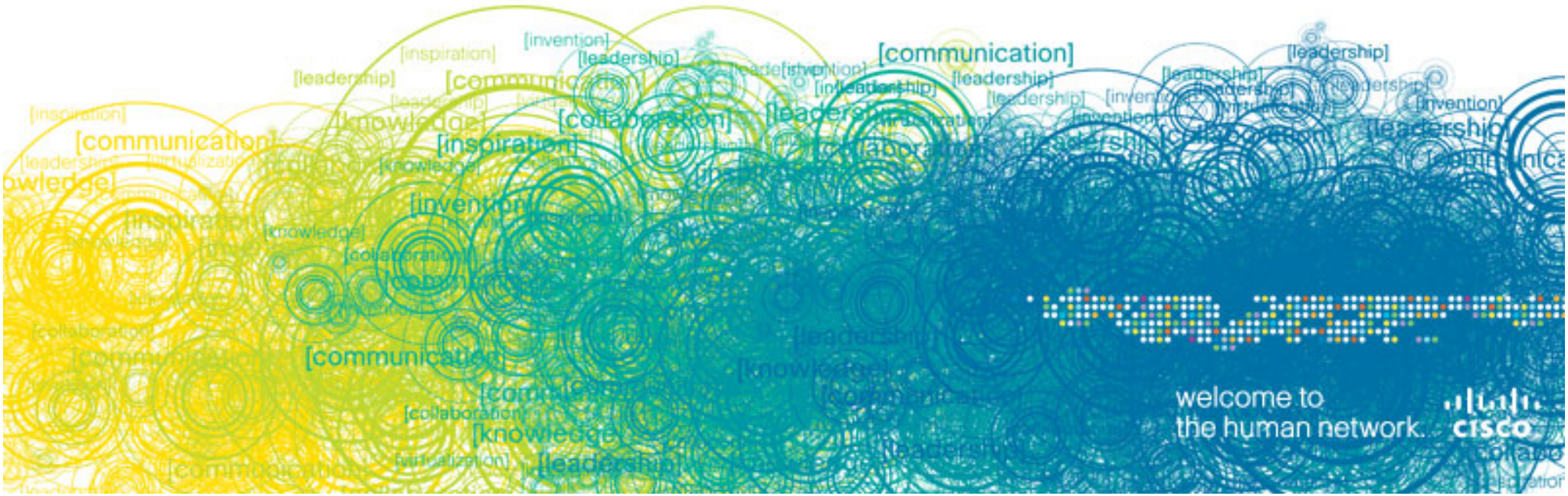
BGP Prefix Origin Validation

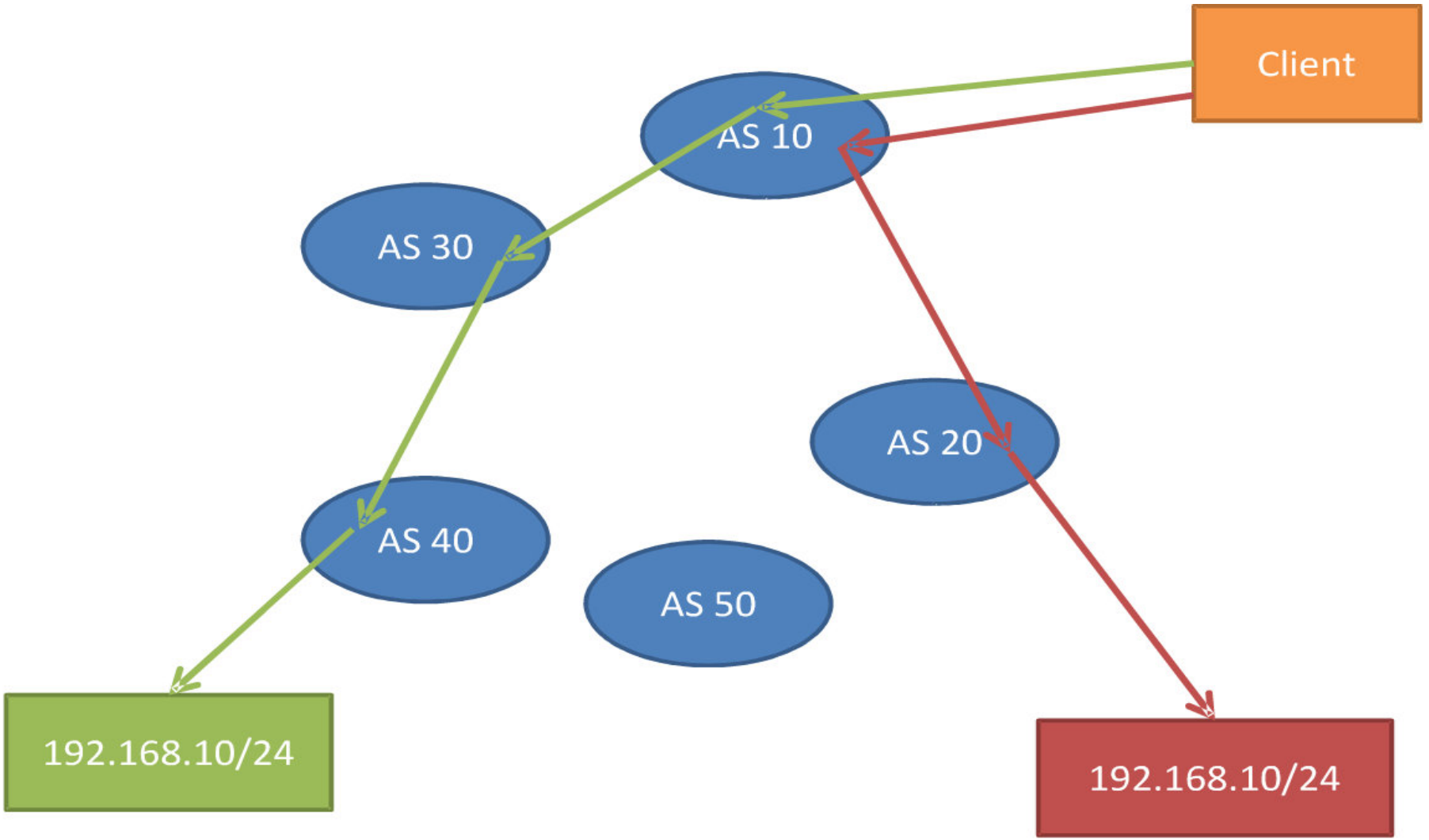
Pradosh Mohapatra



Motivation

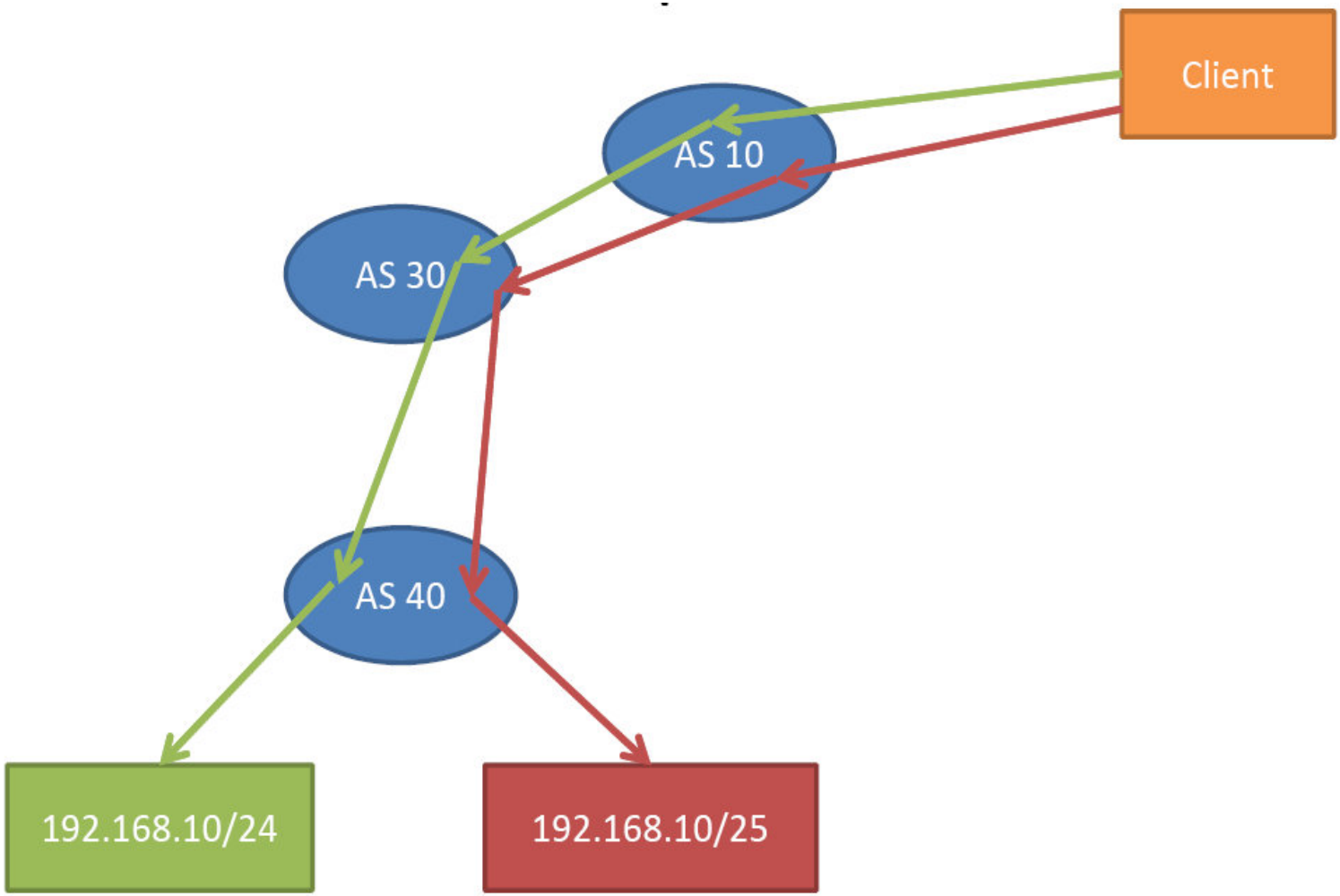
- Any AS can inject any prefix in BGP – prefix hijacking
  - Mistake (most likely)
  - Malicious (could be!)
- Hijacking manifestation
  - Announcing someone else's prefix
  - Announcing a more specific of someone else's prefix
- Some real-life incidents: http://www.networkworld.com/news/2009/011509-bgp-attacks.html
- Need a mechanism to differentiate between invalid and legit routes for a BGP destination

Same prefix: shorter AS_PATH wins



Source: nanog 46 preso

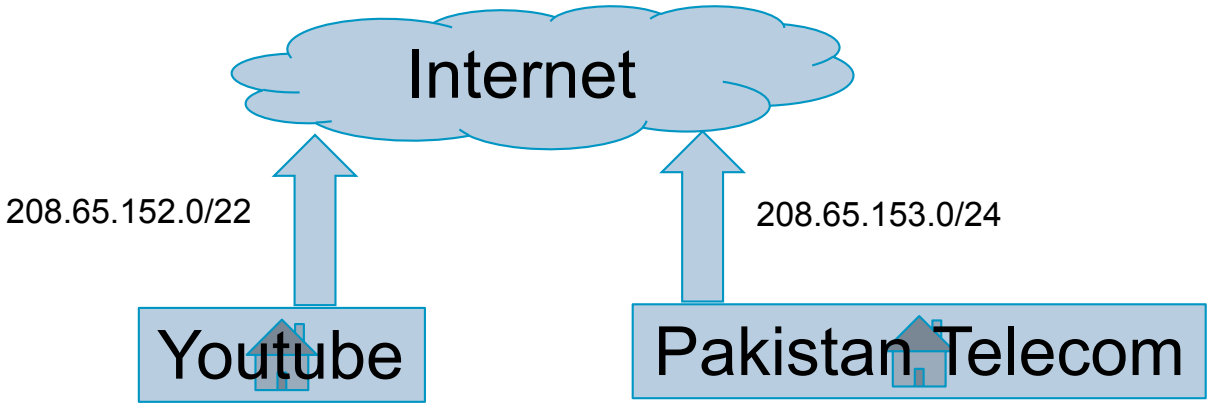
Same prefix: More specific wins



Source: nanog 46 preso

Youtube hijacking example

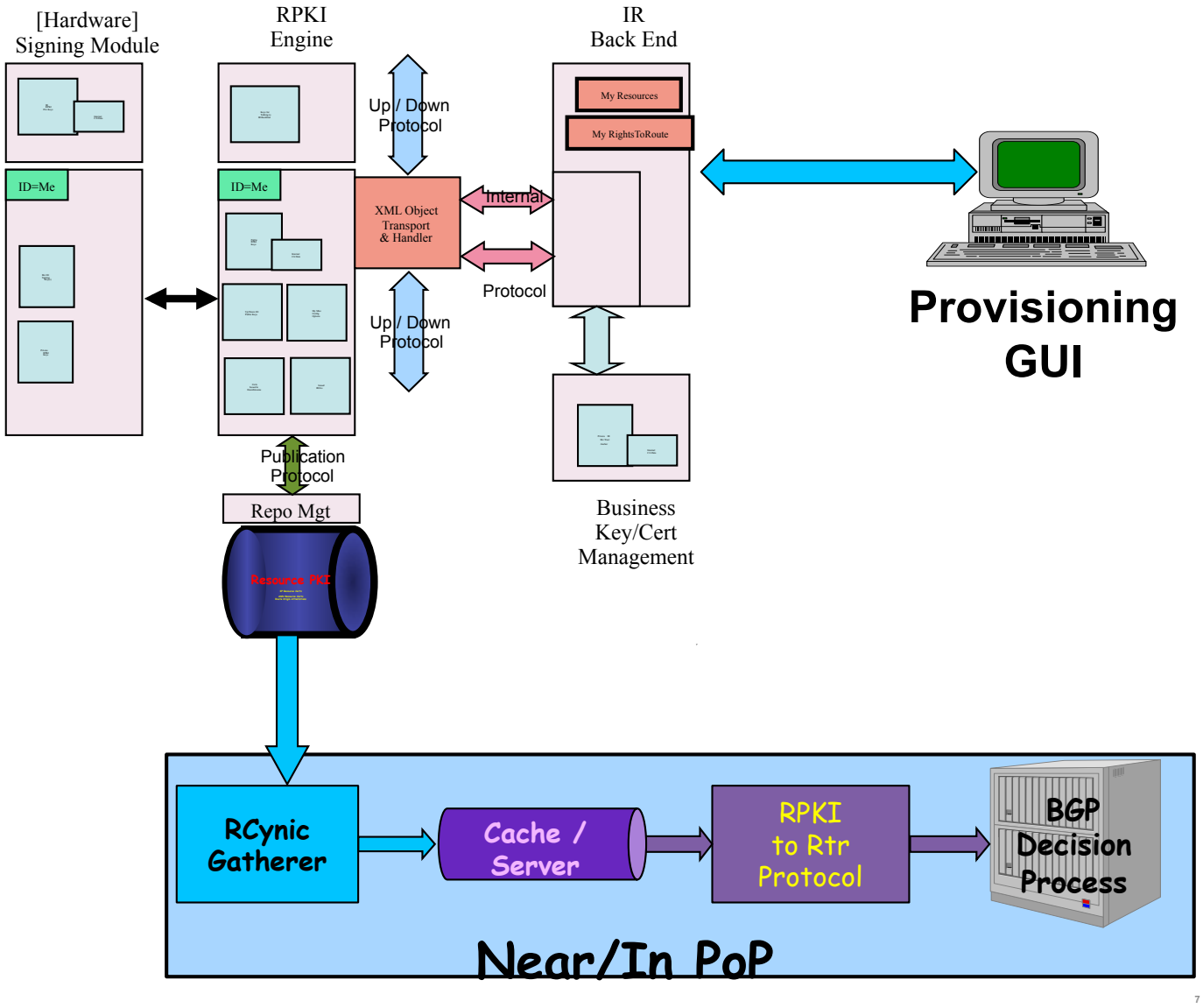
- An example of more specific hijacking stemming from misconfiguration …



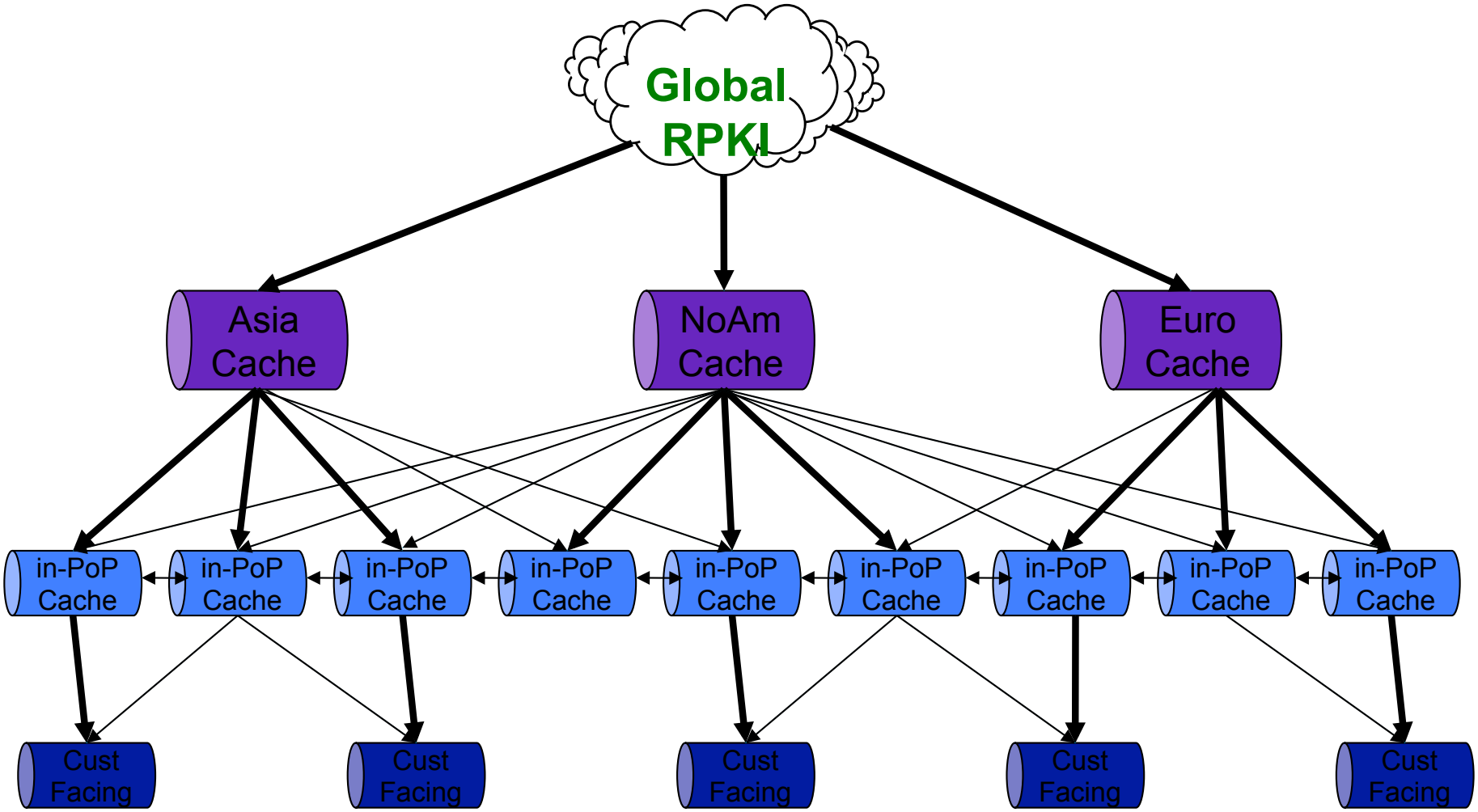
Origin validation framework - components

- RPKI: passive side – object format
  - X.509 certificate with RFC3779 extensions for IP resources (IPAddress and ASN)
  - Route Origin Attestation (ROA) signed object
- RPKI: active side
  - Allocation hierarchy
  - Database maintenance
  - Transaction semantics, certificate checks, …
- Getting data to BGP speaking routers
- BGP operation for origin validation

Complete picture



Extremely Large ISP deployment

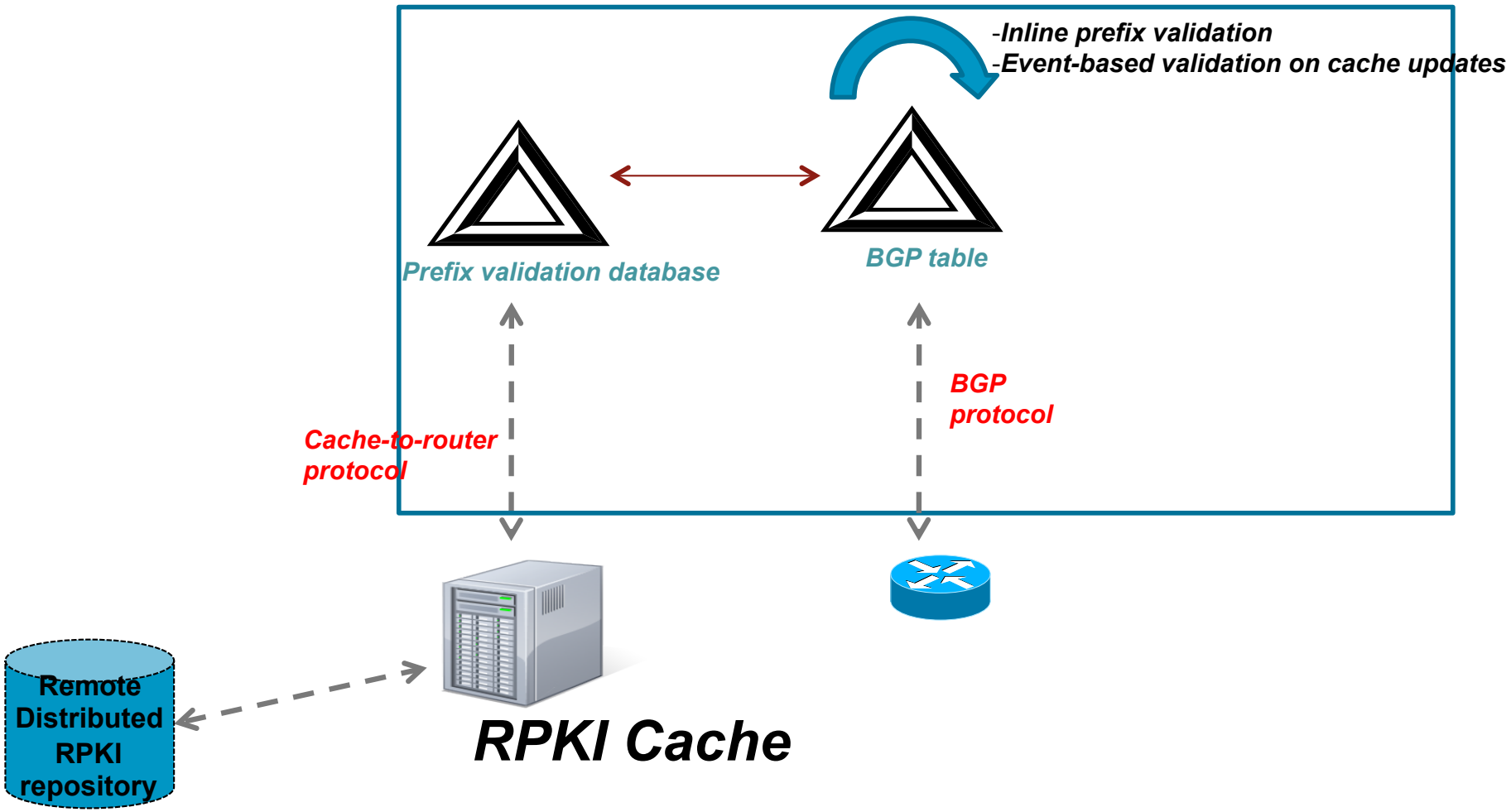


Route origin authorization (ROA)

- ROA is a digitally signed object distributed through the RPKI infrastructure
- Indicates the address prefix holder's explicit authorization that an AS can rightfully originate a prefix
- Format: [AS, {prefix/mask, maxLen}+]

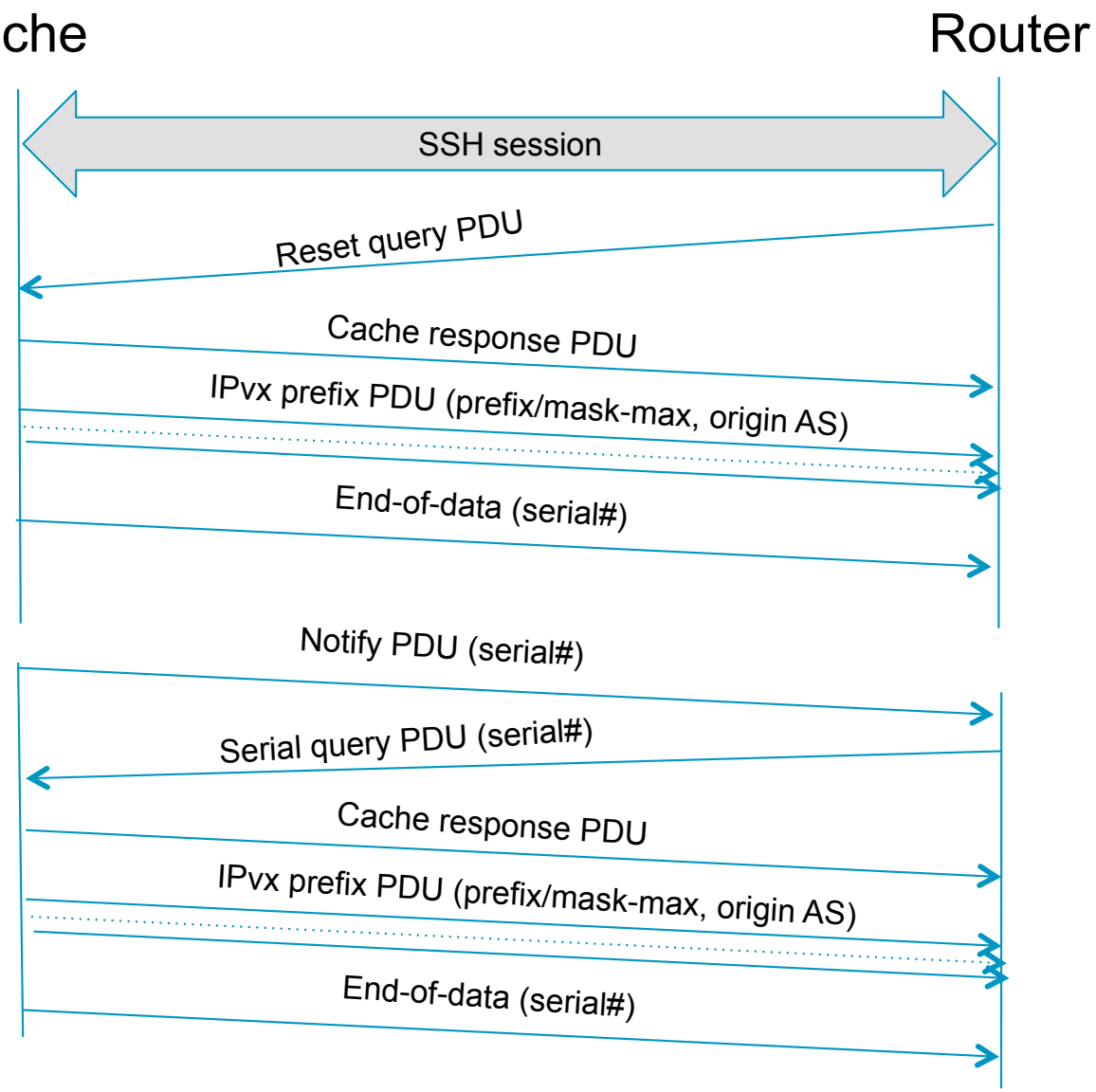
ROA
10.0.0.0/8-16
AS 65431

BGP design



Cache-to-router protocol

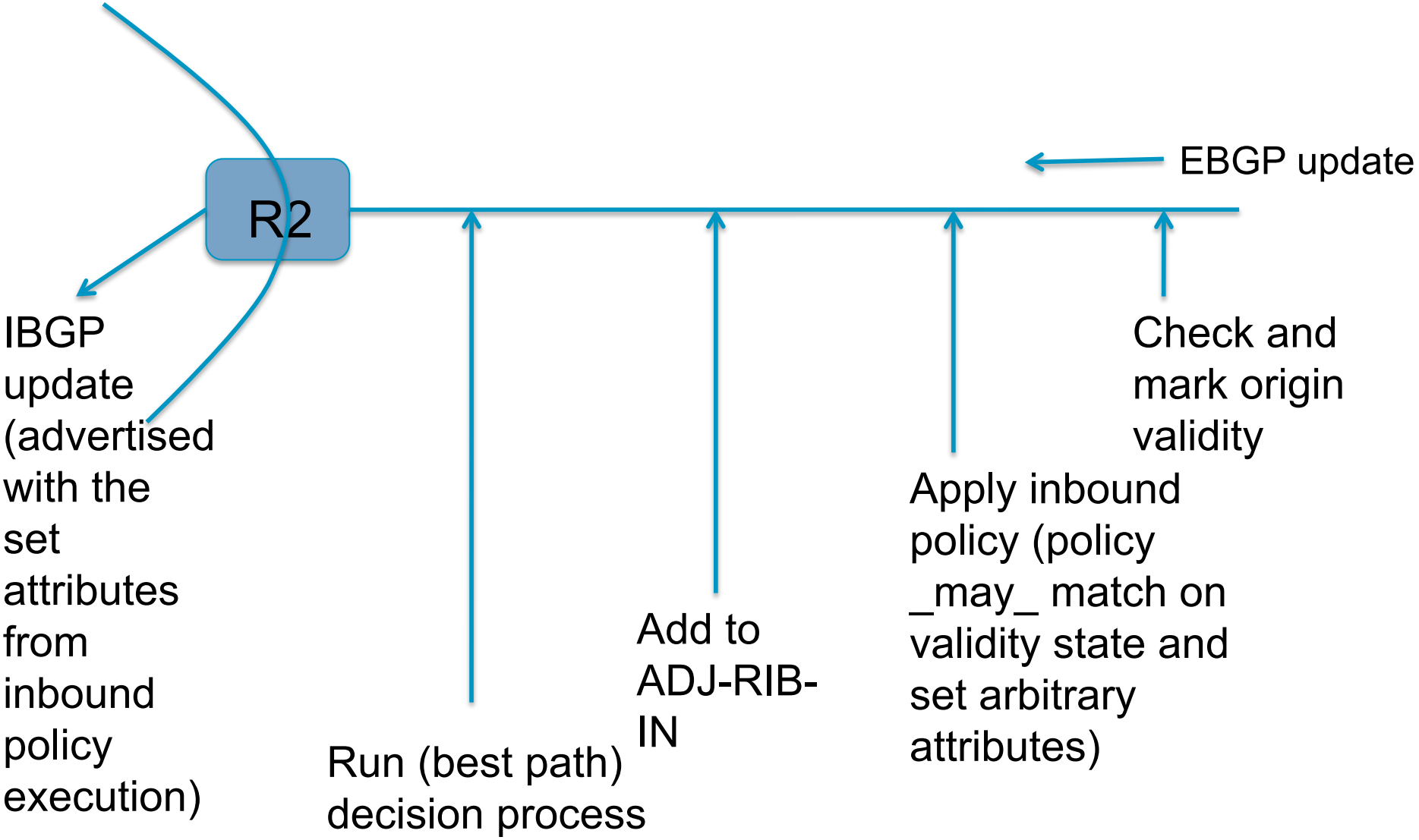
- Persistent SSH session between the router and one or more caches
- Serial# based binary exchange of PDUs containing the prefix<->origin AS mappings
- PDUs formatted as TLVs
- Notification service for changes to the cache database and incremental updates



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 execution



Policy examples

```
route-policy validity-0
       if origin-validation-state is valid then
         set local-preference 100
         else set local-preference 50
   endif
end-policy
```

```
route-policy validity-2
       if origin-validation-state is valid then
         set metric 100
       elseif origin-validate-state is not-found
         set metric 50
       else set metric 25
   endif
end-policy
```

Decision process changes

- Only enabled by configuration
- Before local-preference comparison step
- Path's validation states:

```
typedef enum {
    BGP_PFXV_STATE_VALID = 0,
    BGP_PFXV_STATE_NOT_FOUND = 1,
    BGP_PFXV_STATE_INVALID = 2,
} bgp_pfxv_state_e;
```

- Best path comparison

```
1. INPUT: received path, current bestpath
2. if received path's validation state > current bestpath's validation state
3.     prefer current bestpath
4. else if received path's validation state < current bestpath's validation state
5.     prefer received path
6. else goto next comparison step
7. endif
8. <rest of the tie breaking steps of BGP decision process>
```

Policy overrides

- Disable/enable prefix validation marking [globally, per EBGP peer, for a set of prefixes]
- Enable/disable validation state comparison in decision process [globally, per EBGP peer, 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

Extended community

- Prefix validation marking done only for EBGP updates
- Need a way to carry the marking across IBGP mesh so that other speakers take the correct/consistent best path decision
- Carry the validation state in an opaque extended community (non-transitive)

Status

- Prototype code for the routers available on IOS and IOS-XR

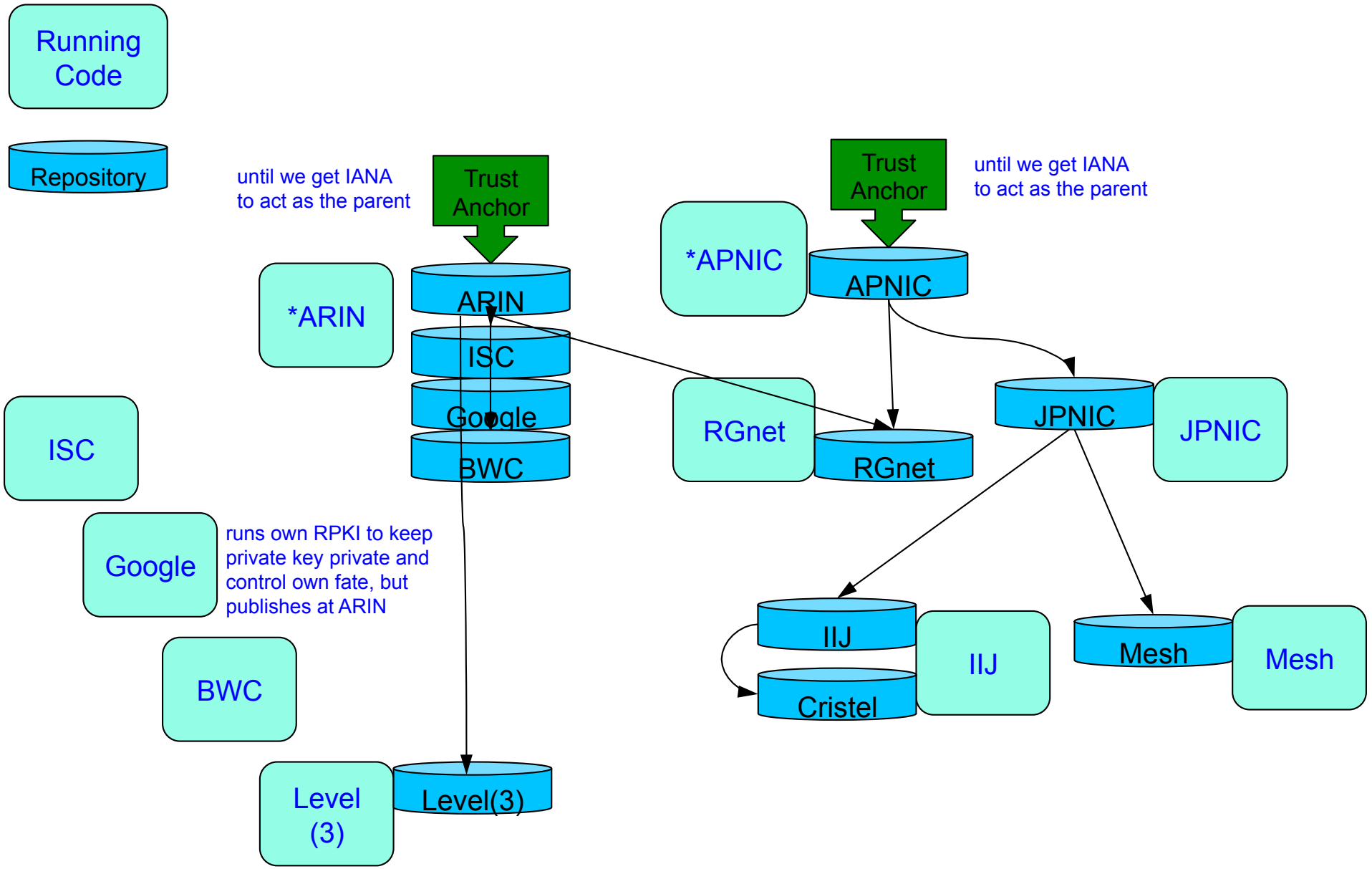
  Contact Ed Kern (ejk@cisco.com) if interested to play

- RPKI full implementation available as open source

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

  Mailing list: rpki-testbed@rpki.net

Open Test bed



Router configuration commands

- router bgp <as#>

```
bgp rpki cache <cache name> <port#> refresh-time <time>

bgp origin-validation {disable}

bgp bestpath compare-validation-state {allow-invalid |
disallow-not-valid}
```

Router show commands

```
RP/0/1/CPU0:r0.dfw#show bgp rpki prefix-validation database
Network           Maxlen          Origin-AS    Color      Source
64.9.224.0/19     24              15169        0          0
72.14.224.0/24    24              36384        0          0
72.14.230.0/24    24              36384        0          0
149.20.0.0/16     16              1280         0          0
192.5.4.0/24      24              3557         0          0
192.5.5.0/24      24              3557         0          0
192.158.248.0/24  24              27318        0          0
192.158.249.0/24  24              27319        0          0
192.158.250.0/24  24              27320        0          0
192.158.251.0/24  24              27321        0          0
192.158.252.0/24  24              27322        0          0
192.228.80.0/24   24              30122        0          0
192.228.81.0/24   24              30123        0          0
192.228.82.0/24   24              30126        0          0
192.228.83.0/24   24              30127        0          0
192.228.84.0/24   24              30124        0          0
192.228.85.0/24   24              30125        0          0
192.228.86.0/24   24              30128        0          0
192.228.87.0/24   24              30129        0          0
192.228.88.0/24   24              30130        0          0
192.228.89.0/24   24              8674         0          0
192.228.90.0/24   24              2500         0          0
192.228.91.0/24   24              27319        0          0
192.228.92.0/24   24              30131        0          0
199.6.0.0/24      24              33071        0          0
199.6.1.0/24      24              30132        0          0
199.6.2.0/24      24              30133        0          0
199.6.3.0/24      24              33073        0          0
199.6.4.0/24      24              30134        0          0
199.6.5.0/24      24              33072        0          0
199.6.6.0/24      24              33074        0          0
199.6.7.0/24      24              33075        0          0
199.6.8.0/24      24              33076        0          0
199.6.9.0/24      24              33077        0          0
199.6.10.0/24     24              33078        0          0
199.6.11.0/24     24              33079        0          0
199.6.12.0/24     24              33080        0          0
199.6.13.0/24     24              33081        0          0
199.6.14.0/24     24              33082        0          0
204.152.184.0/21  21              1280         0          0

RP/0/1/CPU0:r0.dfw#
```

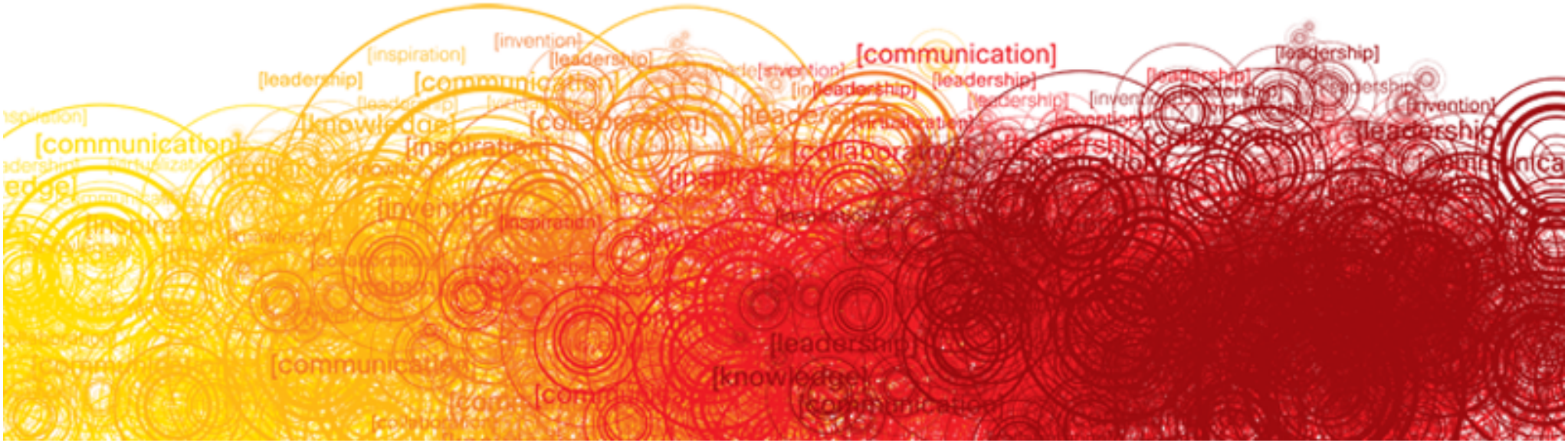
Router show commands

```
RP/0/0/CPU0:cons-ejk-xr#show bgp 199.6.7.0/24
BGP routing table entry for 199.6.7.0/24
Versions:
  Process           bRIB/RIB  SendTblVer
  Speaker                 0           0
Last Modified: Sep 30 09:58:36.715 for 00:04:36
Paths: (1 available, no best path)
  Not advertised to any peer
  Path #1: Received by speaker 0
  4128 25973 3549 16471 33075, (received & used)
    157.238.224.150 (inaccessible) from 157.238.224.150 (198.180.152.251)
      Origin IGP, localpref 100, valid, external, origin validity state: valid
                                                  ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
```

Reference

- *draft-ietf-sidr-arch*
- *draft-pmohapat-sidr-pfx-validate*
- *draft-ymbk-rpki-rtr-protocol*
- *draft-pmohapat-sidr-origin-validation-signaling-00.txt*

Questions?


[invention]
[inspiration]
[leadership]
[communication]
[knowledge]
[collaboration]

RPKI