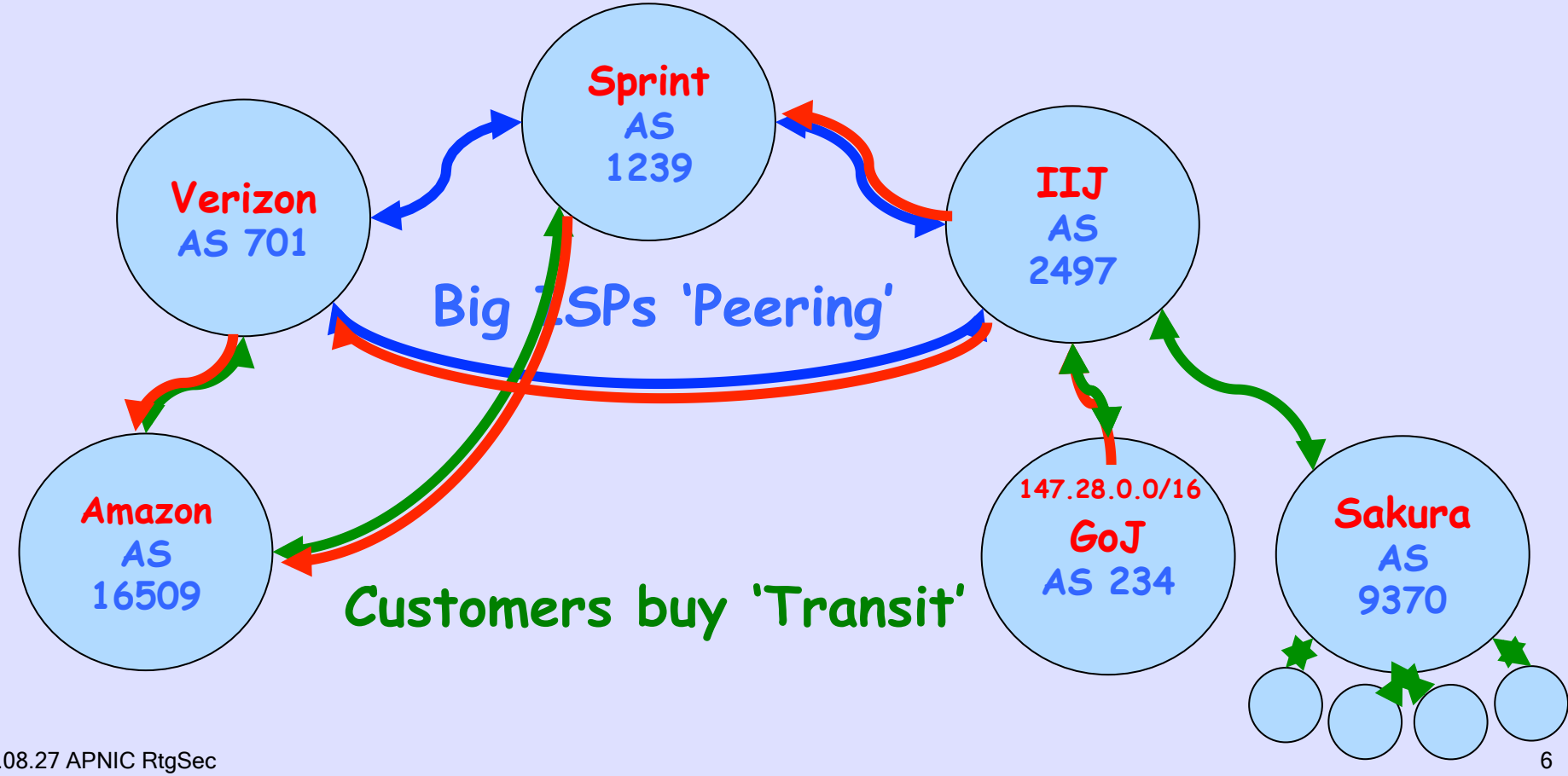


What is an AS?

An ISP or End Site



An IP Prefix is Announced & Propagated



From Inside a Router

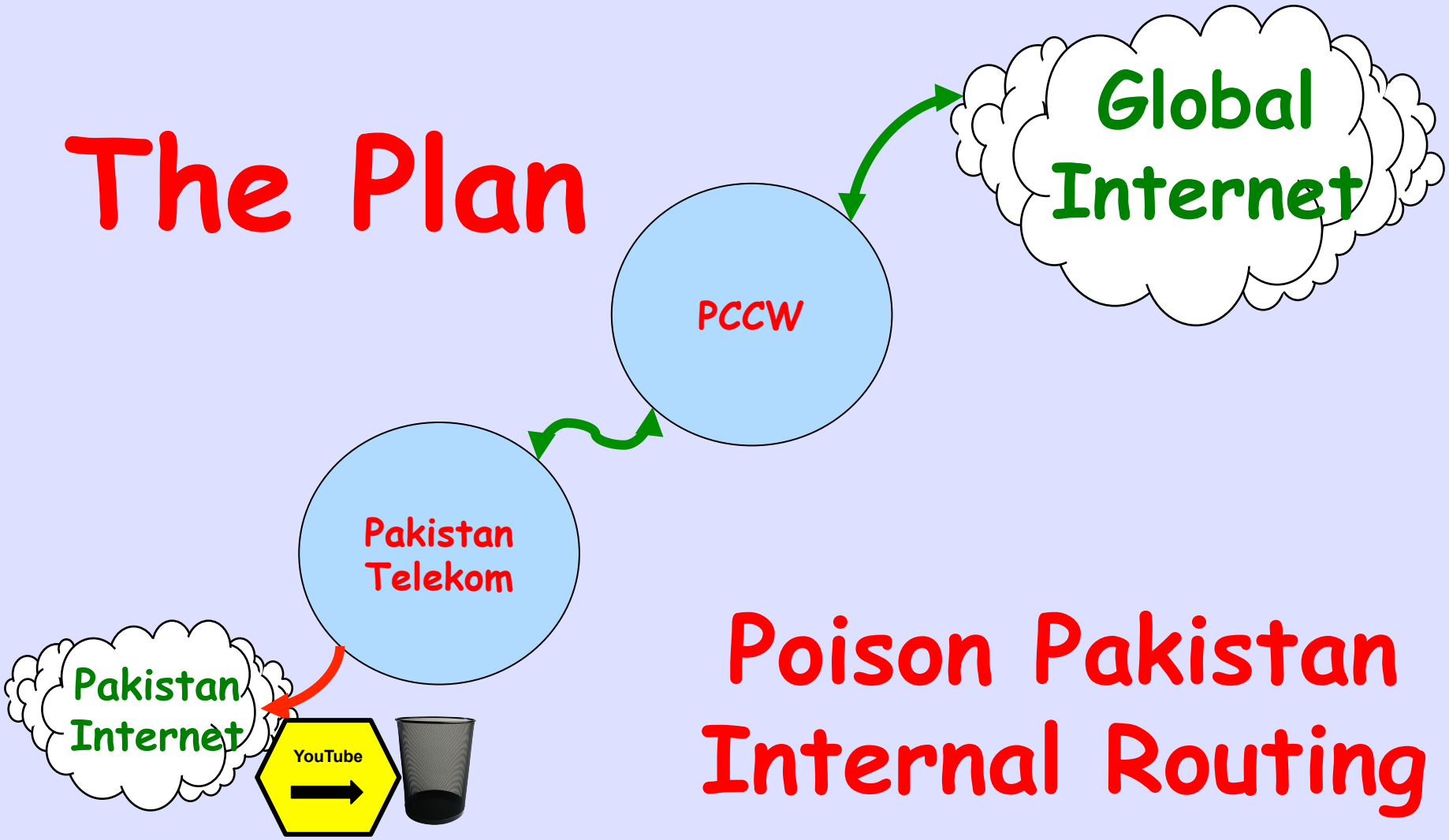
BGP routing table entry for 147.28.0.0/16

Of Course it's Uglier ☺

```
r1.iad#sh ip bgp 147.28.0.0/16
BGP routing table entry for 147.28.0.0/16, version 21440610
Paths: (2 available, best #1, table default)
  Advertised to update-groups:
     1
  Refresh Epoch 1
  16509   1239   2497   234
    144.232.18.81 from 144.232.18.81 (144.228.241.254)
      Origin IGP, metric 841, localpref 100, valid, external, best
      Community: 3297:100 3927:380
      path 67E8FFCC RPKI State valid
  Refresh Epoch 1
  16509   701  2497   234
    129.250.10.157 (metric 11) from 198.180.150.253 (198.180.150.253)
      Origin IGP, metric 95, localpref 100, valid, internal
      Community: 2914:410 2914:1007 2914:2000 2914:3000 3927:380
      path 699A867C RPKI State valid
```

The YouTube Incident

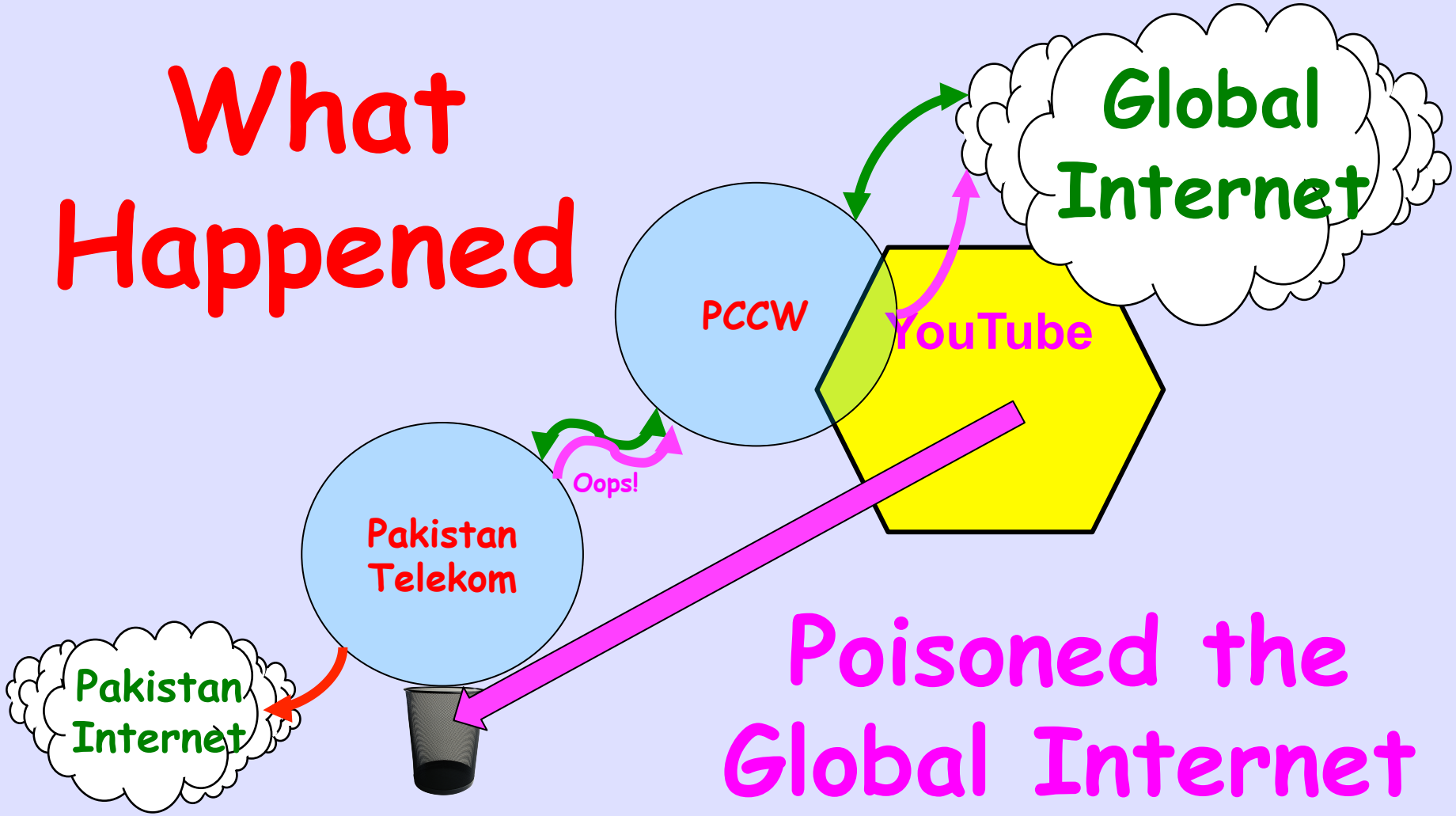
The Plan



Poison Pakistan Internal Routing

The YouTube Incident

What Happened



Poisoned the Global Internet

We Call this *Mis-Origination*

a Prefix is Originated by an AS Which Does Not Own It

I Do Not Call it *Hijacking*

Because that *Assumes* Negative Intent

And These Accidents Happen Every Day

Usually to Small Folk Sometimes to Large

So,

What's the Plan?

Three Pieces

- **RPKI** – Resource Public Key Infrastructure, the Certificate Infrastructure to Support the other Pieces (starting last year)
- **Origin Validation** – Using the RPKI to detect and prevent mis-originations of someone else's prefixes (early 2012)
- **AS-Path Validation** AKA **BGPsec** – Prevent Attacks on BGP (future work)

Why Origin Validation?

- Prevent YouTube accident & Far Worse
- 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', the third step, a few years away

We Need to be Able to Authoritatively Prove Who Owns an IP Prefix And What AS(s) May Announce It

Prefix Ownership Follows the Allocation Hierarchy

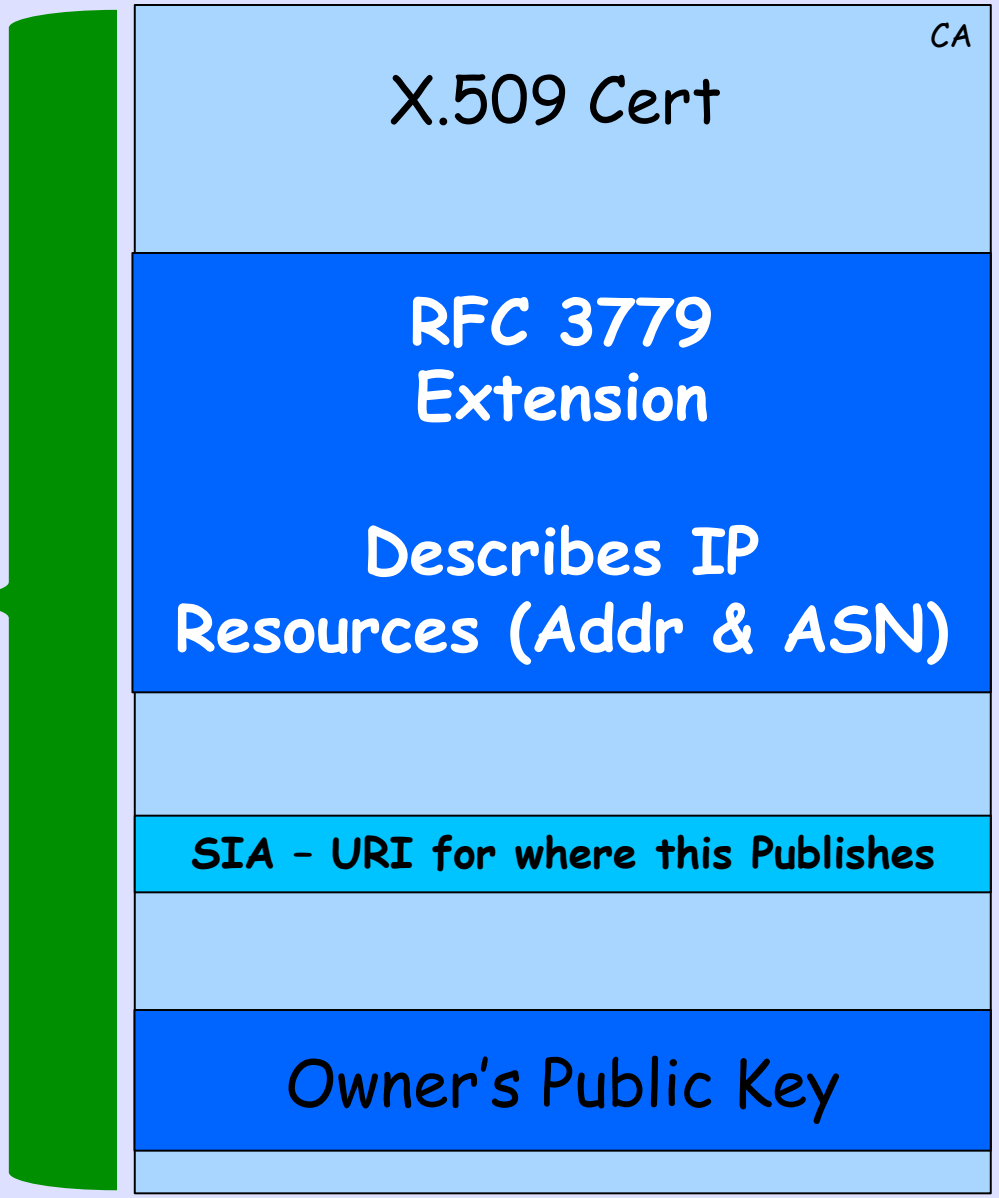
IANA, RIRs, ISPs, …

Resource Public Key Infrastructure (RPKI)

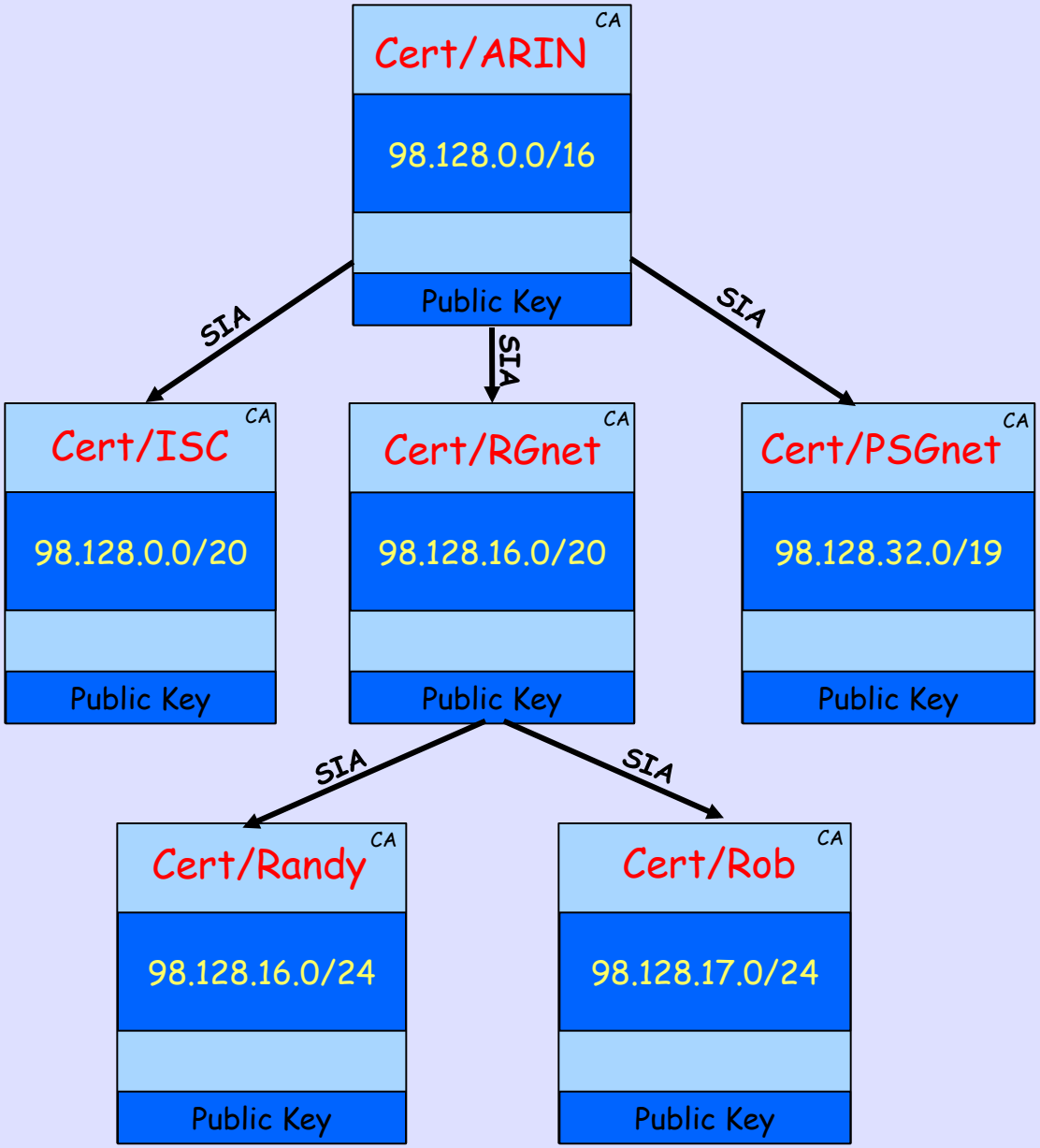
X.509 RPKI Being Developed & Deployed by IANA, RIRs, and Operators

X.509 Certificate w/ 3779 Ext

Signed by Parent's Private Key

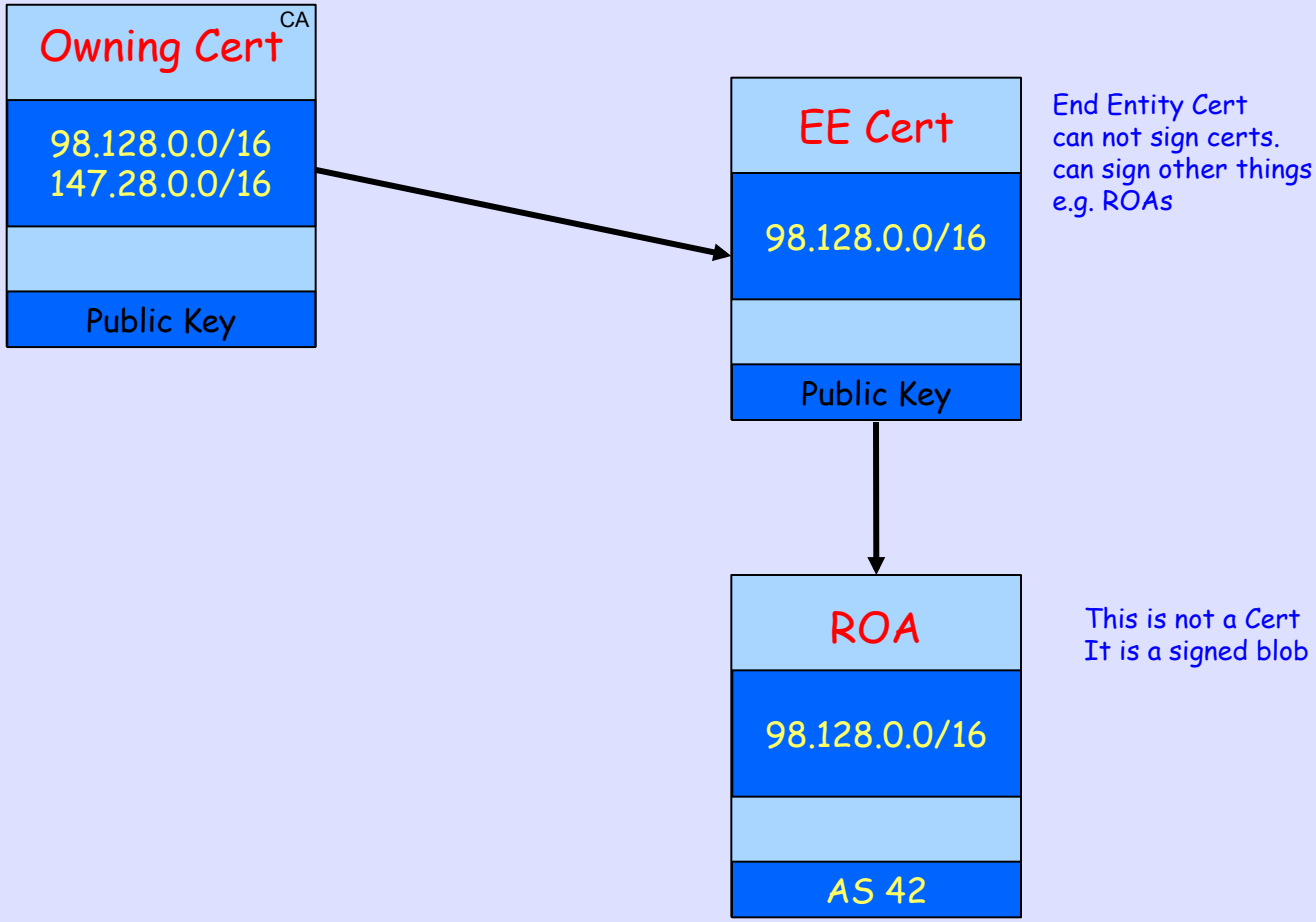


Certificate Hierarchy follows Allocation Hierarchy

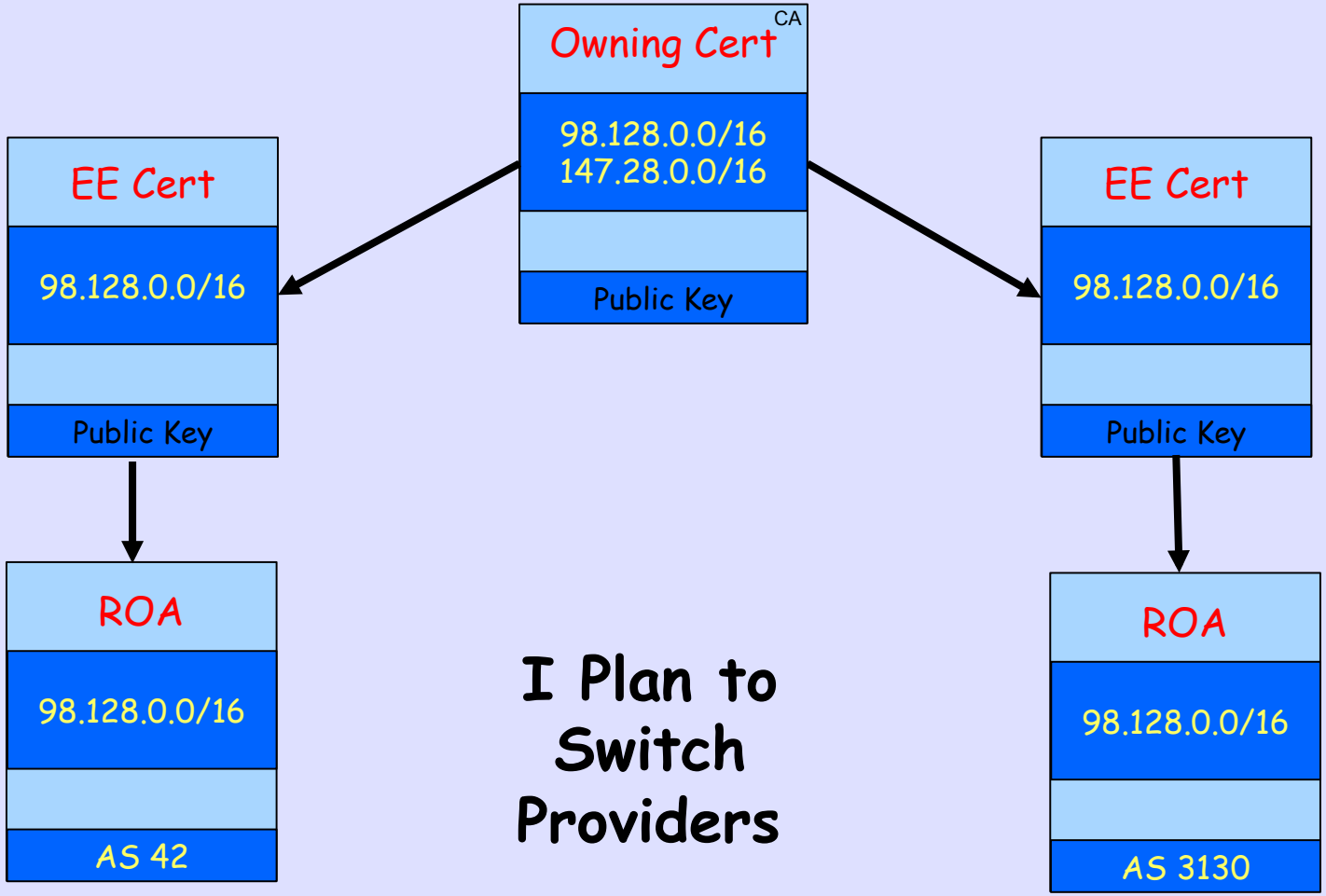


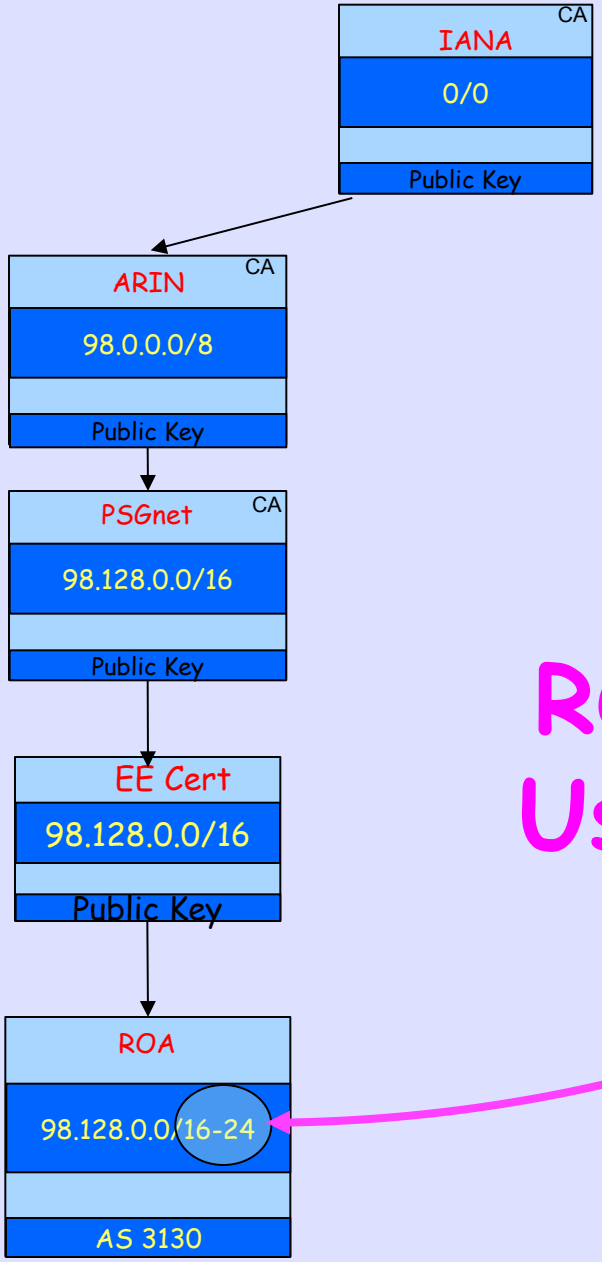
That's Who Owns It
but
Who May Route It?

Route Origin Authorization (ROA)



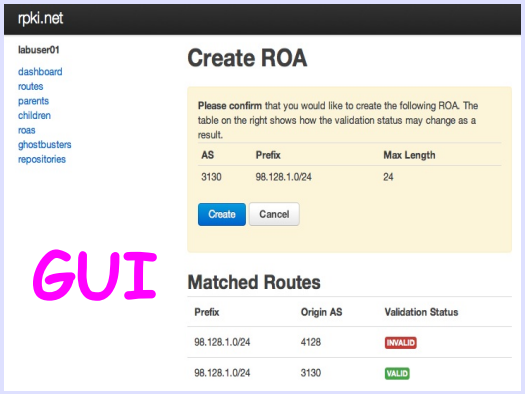
Multiple ROAs Make Before Break





ROA Aggregation Using Max Length

RPKI-Based Origin Validation



GUI

Up / Down to Parent

RPKI Certificate Engine

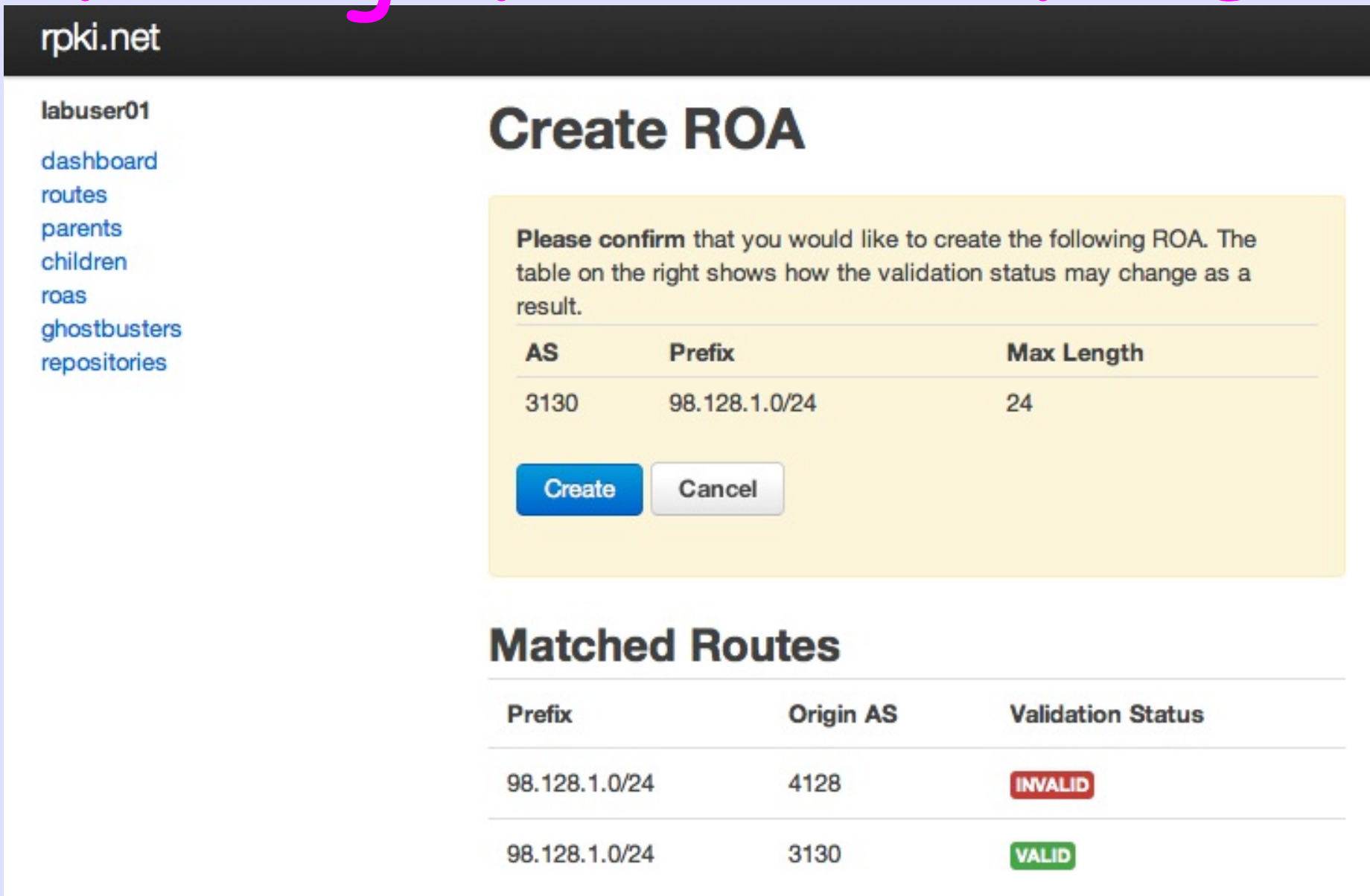
Up / Down to Child

Publication Protocol

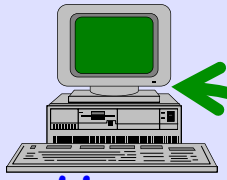
Resource PKI

IP Resource Certs
ASN Resource Certs
Route Origin Attestations

Warning What ROA Will Do



A Usage Scenario



User
Web GUI

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

Resources [OrgID]

Registry's Database(s)

My RightsToRoute

Delegations to Custs

Internal

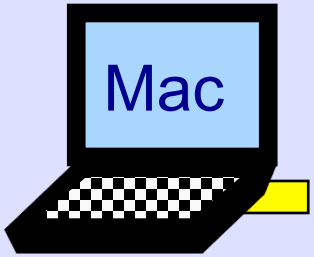
Protocol

RPKI Certificate Engine

Publication Protocol

Publication Point

Contract Out To Google? ☺

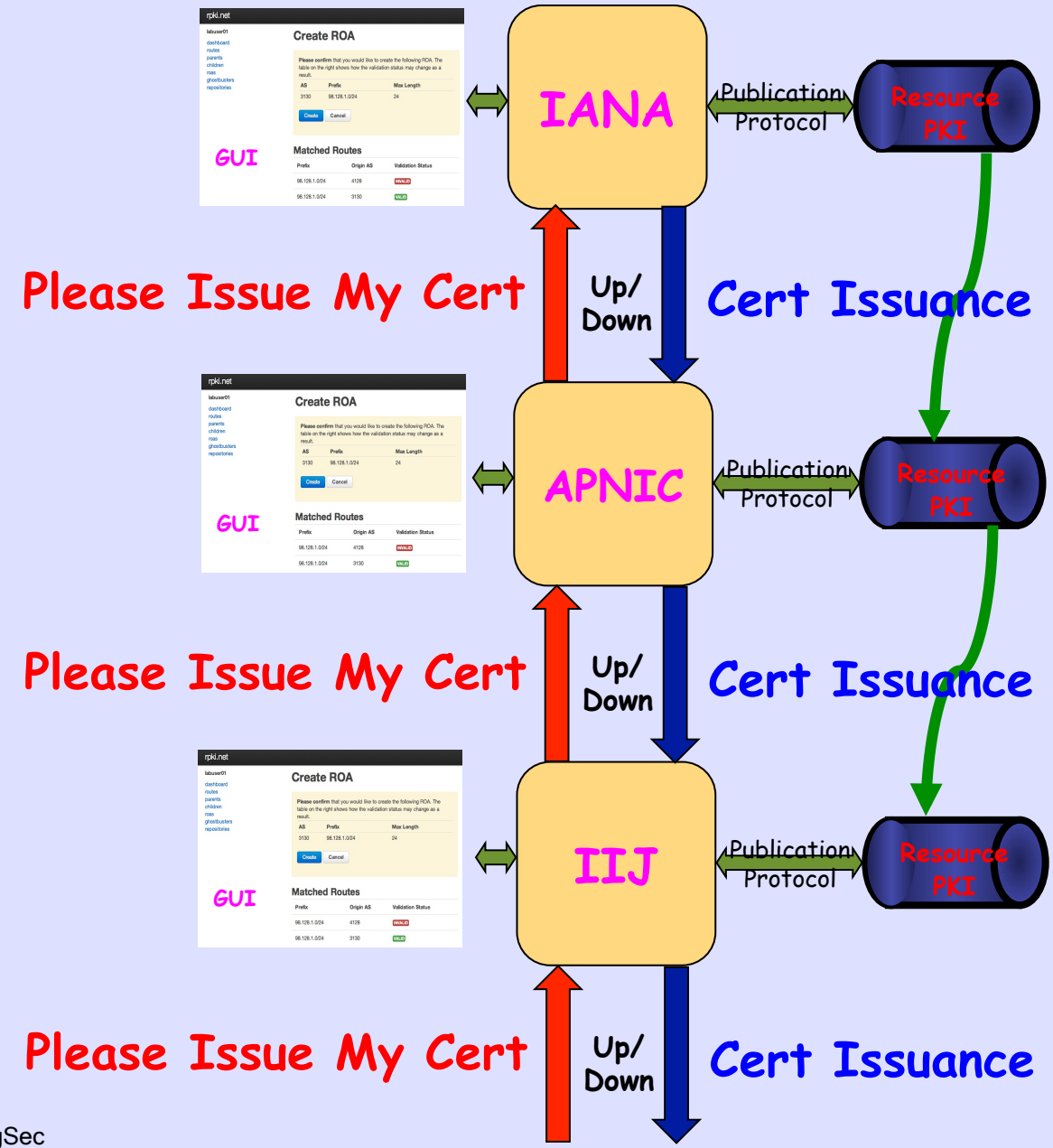


Registry Internal GUI & Management

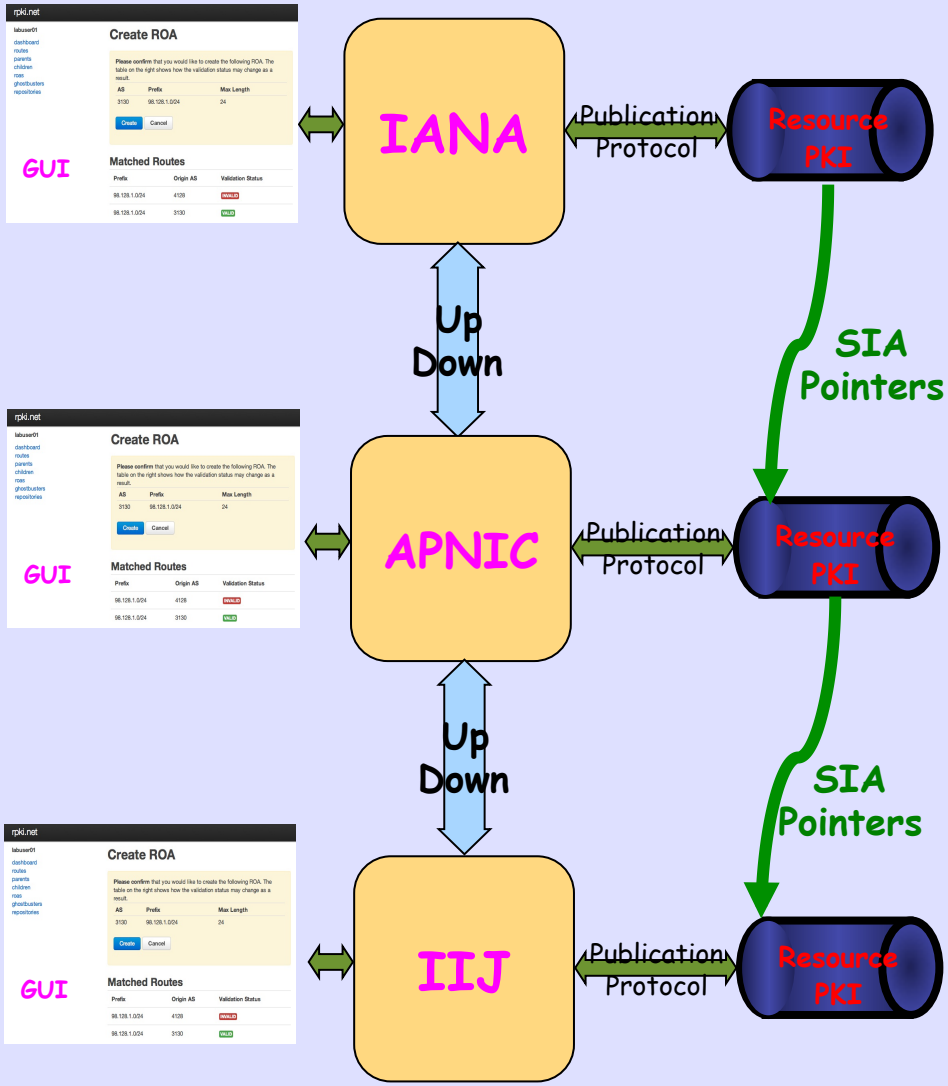
Up / Down Protocol

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

Issuing Parties

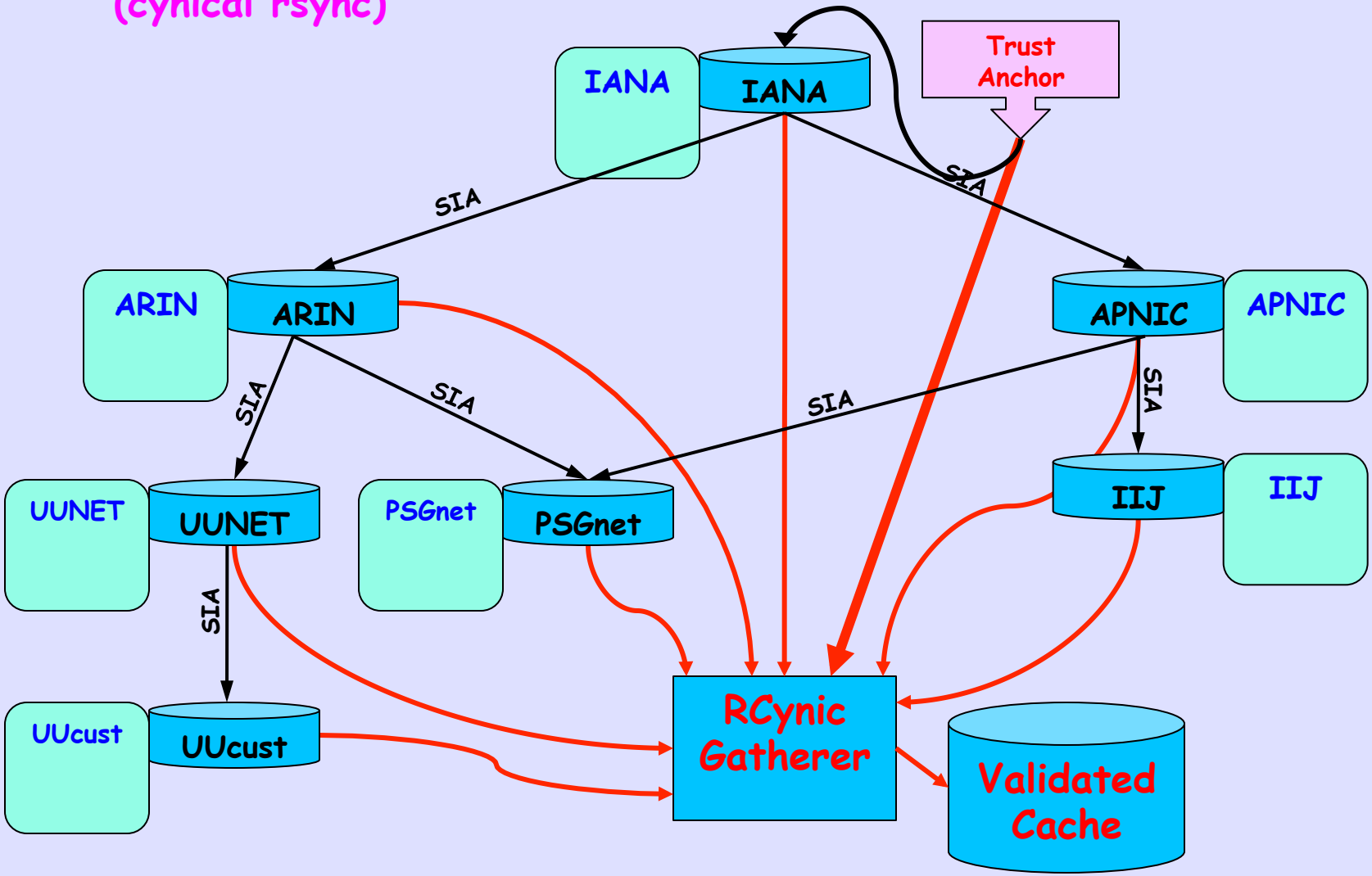


Issuing Parties

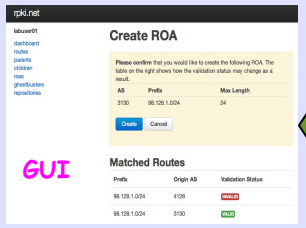


RCynic Cache Gatherer

(cynical rsync)



Issuing Parties Relying Parties



IANA

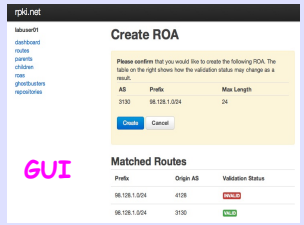
Publication Protocol

Trust Anchor

Resource PKI

Up Down

SIA Pointers



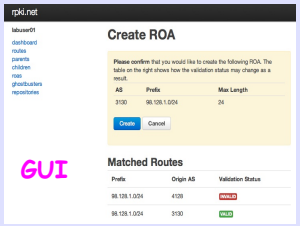
APNIC

Publication Protocol

Resource PKI

Up Down

SIA Pointers



IIJ

Publication Protocol

Resource PKI

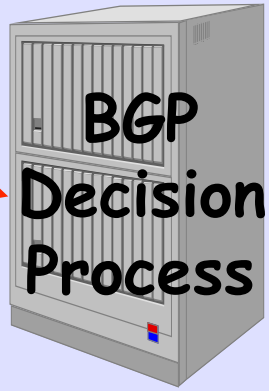
RCynic Gatherer

Validated Cache

Pseudo IRR

```
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 20110606
source:   RPKI
```

NOC Tools



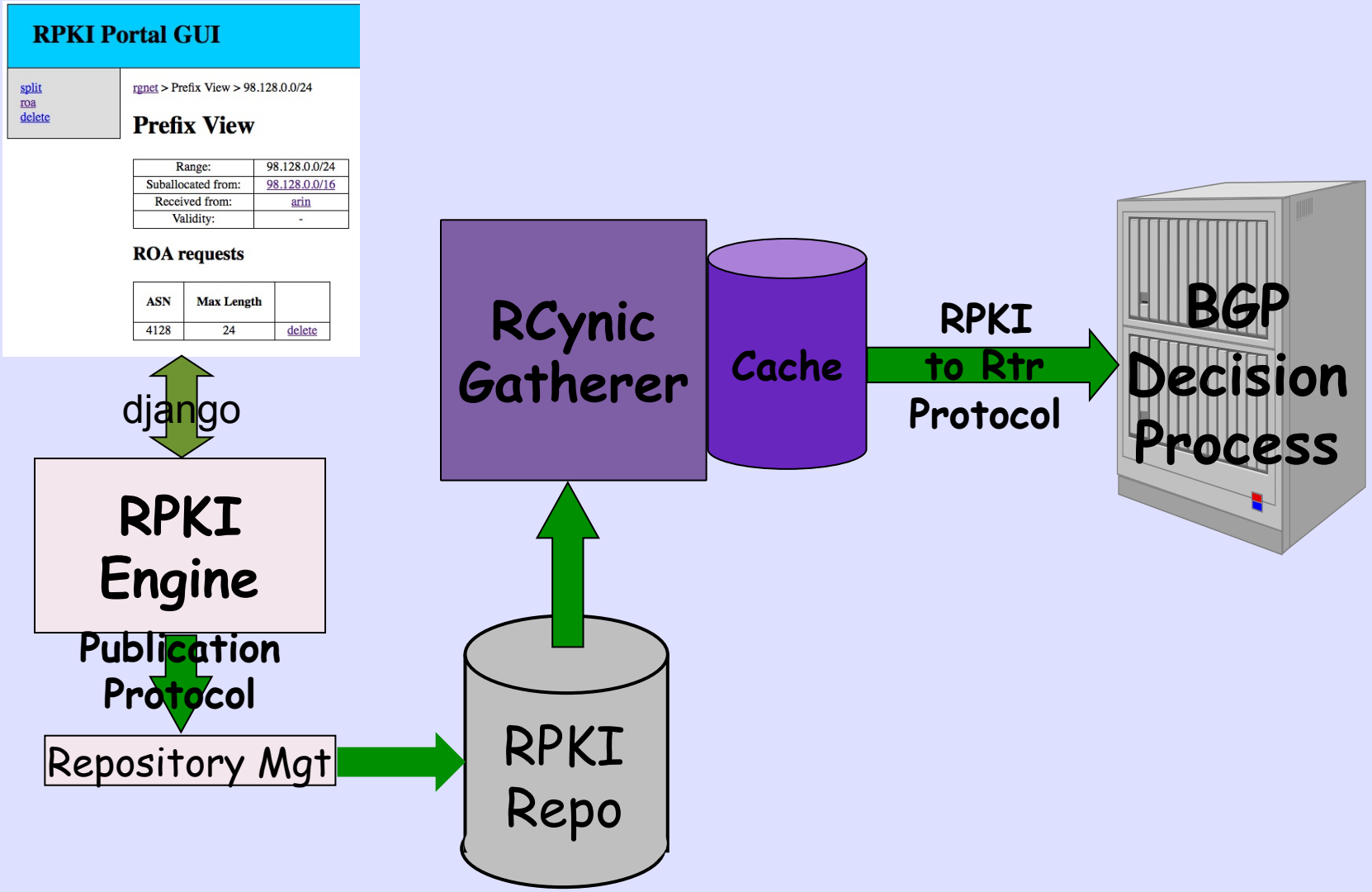
RPSL Your WorkFLow?

```
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 20110606
source:   RPKI
```

CSV Your WorkFlow?

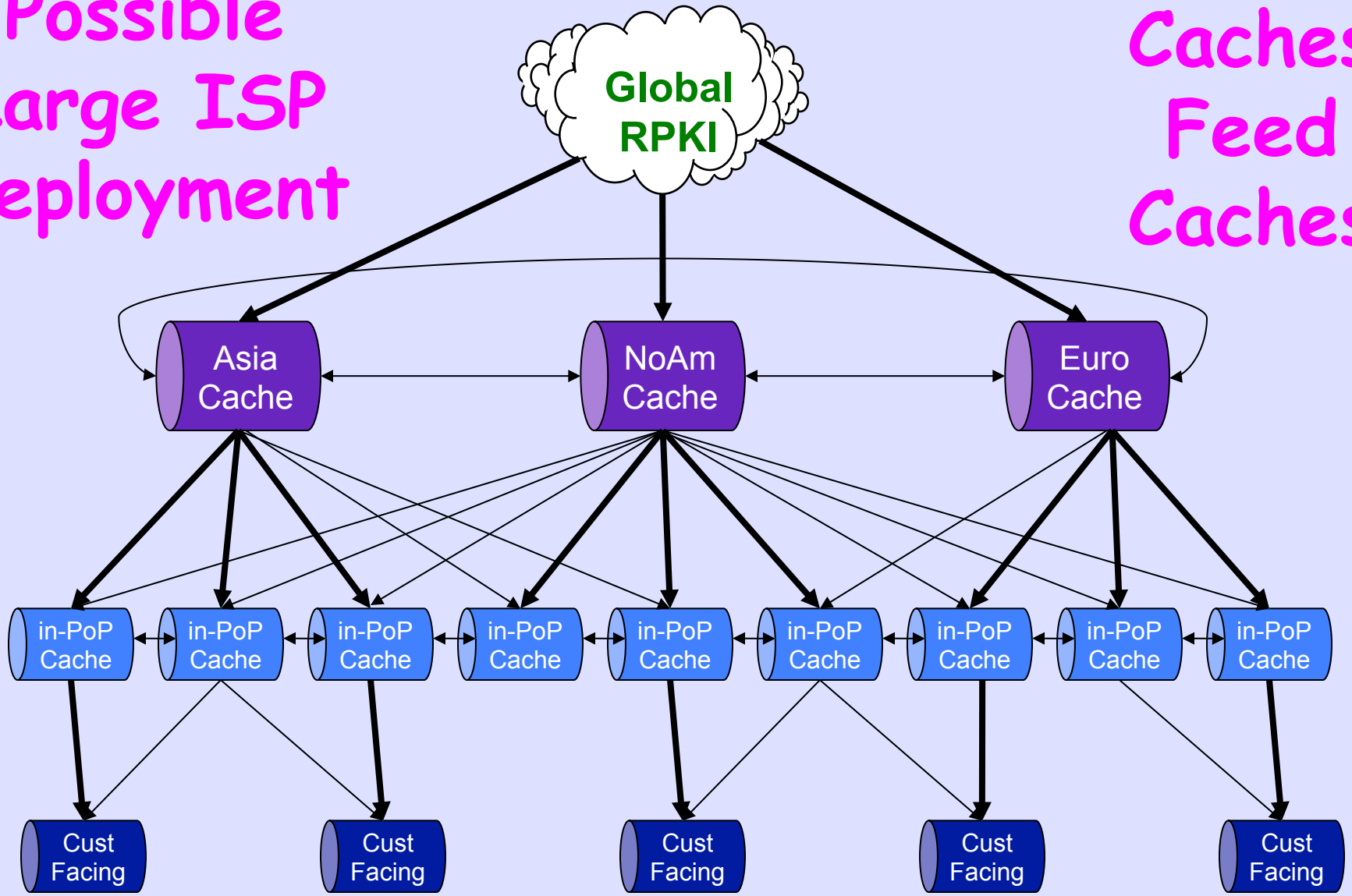
```
67.21.36.0/24    3970
192.169.0.0/23   3970
207.34.0.0/24    3970
216.21.0.0/24    3970
216.21.14.0/24   3970
216.21.16.0/24   3970
216.151.34.0/24 3970
147.28.0.0/16    3130
192.83.230.0/24 3130
```

RPKI-Rtr Protocol

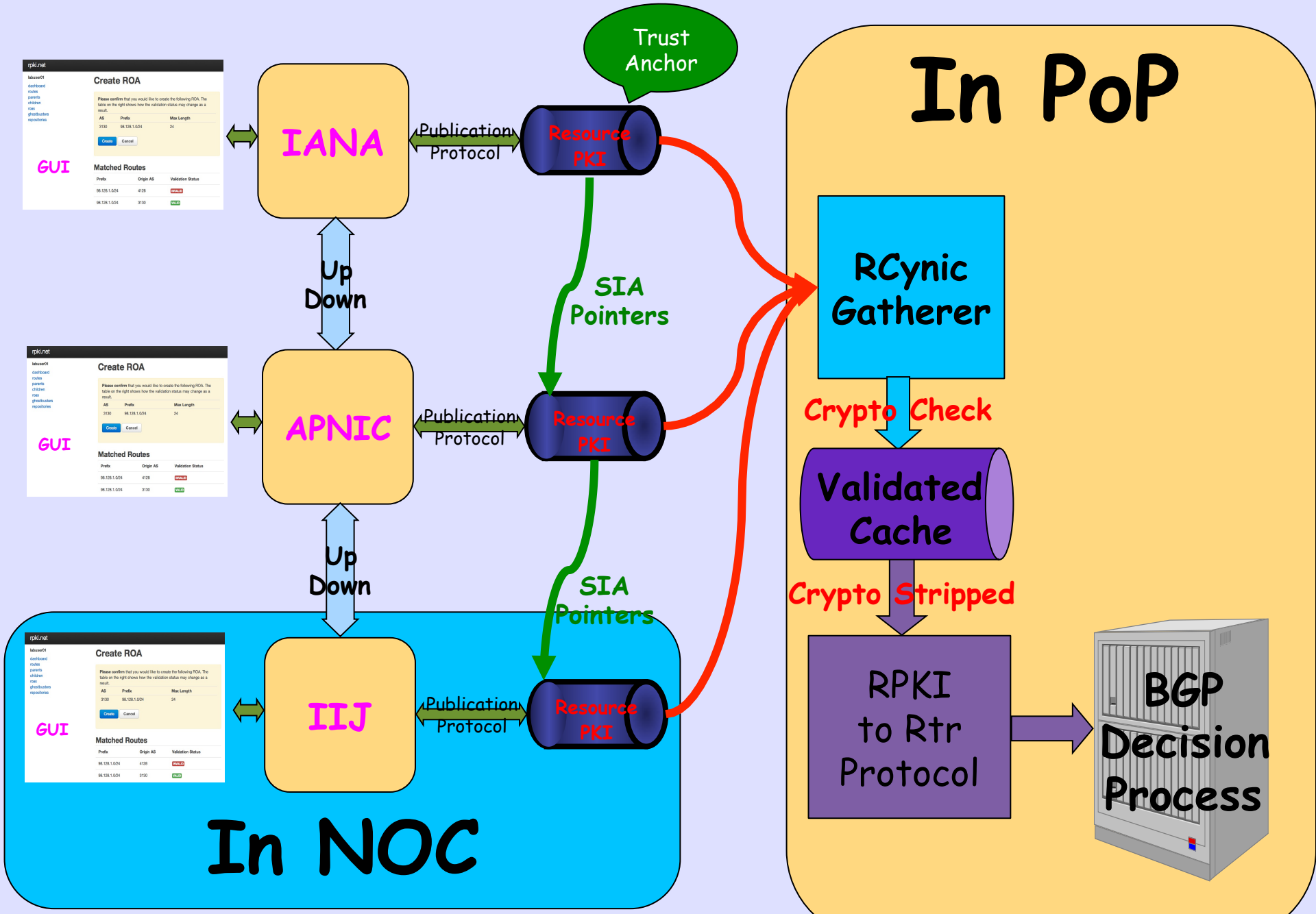


Possible Large ISP Deployment

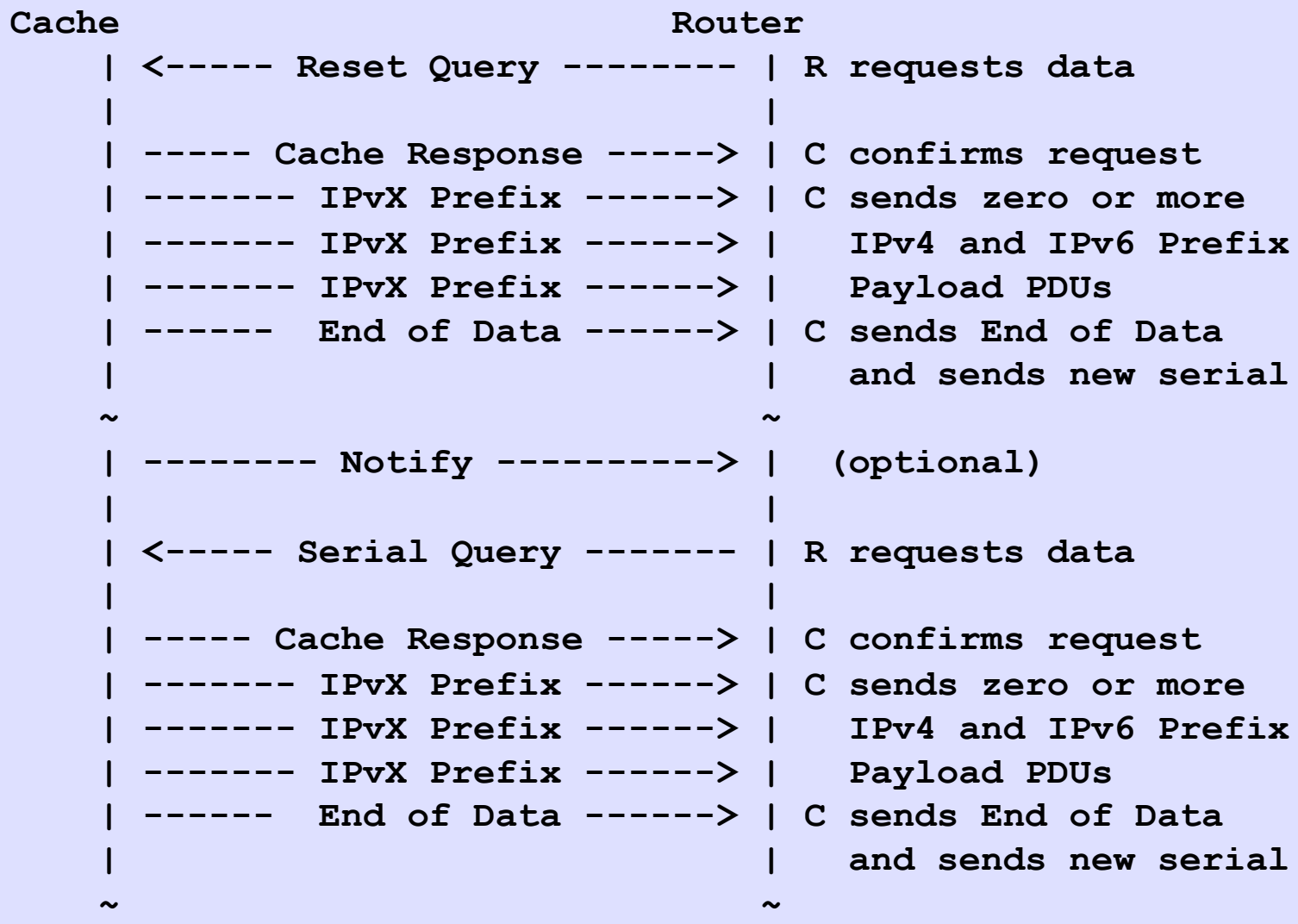
Caches Feed Caches



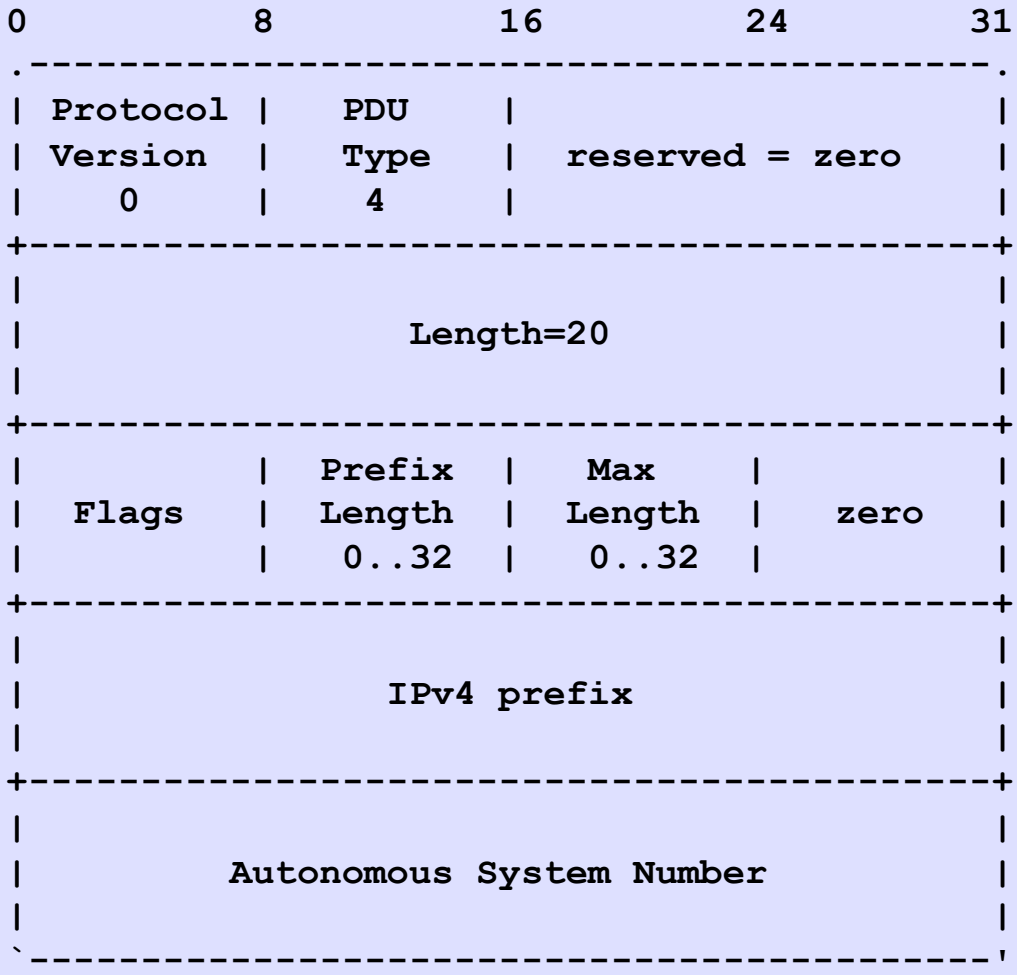
How Do ROAs Affect BGP Updates?


Trust Anchor
In PoP
IANA
APNIC
IIJ
GUI
Publication Protocol
Resource PKI
Up Down
SIA Pointers
RCynic Gatherer
Crypto Check
Validated Cache
Crypto Stripped
RPKI to Rtr Protocol
BGP Decision Process
In NOC

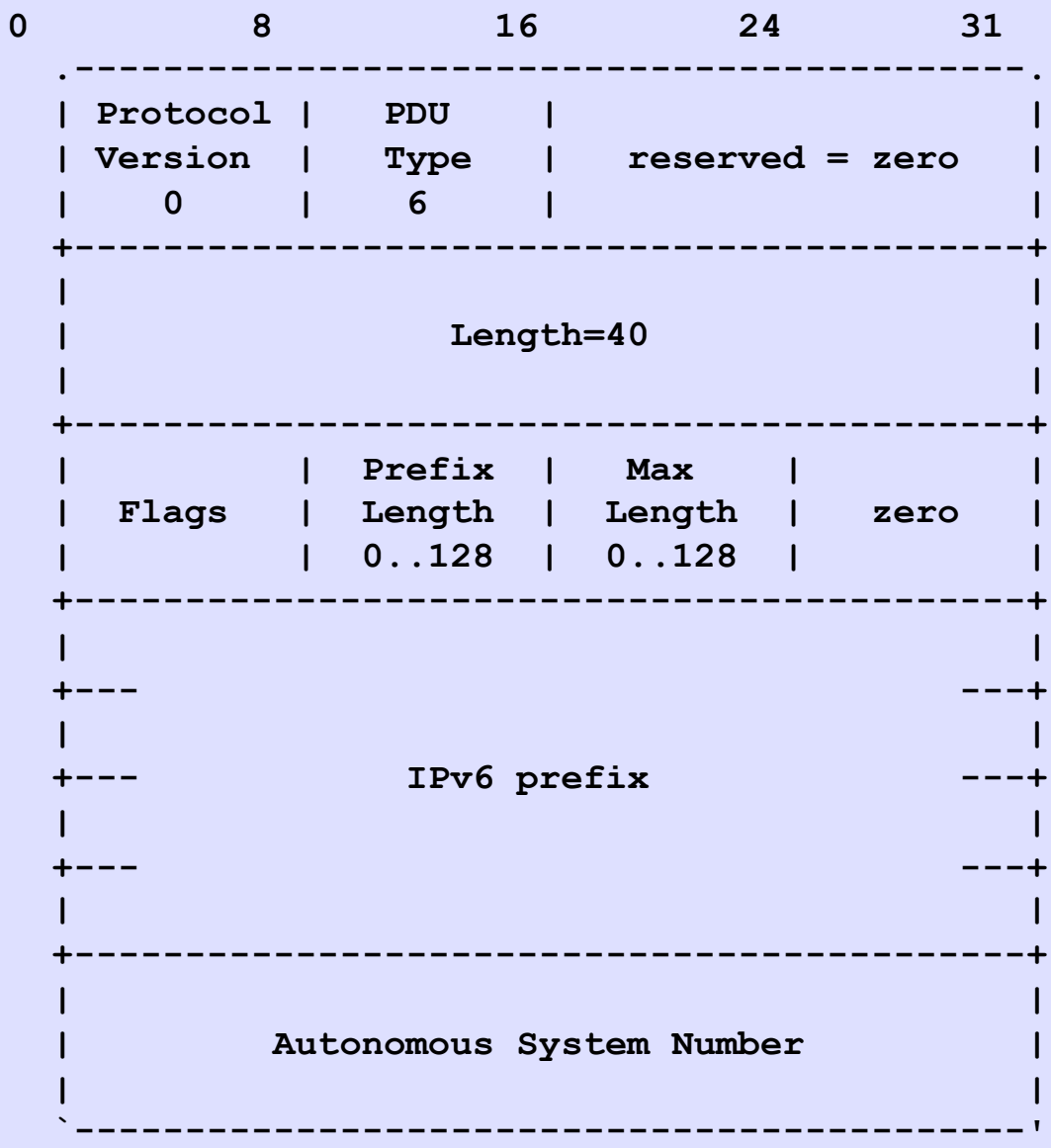
Typical Exchange



IPv4 Prefix

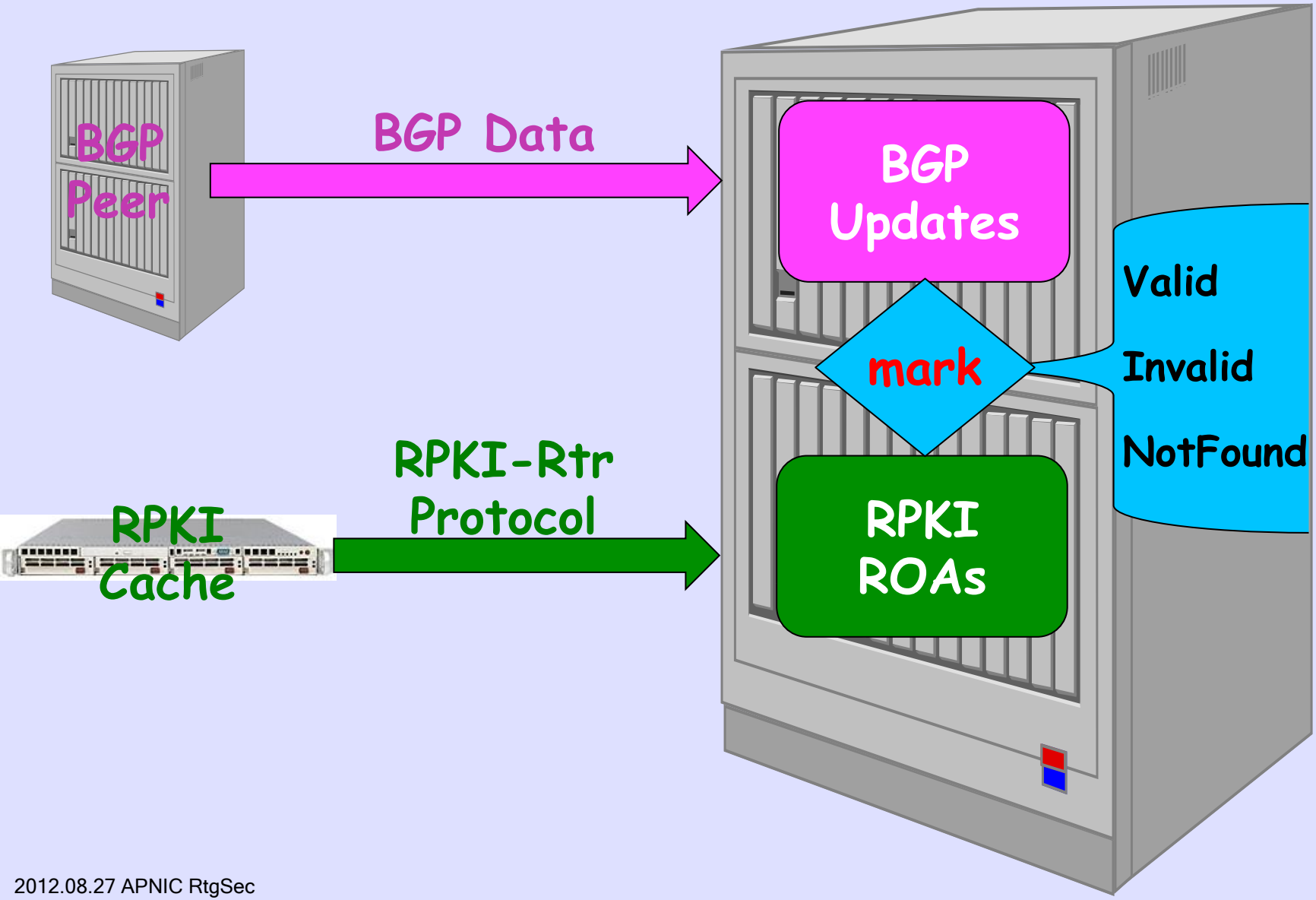

0
8
16
24
31
Protocol Version
0
PDU Type
4
reserved = zero
Length=20
Flags
Prefix Length
0..32
Max Length
0..32
zero
IPv4 prefix
Autonomous System Number

IPv6 Prefix


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

BGP Updates are compared with ROA Data loaded from the RPKI

Marking BGP Updates



Configure Router to Get ROAs

```
router bgp 3130
 …
 bgp rpki server tcp 198.180.150.1 port 42420 refresh 3600
 bgp rpki server tcp 147.28.0.35 port 93920 refresh 3600
 …
```

Check Server

```
r0.sea#show ip bgp rpki servers
BGP SOVC neighbor is 198.180.150.1/42420 connected to port 42420
Flags 0, Refresh time is 120, Serial number is 1304239609
InQ has 0 messages, OutQ has 0 messages, formatted msg 345
Session IO flags 3, Session flags 4008
 Neighbor Statistics:
  Nets Processed 624

Connection state is ESTAB, I/O status: 1, unread input bytes: 0
Connection is ECN Disabled
Mininum incoming TTL 0, Outgoing TTL 255
Local host: 199.238.113.10, Local port: 57932
Foreign host: 198.180.150.1, Foreign port: 42420
Connection tableid (VRF): 0
```

Look at Table

```
r0.sea#sh ip bgp rpki table
80 BGP sovc network entries using 7040 bytes of memory
86 BGP sovc record entries using 1720 bytes of memory
```

Network	Maxlen	Origin-AS	Neighbor
67.21.36.0/24	24	3970	198.180.150.1/42420
98.128.0.0/24	24	3130	198.180.150.1/42420
98.128.0.0/24	24	666	198.180.150.1/42420
98.128.0.0/16	16	3130	198.180.150.1/42420
98.128.3.0/24	24	3130	198.180.150.1/42420
98.128.4.0/24	24	3130	198.180.150.1/42420
98.128.5.0/24	24	3130	198.180.150.1/42420
98.128.6.0/24	24	3130	198.180.150.1/42420
98.128.7.0/24	24	65107	198.180.150.1/42420
98.128.9.0/24	24	3130	198.180.150.1/42420
98.128.10.0/24	24	3130	198.180.150.1/42420

Look at BGP Table

```
r0.sea#sh ip bgp
     Network          Next Hop            Metric LocPrf Weight Path
V*>  98.128.28.0/24   0.0.0.0                  0         32768 i
V*>  98.128.29.0/24   0.0.0.0                  0         32768 i
V*>  98.128.30.0/24   0.0.0.0                  0         32768 i
V*>  98.128.31.0/24   0.0.0.0                  0         32768 i
N*>  98.129.0.0/16    199.238.113.9           62             0 2914 12179 33070 33070 i
N*                    129.250.11.41           67             0 2914 12179 33070 33070 i
V*>i 98.130.0.0/16    206.81.80.40           789     90      0 6939 32392 i
N*                    199.238.113.9           65             0 2914 4436 32392 i
N*                    129.250.11.41           70             0 2914 4436 32392 i
I*>i 98.130.0.0/15    206.81.80.40           789     90      0 6939 32392 i
N*                    199.238.113.9           65             0 2914 4436 32392 i
```

Look at a Prefix

```
R3#show ip bgp 98.128.0.0/24
BGP routing table entry for 98.128.0.0/24, version 360
Paths: (2 available, best #1, table default)
 65000 3130
    10.0.0.1 from 10.0.0.1 (193.0.24.64)
      Origin IGP, localpref 100, valid, external, best
      path 680D859C RPKI State valid
  65001 4128
    10.0.1.1 from 10.0.1.1 (193.0.24.65)
      Origin IGP, localpref 100, valid, external
      path 680D914C RPKI State invalid
```

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, same as today

Valid!

```
r0.sea#show bgp 192.158.248.0/24
BGP routing table entry for 192.158.248.0/24, version 3043542
Paths: (3 available, best #1, table default)
 6939 27318
    206.81.80.40 (metric 1) from 147.28.7.2 (147.28.7.2)
      Origin IGP, metric 319, localpref 100, valid, internal,
best
      Community: 3130:391
      path 0F6D8B74 RPKI State valid
 2914 4459 27318
    199.238.113.9 from 199.238.113.9 (129.250.0.19)
      Origin IGP, metric 43, localpref 100, valid, external
      Community: 2914:410 2914:1005 2914:3000 3130:380
      path 09AF35CC RPKI State valid
```

Invalid!

```
r0.sea#show bgp 198.180.150.0
BGP routing table entry for 198.180.150.0/24, version 2546236
Paths: (3 available, best #2, table default)
  Advertised to update-groups:
     2           5           6           8
  Refresh Epoch 1
  1239 3927
    144.232.9.61 (metric 11) from 147.28.7.2 (147.28.7.2)
      Origin IGP, metric 759, localpref 100, valid, internal
      Community: 3130:370
      path 1312CA90 RPKI State invalid
```

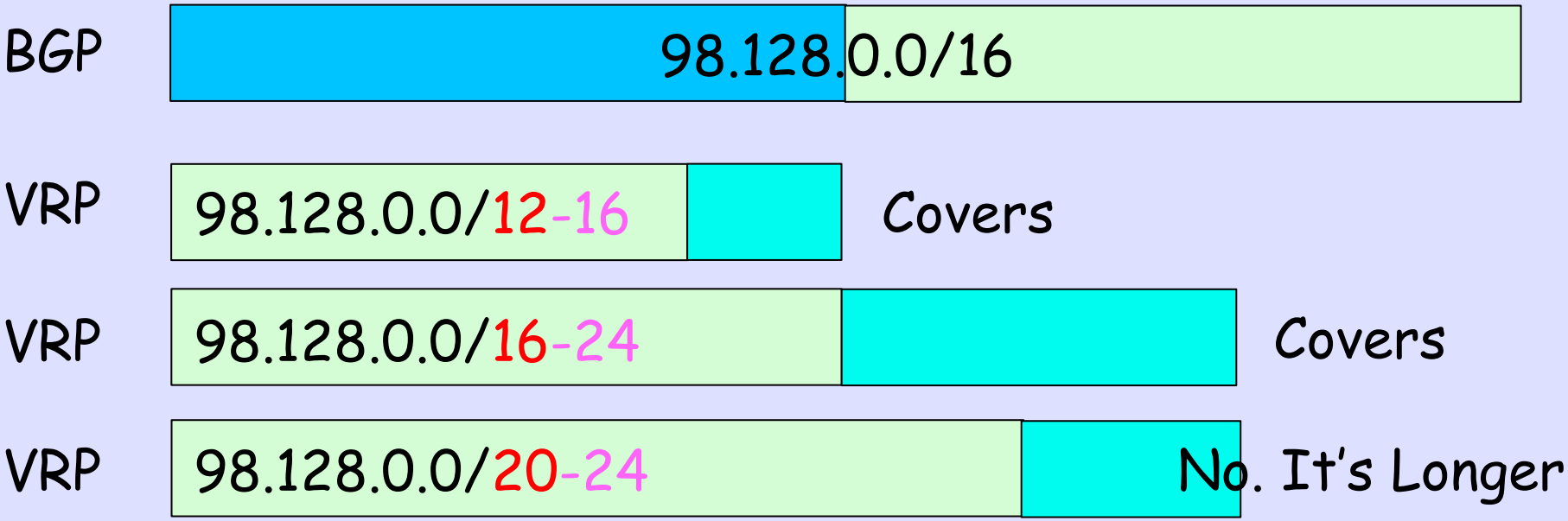
NotFound

```
r0.sea#show bgp 64.9.224.0
BGP routing table entry for 64.9.224.0/20, version 35201
Paths: (3 available, best #2, table default)
  Advertised to update-groups:
     2           5           6
  Refresh Epoch 1
  1239 3356 36492
    144.232.9.61 (metric 11) from 147.28.7.2 (147.28.7.2)
      Origin IGP, metric 4, localpref 100, valid, internal
      Community: 3130:370
      path 11861AA4 RPKI State not found
```

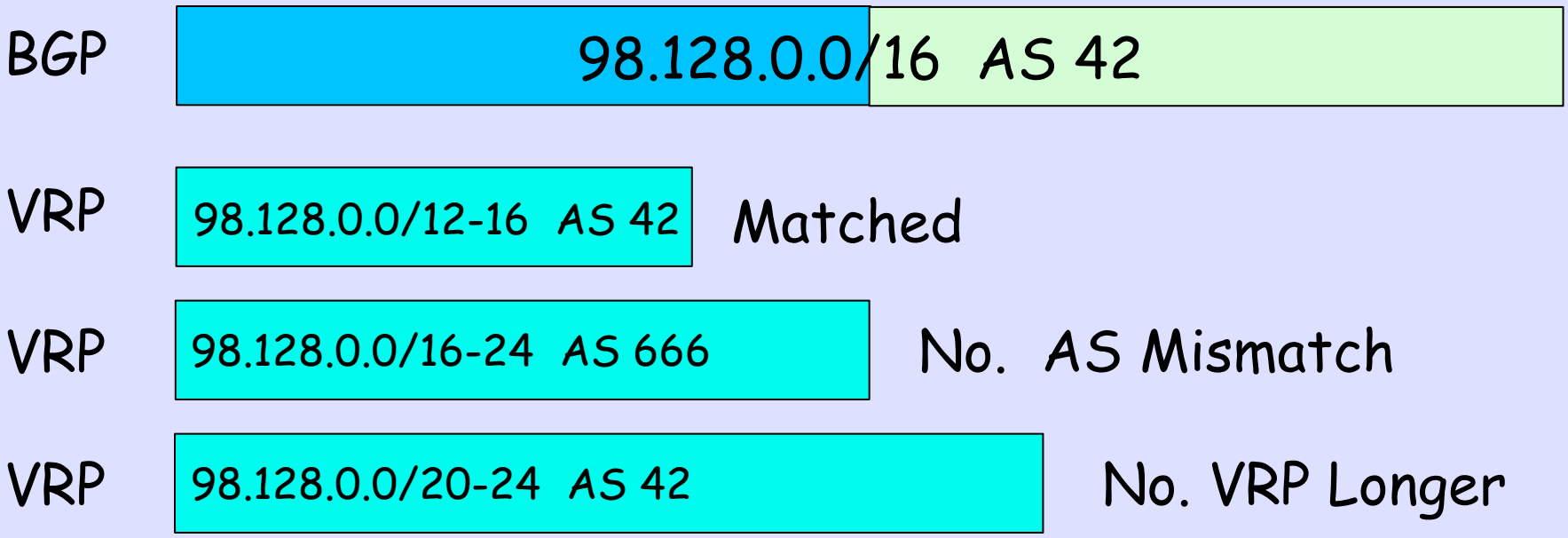
What are the BGP / VRP[1] Matching Rules?

[1] Validated ROA Payload

A Prefix is Covered by a VRP when the VRP prefix length is less than or equal to the Route prefix length



Prefix is **Matched** by a VRP when the Prefix is Covered by that VRP , prefix length is less than or equal to the VRP max-len, and the Route Origin AS is equal to the VRP's AS

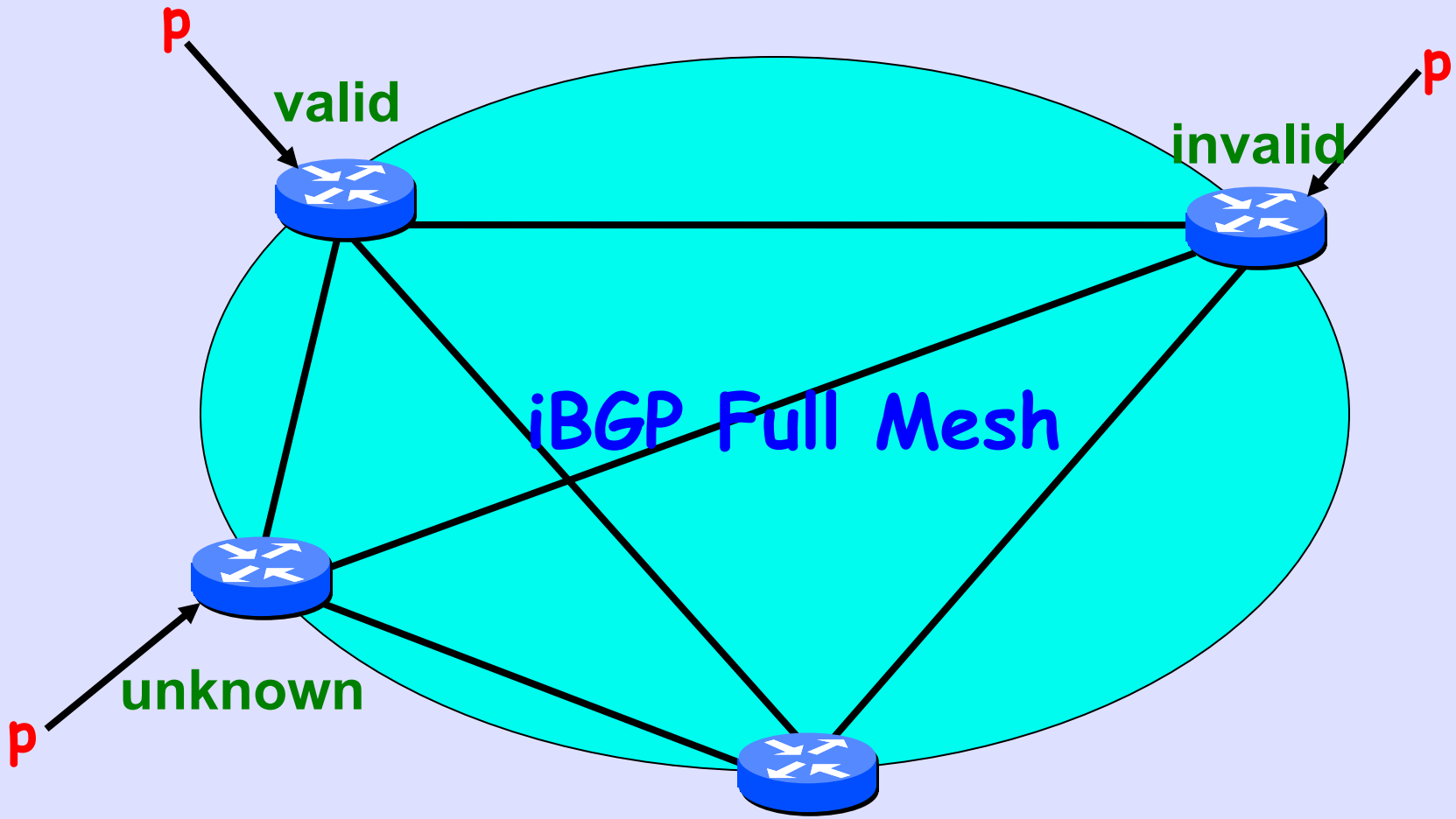


Matching and Validity

VRP_0	98.128.0.0/16-24 AS 6
VRP_1	98.128.0.0/16-20 AS 42

BGP	98.128.0.0/12	AS 42	NotFound, shorter than VRPs
BGP	98.128.0.0/16	AS 42	Valid, Matches VRP_1
BGP	98.128.0.0/20	AS 42	Valid, Matches VRP_1
BGP	98.128.0.0/24	AS 42	Invalid, longer than VRP with AS 42
BGP	98.128.0.0/24	AS 6	Valid, Matches VRP_0

iBGP Hides Validity State



which do i choose?
why do i choose it?

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

Fairly Secure

```
route-map validity permit 10
  match rpki valid
  set local-preference 100
route-map validity permit 20
  match rpki not-found
  set local-preference 50
! invalid is dropped
```

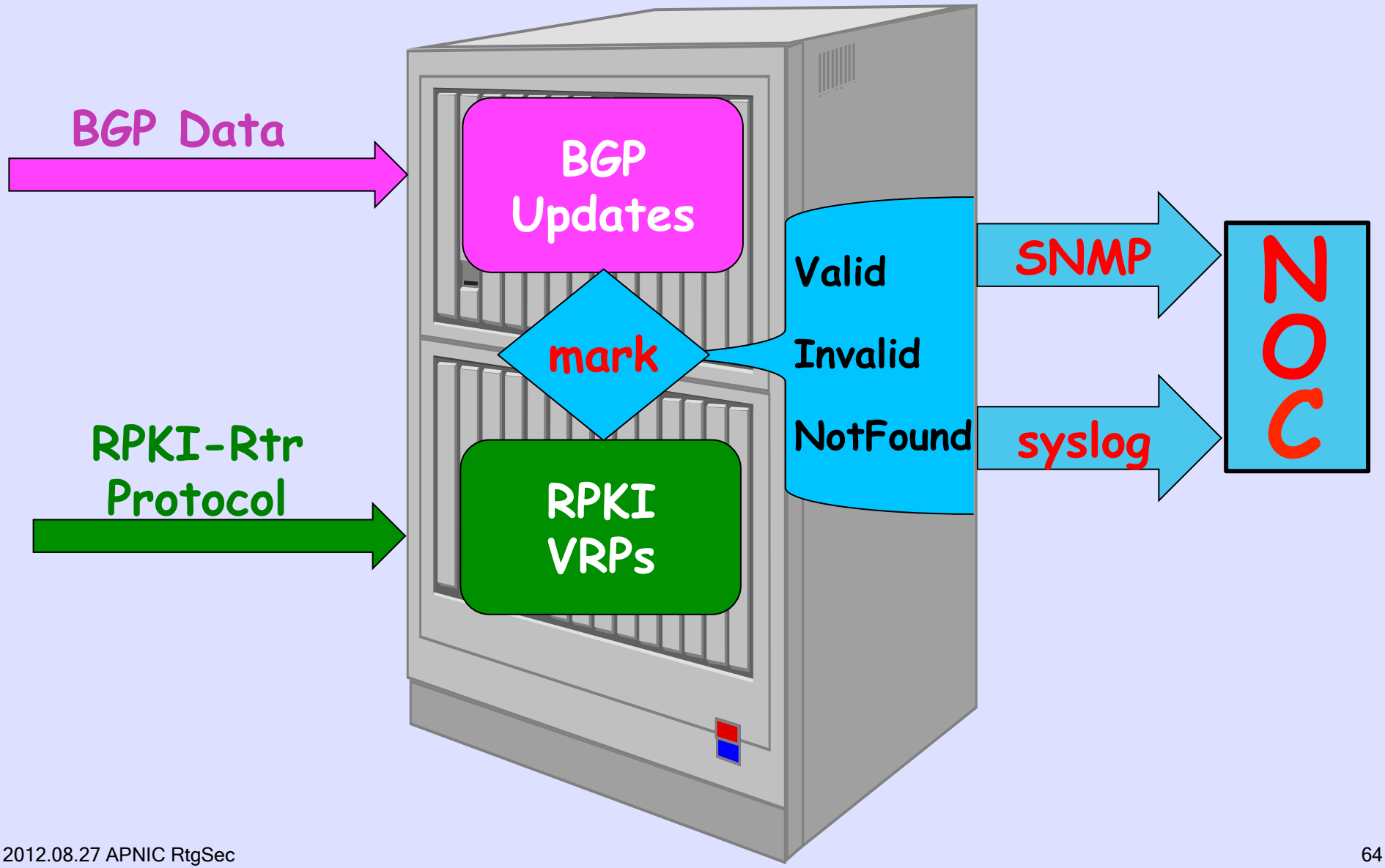
Paranoid

```
route-map validity permit 42
   match rpki valid
   set local-preference 110
! everything else dropped
```

Set a Community

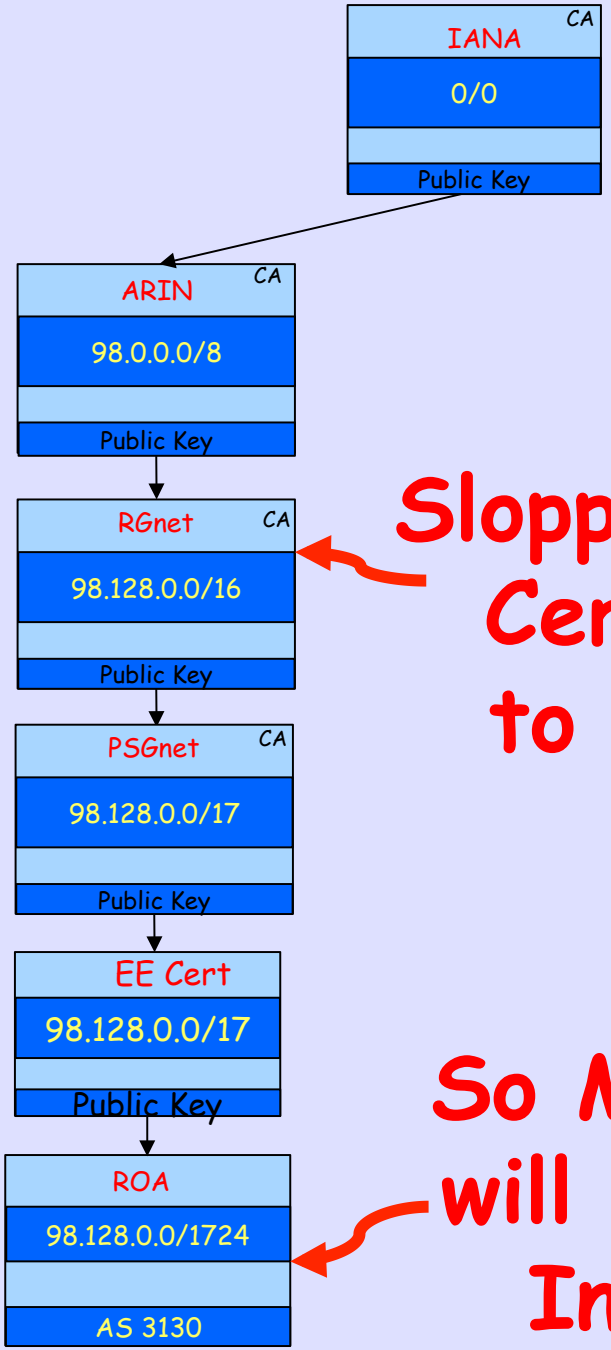
```
route-map validity permit 10
  match rpki valid
  set community 3130:400
route-map validity permit 20
  match rpki invalid
  set community 3130:200
route-map validity permit 30
  set community 3130:300
```

And it is All Monitored



The Big Speedbump

Up-Chain Expiration



These are not Identity Certs

Sloppy Admin, Cert Soon to Expire!

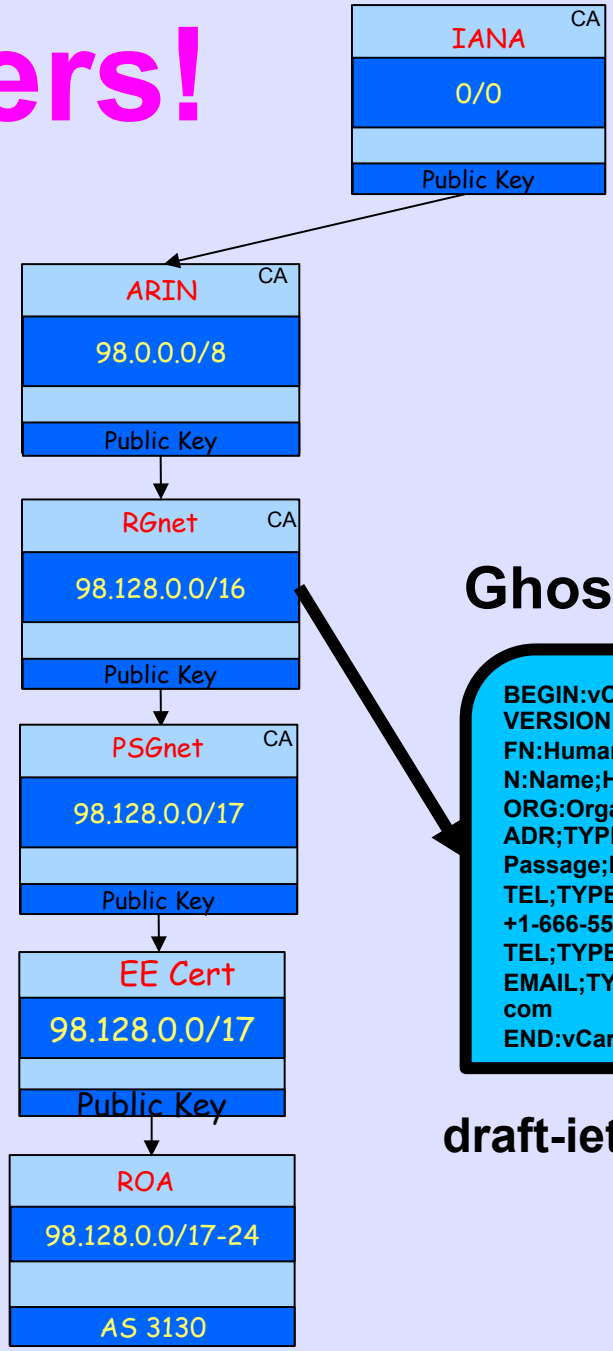
So My ROA will become Invalid!

ROA Invalid but I Can Route

- The ROA will become Invalid
- My announcement will just become NotFound, not Invalid
- Unless my upstream has a ROA for the covering prefix, which is likely

So Who You Gonna Call?

Ghostbusters!



Ghostbusters Record

BEGIN:vCard
VERSION:3.0
FN:Human's Name
N:Name;Human's;Ms.;Dr.;OCD;ADD
ORG:Organizational Entity
ADR;TYPE=WORK:;;42 Twisty
Passage;Deep Cavern; WA; 98666;U.S.A.
TEL;TYPE=VOICE,MSG,WORK:
+1-666-555-1212
TEL;TYPE=FAX,WORK:+1-666-555-1213
EMAIL;TYPE=INTERNET:human@example.
com
END:vCard

draft-ietf-sidr-ghostbusters

But in the End, You Control Your Policy

"Announcements with Invalid origins MAY be used, but SHOULD be less preferred than those with Valid or NotFound."

-- draft-ietf-sidr-origin-ops

But if I do not reject Invalid, what is all this for?

Open Source (BSD Lisc) Running Code

https://rpki.net/

Shipping Router Code

Talk to C & J

BGPsec AS-Path Validation

Future Work

Origin Validation is Weak

- RPKI-Based Origin Validation only stops accidental misconfiguration, which is very useful. But ...
- A malicious router may announce as any AS, i.e. forge the ROAed origin AS.
- This would pass Origin ROA Validation.

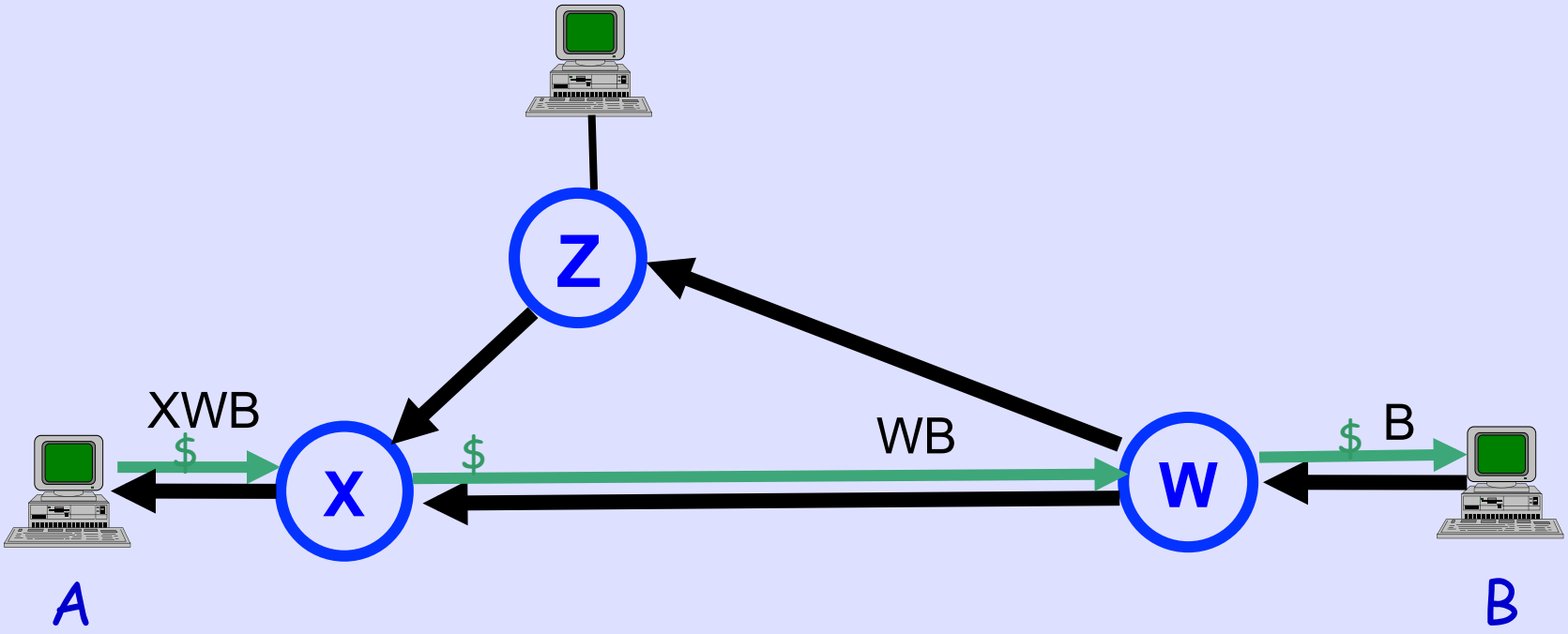
Full Path Validation

- Rigorous per-prefix AS path validation is the goal
- Protect against origin forgery and AS-Path monkey in the middle attacks
- Not merely showing that a received AS path is not impossible

Protocol Not Policy

- We can not know intent, **should** Mary have announced the prefix to Bob
- But Joe can formally validate that Mary **did** announce the prefix to Bob
- Policy on the global Internet changes every 36ms, new peers, new customers, new circuits, etc.
- We already have a protocol to distribute policy or its effects, it is called BGP
- BGPsec validates that the protocol has not been violated, and is not about intent or business policy

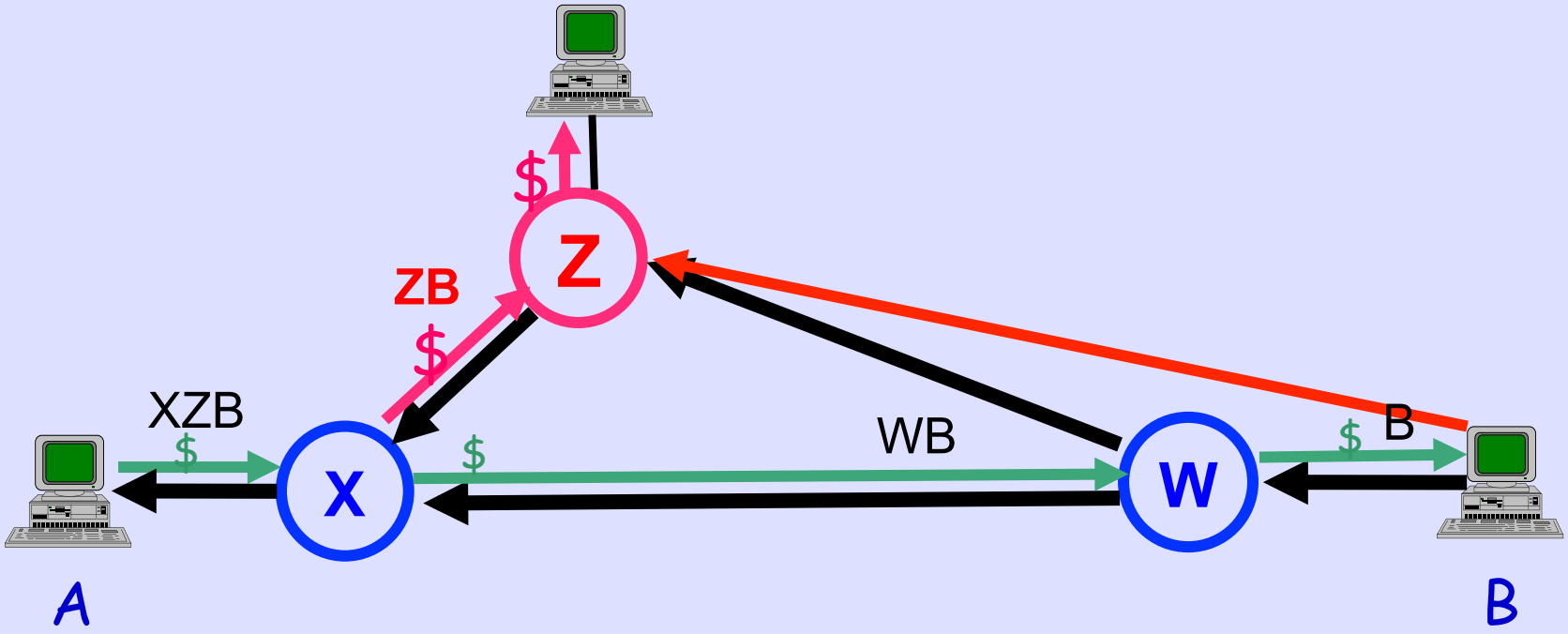
Our Parents' Internet



Routing Announcements

Packet Data Flows

Path Shortening Attack



Expected Path - A->X->W->B

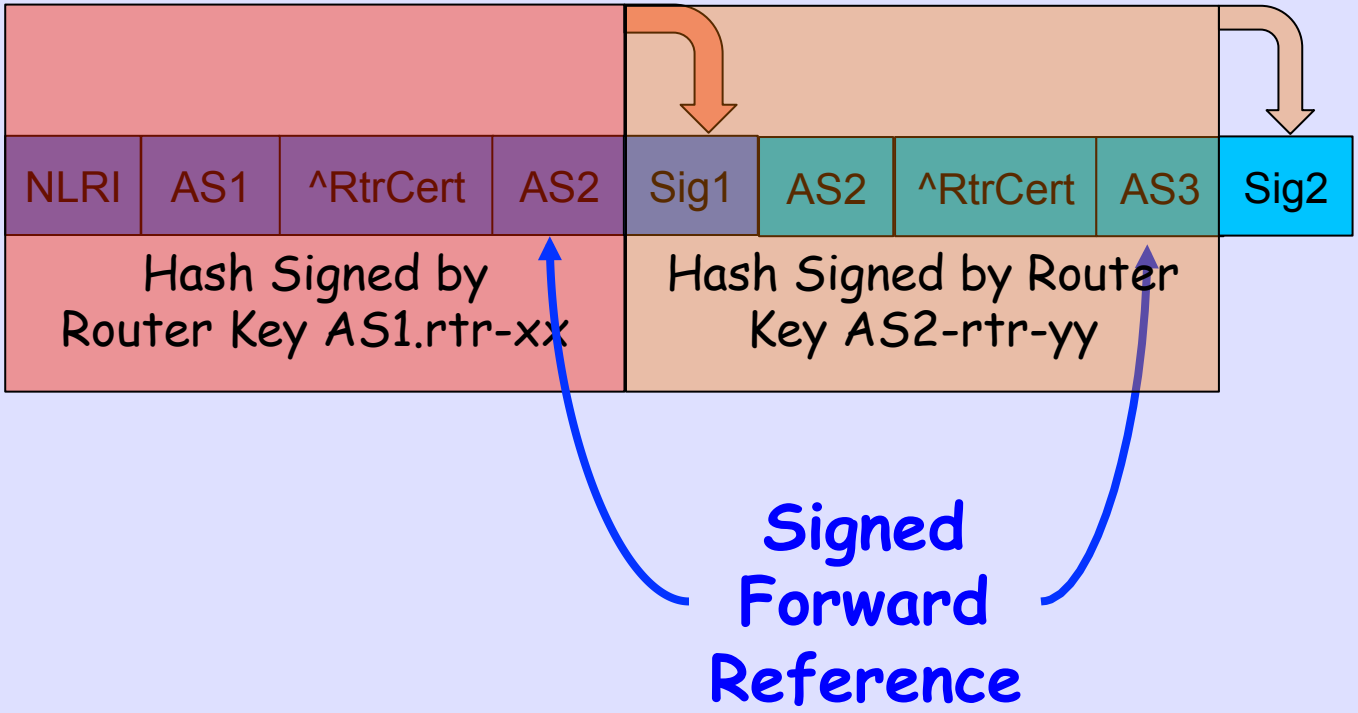
Diverted Path - A->X->Z->W->B

There Are Many Many Other Attacks

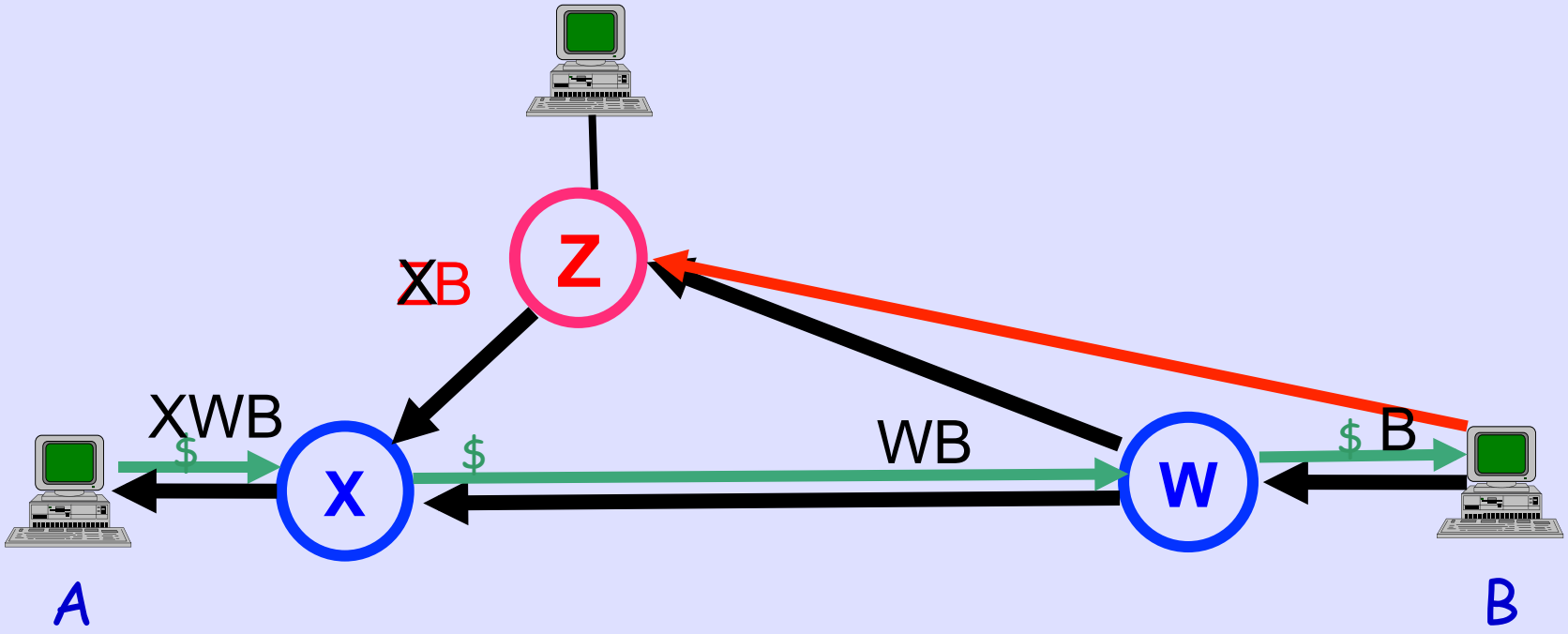
Forward Path Signing

AS hop N signing (among other things) that it is sending the announcement to AS hop N+1 by AS number, is believed to be fundamental to protecting against monkey in the middle attacks

Forward Path Signing



Forward-Signing



B cryptographically signs the message to W **Sb(B->W)**

W signs messages to X and Z encapsulating B's message

Sw(W->X (**Sb(B->W)**)) and Sw(W->Z (**Sb(B->W)**))

X signs the message to A Sx(X->A (Sw(W->X (**Sb(B->W)**))))

Z can only sign Sz(Z->X (Sw(W->Z (**Sb(B->W)**))))

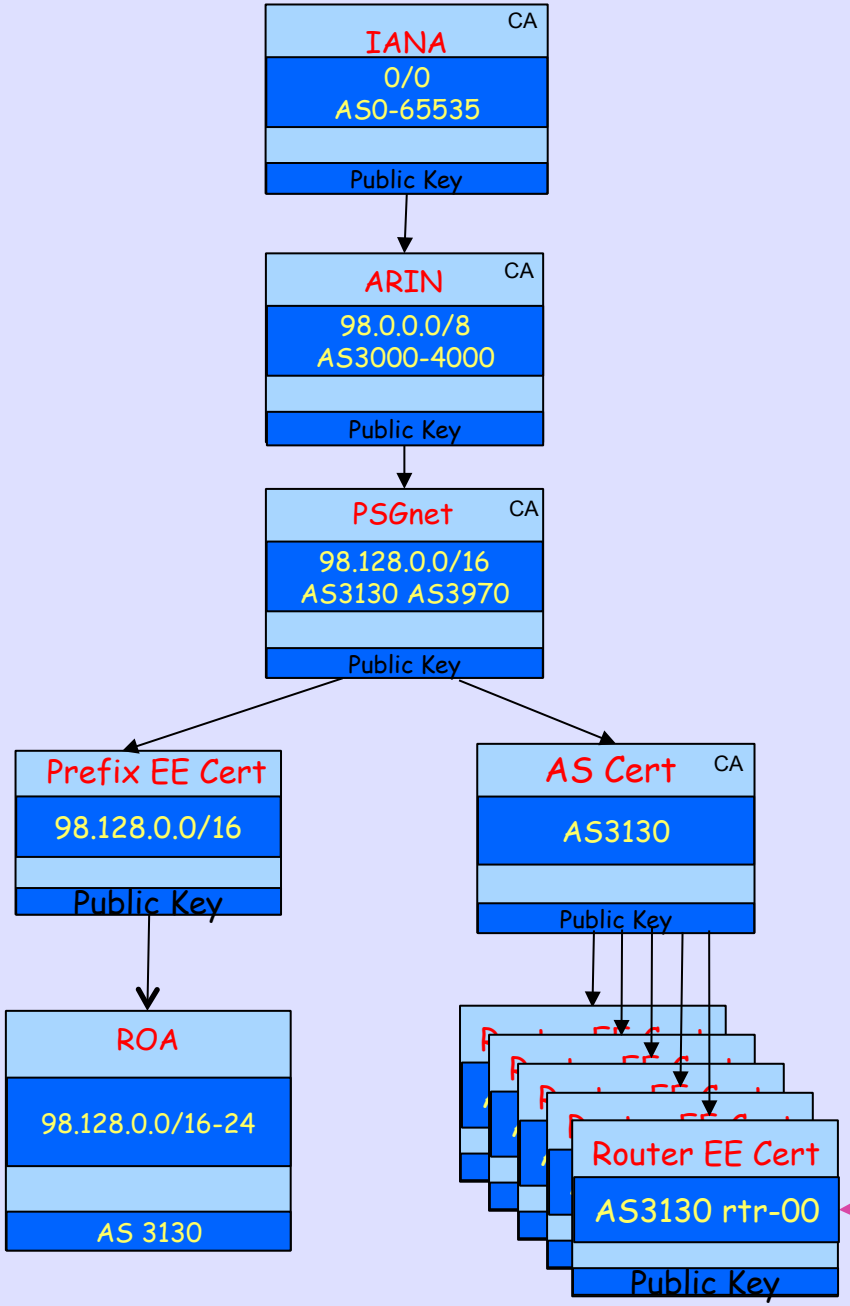
Capability Negotiation

- It is assumed that consenting routers will use BGP capability exchange to agree to run BGPsec between them
- The capability will, among other things remove the 4096 PDU limit for updates
- If BGPsec capability is not agreed, then only traditional BGP data are sent

Per-Router Keys

- Needed to deal with compromise of one router exposing an AS's private key
- Implies a more complex certificate and key distribution mechanism
- A router could generate key pair and send certificate request to RPKI for signing
- Certificate, or reference to it, must be in each signed path element
- If you want one per-AS key, share a router key

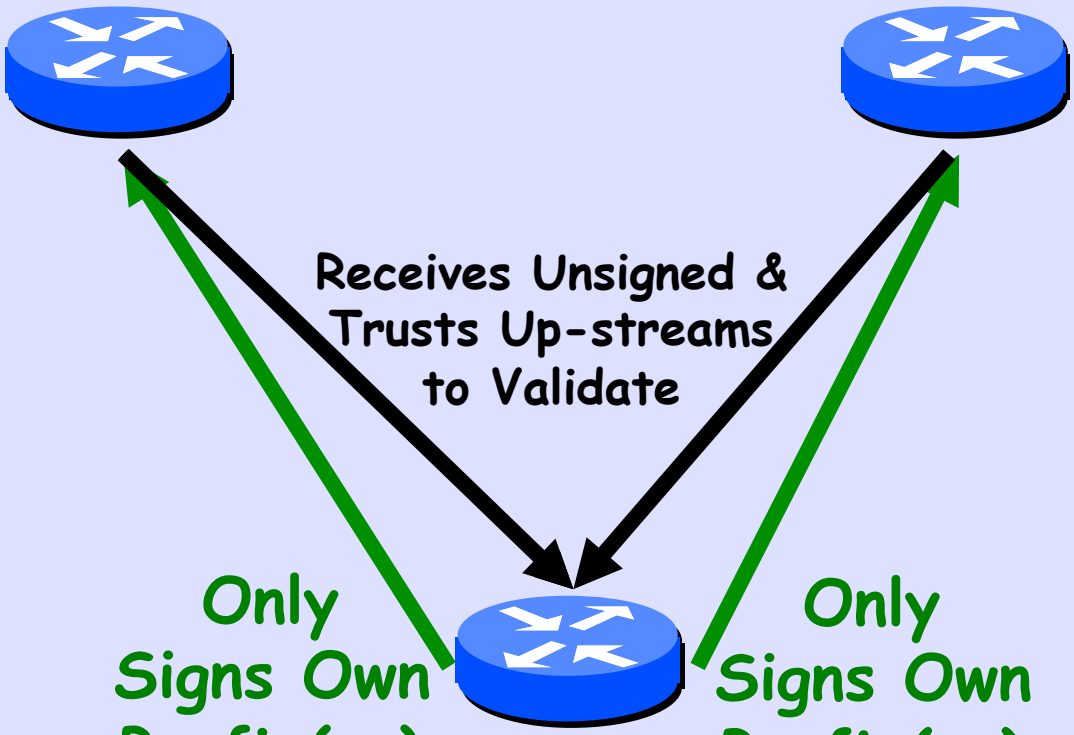
Cert / Key Structure for an ISP



Only at Provider Edges

- This design protects only inter-domain routing, not IGPs, not even iBGP
- BGPsec will be used inter-provider, only at the providers' edges
- Of course, the provider's iBGP will have to carry the BGPsec information
- Providers and inter-provider peerings might be heterogeneous

Simplex End Site



Very few signatures! No verification
Only Needs to Have Own Private Key
No Other Crypto or RPKI Data
No Hardware Upgrade!!

Incremental Deployment

Incrementally deployable in today's Internet, and does not require global deployment, a flag day, etc.

No Increase of Operator Data Exposure

- Operators wish to minimize any increase in visibility of information about peering and customer relationships etc.
- No IRR-style publication of customer or peering relationships is needed

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). **[0]**

  [0] – they Take your Scissors Away and we turn them into plowshares

- **ARIN**
- **Internet Initiative Japan & ISC**
- **Cisco, Juniper, Google, NTT, Equinix**