Understand BGP RPKI With XR7 Cisco8000 Whitepaper

Contents

Introduction
Background Information
Preface
Scope
Prerequisites
Disclaimer
BGP Problems due to Bad Prefix Advertisement
Route Hijacking
Degrade System Performance
Sub-Prefix Hijacking
RPKI
Validator
BGP RPKI Demonstration
Topology
Configure
BGP RPKI Session
ROA Downloads on Router
Verify
Enabling Origin-As Validity
Prefix Validity States
1. 203.0.113.0/24 - Valid
2. 203.0.113.1/24 - Invalid
3. 192.168.122.1/32 Not Found
Allow Invalid Prefix
Manual ROA Configuration on Router
Route-Policy and Prefix Validation State
Share Prefix Validation Information through Extended Community
Recommendations for BGP RPKI Implementation
Good Practices for ROA Creation
Performance Impact of RPKI on XR BGP Routers
Effect of ROA Update on CPU with Route-Policy
Minimize CPU Impact Caused by ROA Update
BGP RPKI Memory Footprint
Scenario 1. Three RPKI Servers Configured on Router
Scenario 2. Single RPKI Servers Configured on Router

Introduction

This document describes the Border Gateway Protocol (BGP) Resource Public Key Infrastructure (RPKI) feature on the Cisco IOS® XR Platform.

Background Information

Preface

This document discusses the BGP RPKI feature and how it safeguards BGP with routers against false/malicious BGP prefix updates.

Scope

This document uses Cisco 8000 with XR 7.3.1 release for demonstration. However, BGP RPKI is a platform-independent feature, the concepts discussed in this document applies to other Cisco platforms (with Cisco IOS, Cisco IOS-XE .) with appropriate equivalent CLI conversions. This document does not cover the procedure to add Route Origin Authorizations (ROAs) on regional internet registries.

Prerequisites

The reader needs knowledge of BGP protocol.

Disclaimer

Any Internet Protocol (IP) addresses used in this document are not intended to be actual addresses. Any examples, command display output, and figures included in the document are shown for illustrative purposes only. Any use of actual IP addresses in illustrative content is unintentional and coincidental.

BGP Problems due to Bad Prefix Advertisement

BGP serves as the backbone of internet traffic. Even though it is the most important component of internet core, it lacks the capability to verify if the ingress BGP announcement originated from an authorised autonomous system or not.

This limitation of BGP makes it an easy candidate for various kind of attacks. One common attack is called ‘route hijack’. This attack can be used exploited to:

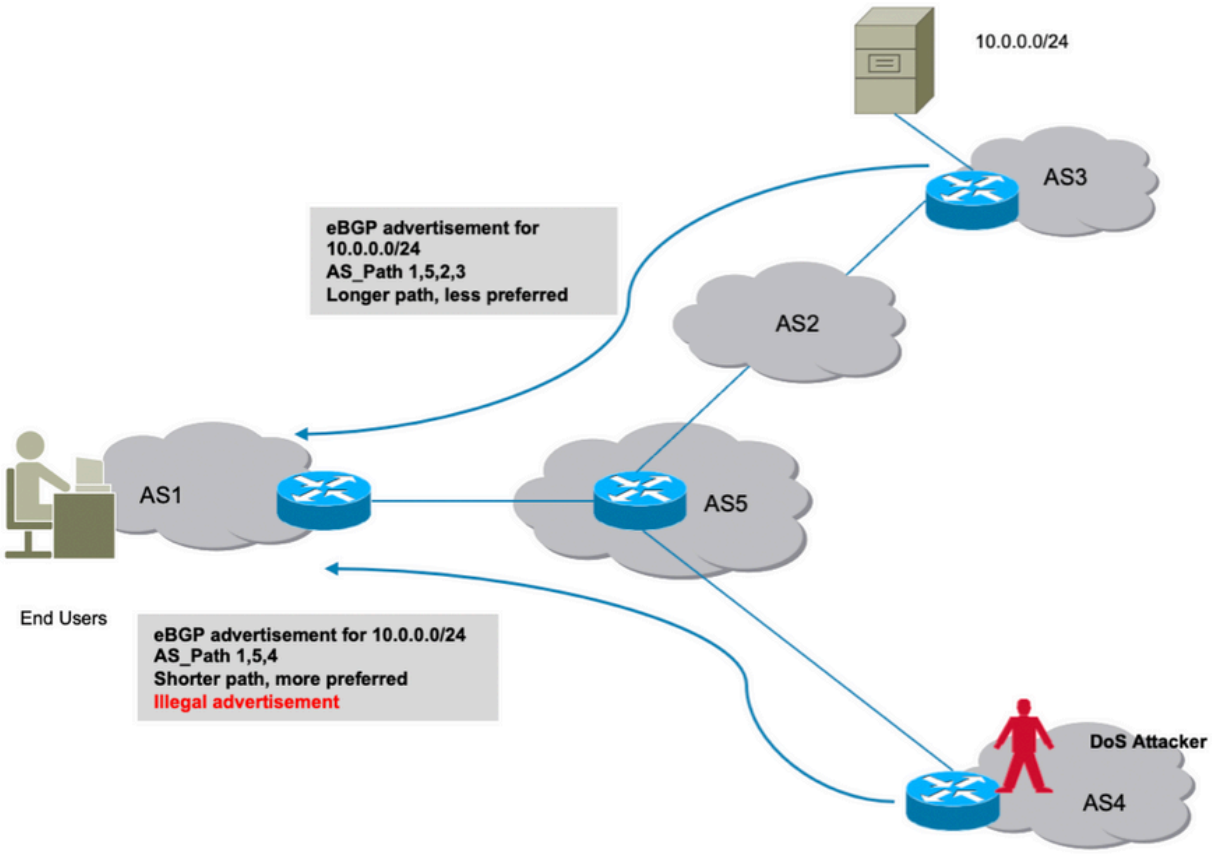
- Steal IPs to send spam results in IP gets rejected and hence denial of service.
- Spy on traffic to obtain sensitive information like passwords.
- Disruptions due to incorrect configurations by administrator.
- Prevent delivery of traffic by with up fake servers ensuesin denial of service.

Denial of Service attack (commonly known as DoS) is a malicious attempt to disrupt normal traffic to a router, switch, server and so on. There are variety of DoS attacks and few are discussed here.

Route Hijacking

Consider the scenario shown here. Autonomous System 3 (AS3) sends out legal BGP advertisement for its prefix 10.0.0.0/24. By BGP's design, there is nothing in BGP that would prevent an attacker from advertising same prefix to the internet.

As shown, attacker in AS4 advertises the same prefix 10.0.0.0/24. BGP best path algorithm prefers a path with shorter AS_Path. AS_Path 1,5,4 wins over longer path via AS 1,5,2,3. Therefore, the traffic from the clients will now be redirected to attacker's environment and can be black holed which ensues in denial of service to end clients.

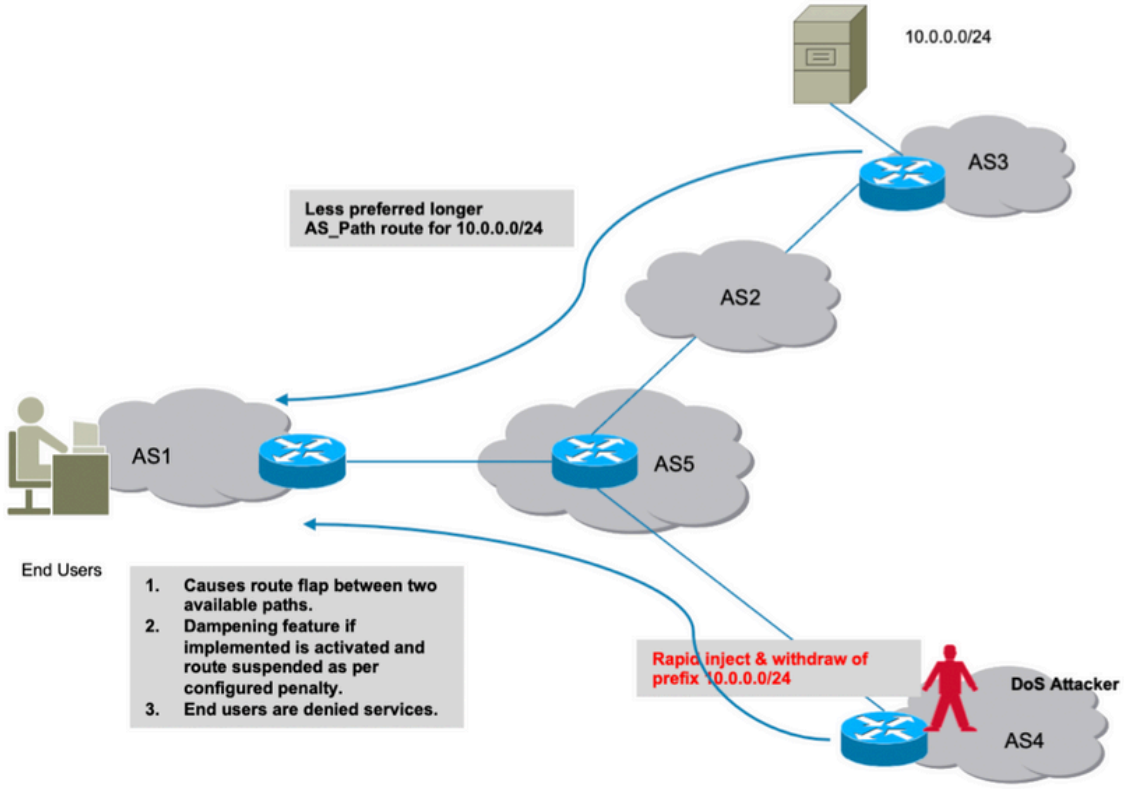


Route hijack

Degrade System Performance

This section discusses another way in which services can be denied. If Cisco's BGP route dampening feature is configured, it could be exploited if the attacker introduces rapid route flaps in the network causing a constant churn.

The dampening feature will impose penalties on the legitimate route and will make it unavailable for the actual traffic. In addition, this kind of unethically induced flaps will cause strain on the router's resources like CPU, memory and so on.

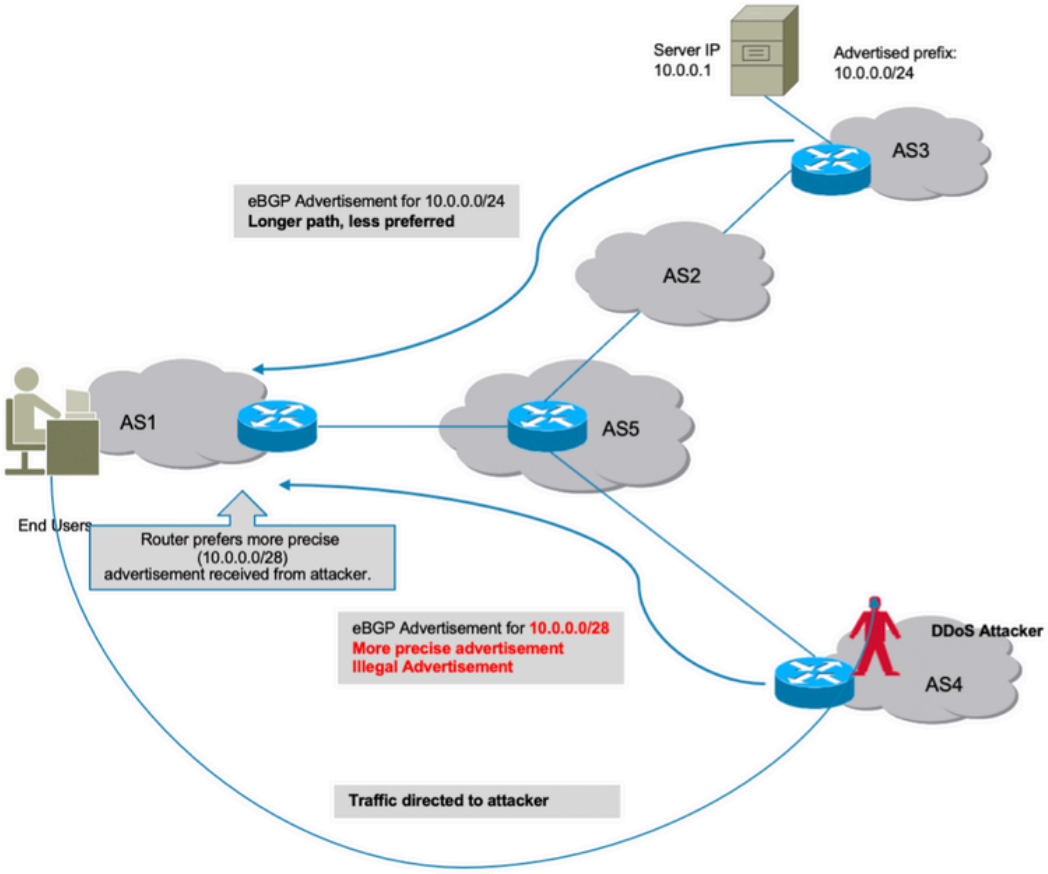


Route Dampening

Sub-Prefix Hijacking

As discussed in the previous section, how an attacker can originate a prefix illegally and cause a disruption of traffic. Unfortunately, a disruption is not the only cause of concern. In such attacks, actual data can be compromised wherein an attacker can scan received data for unethical use.

Similarly, hijacking of a route could be done by illegally advertising a more precise route. BGP prefers prefixes that are a longer match and this behaviour can be wrongly exploited as shown in the image.



Sub-prefix hijack

All the attacks that are discussed stem from the fact that BGP couldn't identify if the origin AS of these maliciously advertised prefixes was valid or not. In order to fix this, a 'true' and 'trusted' source of data is needed which a router can keep in its database. Then upon every receipt of a new advertisement, the router now becomes capable of cross verifying prefix's AS origin information received from BGP peer with its local database information from the validator.

Thus, the router is able to distinguish the good advertisements from the bad (illegal) ones and the capability to avoid all attacks discussed previously are inherently added on the router. BGP RPKI provides the needed trusted source of information.

RPKI

RPKI makes use of a repository that contains ROAs. A ROA contains information about prefix and their associated BGP AS number. Route origin authorization is a cryptographically signed statement.

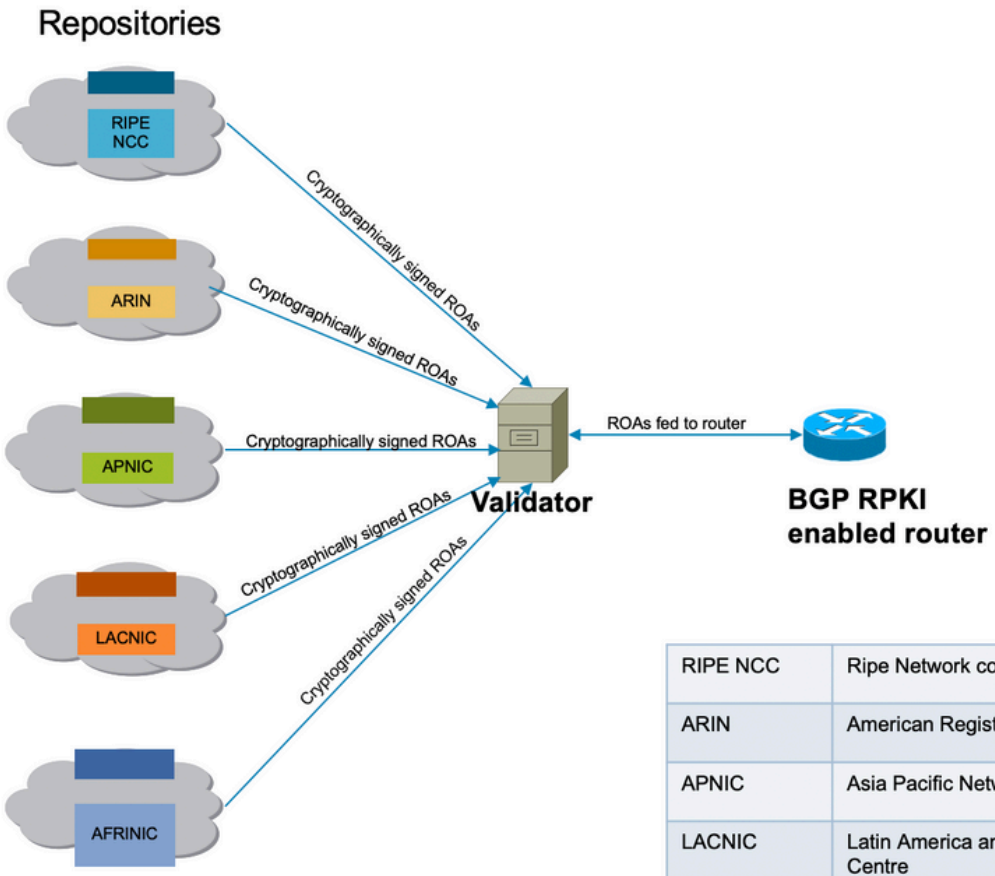
The 5 Regional Internet Registries (RIRs) are the trust anchors of the RPKI. Internet Assigned Numbers Authority (IANA) is the top of the tree that hands out IP prefixes. The RIRs are next in the hierarchy. They assign sub-prefixes to Local Internet Registries (LIRs) and large Internet Service Provider (ISP). They sign a Certificate for these prefixes. The next level allocates sub-prefixes of those and uses the certificates from above to sign their own certificates to certify their own allocations. They typically use their own publication points to host the certificates and ROAs. Each certificate lists the publication points of the child certificates it signs. Thus, RPKI forms a tree of certificates that mirrors the tree of IP address allocations. The RPKI validators owned by the

relying parties poll all the publication points to find updated certificates and ROAs (and CRLs and manifests). They start at the trust anchors and follow the links to the publication points of the child certificates.

ROAs are entered in the repository through RIRs but the same can be done through other registries (national or local). This responsibility can also be delegated to ISPs with proper supervision and verification by RIRs.

At this moment, there are five ROA repositories maintained by RIPE NCE, ARIN, APNIC, LACNIC and AFRINIC.

A validator present in the network communicates with these repositories and downloads a trusted ROA database to build its cache. This is a coalesced copy of the RPKI, which is periodically fetched/refreshed directly or indirectly from the Global RPKI. Validator then feeds this information to the routers enabling them to compare the incoming BGP announcements with the RPKI table in order to make a safe decision.



RIPE NCC	Ripe Network coordination center
ARIN	American Registry for Internet Numbers
APNIC	Asia Pacific Network Information Centre
LACNIC	Latin America and Caribbean Network Information Centre
AFRINIC	African Network Information Centre

RPKI infrastructure connectivity

Validator

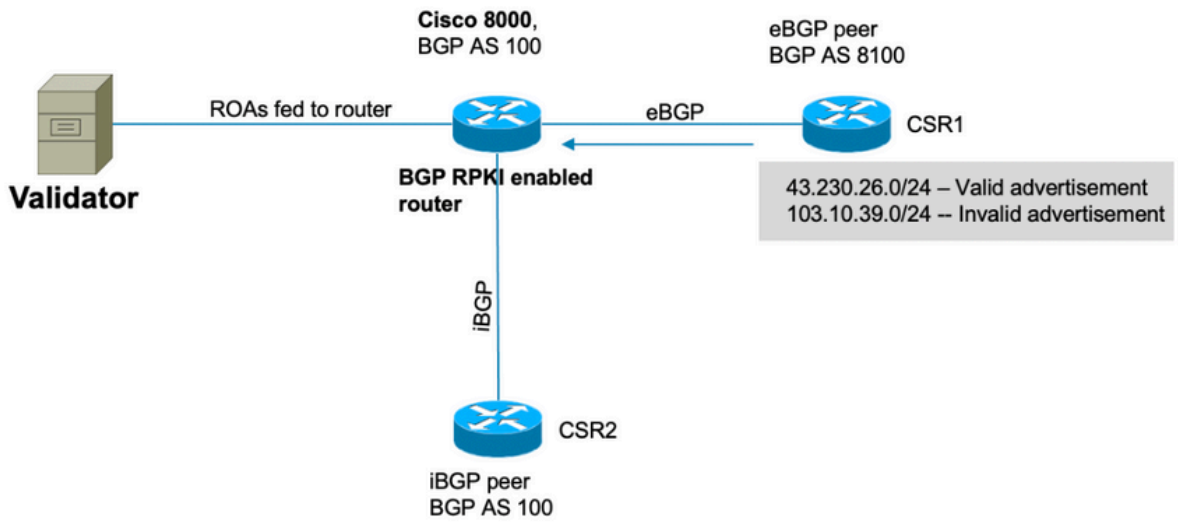
This demonstration makes use of the RIPE validator. The validator will communicate with the router by establishing a TCP session. In this demonstration, the validator listens on its IP 192.168.122.120 and port 3323.

```
routinator server --rtr 192.168.122.120:3323 --refresh=900
```

IANA has specified port 3323 for this communication. The refresh timer defines the time interval after which local repository will be synced and updated to stay updated.

BGP RPKI Demonstration

Topology



Topology

> **Note**: This demonstration uses random Public AS number and prefixes simply for the sake of explaining BGP RPKI mechanics. Public IPs are used due to RPKI primarily is meant for public prefix protection and all ROAs created on RIRs are public prefixes. Lastly, none of the actions, configurations etc. described in this document affects these public IPs and AS in any way.

Configure

```
router bgp 100

bgp router-id 10.1.1.1

rpki server 192.168.122.120

transport tcp port 3323

refresh-time 900



address-family ipv4 unicast
```

```
 !
 neighbor 10.0.12.2
 remote-as 8100
 address-family ipv4 unicast
  route-policy Pass in
  route-policy Pass out
 !
 !
 neighbor 10.0.13.3
 remote-as 100
 address-family ipv4 unicast
 !
 !
// ‘Pass’ is a permit all route-policy.
```

BGP RPKI Session

The router establishes a TCP session with a validator (IP: 192.168.122.120, port 3323) in order to download the ROA cache to the router’s memory.

```
RP/0/RP0/CPU0:Cisco8000#show bgp rpki server 192.168.122.120
Wed Jan 20 22:54:15.763 UTC

RPKI Cache-Server 192.168.122.120
 Transport: TCP port 3323
 Bind source: (not configured)
 Connect state: ESTAB
 Conn attempts: 1
 Total byte RX: 4428792
 Total byte TX: 1400
 Last reset
   Timest: Jan 20 05:59:58 (16:54:17 ago)
   Reason: protocol error
```

ROA Downloads on Router

Validator feeds the ROA information to the router. This cache is refreshed at periodic intervals in order to minimize the possibility of the router holding stale information. In this demonstration, a refreshing time of 900 seconds has been configured. As shown here, the Cisco 8000 router has downloaded 172632 IPv4 and 28350 IPv6 ROAs from the validator.

```
RP/0/RP0/CPU0:Cisco8000#show bgp rpki server summary

Wed Jan 20 23:01:59.432 UTC

Hostname/Address        Transport       State           Time                ROAs (IPv4/IPv6)

192.168.122.120         TCP:3323        ESTAB           17:00:21        172632/28350

RP/0/RP0/CPU0:Cisco8000#show bgp rpki table ipv4

Wed Jan 20 23:09:26.899 UTC

>>>Snipped output<<<

 Network               Maxlen      Origin-AS         Server

 10.0.0.0/24            24              13335             192.168.122.120

 10.0.4.0/22            22              38803             192.168.122.120

 10.0.4.0/24            24              38803             192.168.122.120

 10.0.5.0/24            24              38803             192.168.122.120

 10.0.6.0/24            24              38803             192.168.122.120

 10.0.7.0/24            24              38803             192.168.122.120

 10.1.1.0/24            24              13335             192.168.122.120

 10.1.4.0/22            22              4134              192.168.122.120

 10.1.16.0/20           20              4134              192.168.122.120

 10.2.9.0/24            24              4134              192.168.122.120

 10.2.10.0/24           24              4134              192.168.122.120

 10.2.11.0/24           24              4134              192.168.122.120

 10.2.12.0/22           22              4134              192.168.122.120

 10.3.0.0/16            16              4134              192.168.122.120

 10.6.0.0/22            24              9583              192.168.122.120
```

Verify

This section demonstrates how BGP RPKI in action and how it prevents the router from wrong/illegal advertisements.

Enabling Origin-As Validity

By default, the router fetches ROAs from the validator but does not begin using them until it is configured to do so. As a result, these prefixes are marked as ‘D’ or disabled.

```
RP/0/RP0/CPU0:Cisco8000#show bgp origin-as validity

Wed Jan 20 23:27:37.268 UTC

BGP router identifier 10.1.1.1, local AS number 100

BGP generic scan interval 60 secs

Non-stop routing is enabled

BGP table state: Active

Table ID: 0xe0000000   RD version: 30

BGP main routing table version 30

BGP NSR Initial initsync version 2 (Reached)

BGP NSR/ISSU Sync-Group versions 0/0

BGP scan interval 60 secs


Status codes: s suppressed, d damped, h history, * valid, > best

              i - internal, r RIB-failure, S stale, N Nexthop-discard

Origin codes: i - IGP, e - EGP, ? - incomplete

Origin-AS validation codes: V valid, I invalid, N not-found, D disabled

   Network            Next Hop            Metric LocPrf Weight Path

D*> 203.0.113.0/24     10.0.12.2                0             0 8100 ?

D*> 203.0.113.1/24     10.0.12.2                0             0 8100 ?

D*> 192.168.122.1/32   10.0.12.2                0             0 8100 ?
```

In order to enable the router for as-origin validity check, activate this command for the concerned address family.

```
router bgp 100

 address-family ipv4 unicast

 bgp origin-as validation enable

 !
```

When you activate this command, it causes the router to scan the prefixes present in its BGP table against the ROA information received from the validator and one of the three states is assigned to prefixes .

```
RP/0/RP0/CPU0:Cisco8000#show bgp origin-as validity

Thu Jan 21 00:04:58.136 UTC

Status codes: s suppressed, d damped, h history, * valid, > best

              i - internal, r RIB-failure, S stale, N Nexthop-discard

Origin codes: i - IGP, e - EGP, ? - incomplete

Origin-AS validation codes: V valid, I invalid, N not-found, D disabled

   Network            Next Hop            Metric LocPrf Weight Path

V*> 203.0.113.0/24     10.0.12.2                0             0 8100 ?

I*  203.0.113.1/24     10.0.12.2                0             0 8100 ?

N*> 192.168.122.1/32   10.0.12.2                0             0 8100 ?
```

In order to enable the router to use prefix validation state information while making the best path calculation, this command is needed. This is not enabled by default as it gives you the option of not using the validity information for best path calculation but still allowing you to use it in route policies which are discussed later in this document.

```
router bgp 100

 address-family ipv4 unicast

 bgp bestpath origin-as use validity

 !
```

Prefix Validity States

There are three states a prefix could be found in.

```
RP/0/RP0/CPU0:Cisco8000#show bgp origin-as validity

Thu Jan 21 00:04:58.136 UTC



Status codes: s suppressed, d damped, h history, * valid, > best

              i - internal, r RIB-failure, S stale, N Nexthop-discard

Origin codes: i - IGP, e - EGP, ? - incomplete

Origin-AS validation codes: V valid, I invalid, N not-found, D disabled

   Network            Next Hop            Metric LocPrf Weight Path

V*> 203.0.113.0/24     10.0.12.2                0             0 8100 ?

I*  203.0.113.1/24     10.0.12.2                0             0 8100 ?

N*> 192.168.122.1/32   10.0.12.2                0             0 8100 ?
```

- Invalid - Indicates the prefix meets either of these two conditions: 1. It matches one or more **Route Origin Authorizations** (ROAs), but there is no ROA match where the origin AS matches the origin AS on the AS-PATH. 2. It matches one or more ROAs at the minimum-length specified in the ROA, but for all ROAs where it matches the minimum length, it is longer than the specified maximum length. Origin AS does not matter for condition #2.
- Valid - Indicates the prefix and AS pair are found in the RPKI cache table.
- Not Found - Indicates the prefix is not among the valid or invalid prefixes.

This section discusses each prefix and its state in detail.

1. 203.0.113.0/24 - Valid

eBGP peer in AS 8100 originated this route and advertised to Cisco8000 node. Since the origin AS (8100) matches the origin AS in ROA (received from validator), this prefix is marked valid and is installed in the router's routing table.

```
RP/0/RP0/CPU0:Cisco8000#show bgp rpki table | in "203.0.113.0|Max"

Thu Jan 21 00:21:26.026 UTC

 Network               Maxlen          Origin-AS         Server

 203.0.113.0/24                24       8100              192.168.122.120
```

The route is installed in the BGP table.

```
RP/0/RP0/CPU0:Cisco8000#show bgp 203.0.113.0/24

Thu Jan 21 05:30:13.858 UTC

BGP routing table entry for 203.0.113.0/24

Versions:

 Process           bRIB/RIB  SendTblVer

 Speaker                 31          31

Last Modified: Jan 21 00:03:33.344 for 05:26:40

Paths: (1 available, best #1)

 Not advertised to any peer

 Path #1: Received by speaker 0

 Not advertised to any peer

 8100

   10.0.12.2 from 10.0.12.2 (192.168.122.105)

     Origin incomplete, metric 0, localpref 100, valid, external, best, group-best

     Received Path ID 0, Local Path ID 1, version 31

     Origin-AS validity: valid
```

Since this is the best BGP prefix and also valid per RPKI, it is successfully installed in the routing table.

```
RP/0/RP0/CPU0:Cisco8000#show route 203.0.113.0/24

Thu Jan 21 00:29:43.667 UTC



Routing entry for 203.0.113.0/24

 Known via "bgp 100", distance 20, metric 0

 Tag 8100, type external

 Installed Jan 21 00:03:33.731 for 00:26:10

 Routing Descriptor Blocks

   10.0.12.2, from 10.0.12.2, BGP external

     Route metric is 0

 No advertising protos.
```

2. 203.0.113.1/24 - Invalid

This prefix is invalid because there is a conflict in the origin AS information contained in ROA and the origin-as information received via BGP message from eBGP peer. 203.0.113.1/24 is received via BGP with origin AS 8100.

```
RP/0/RP0/CPU0:Cisco8000#show bgp origin-as validity invalid

Thu Jan 21 00:34:38.171 UTC

BGP router identifier 10.1.1.1, local AS number 100

BGP generic scan interval 60 secs

Non-stop routing is enabled

BGP table state: Active

Table ID: 0xe0000000   RD version: 33

BGP main routing table version 33

BGP NSR Initial initsync version 2 (Reached)

BGP NSR/ISSU Sync-Group versions 0/0

BGP scan interval 60 secs


Status codes: s suppressed, d damped, h history, * valid, > best

              i - internal, r RIB-failure, S stale, N Nexthop-discard
```

```
Origin codes: i - IGP, e - EGP, ? - incomplete

  Network            Next Hop            Metric LocPrf Weight Path

*  203.0.113.1/24     10.0.12.2                0             0 8100 ?
```

However, the ROA received from the validator shows that this prefix belongs to AS 10021.

```
RP/0/RP0/CPU0:Cisco8000#show bgp rpki table 203.0.113.1/24 max 24

Thu Jan 21 00:37:05.615 UTC


RPKI ROA entry for 203.0.113.1/24-24

 Origin-AS: 10021 from 192.168.122.120

Version: 124211
```

Since AS origin information in the received BGP announcement (AS 8100) did not match the actual AS origin received in ROA (AS 10021), the prefix is marked Invalid and is not installed in the routing table.

```
RP/0/RP0/CPU0:Cisco8000#show bgp 203.0.113.1/24

Thu Jan 21 05:37:26.714 UTC

BGP routing table entry for 203.0.113.1/24

Versions:

 Process           bRIB/RIB  SendTblVer

 Speaker                 32          32

Last Modified: Jan 21 00:03:33.344 for 05:33:53

Paths: (1 available, no best path)

 Not advertised to any peer

 Path #1: Received by speaker 0

 Not advertised to any peer

 8100

   10.0.12.2 from 10.0.12.2 (192.168.122.105)

     Origin incomplete, metric 0, localpref 100, valid, external

     Received Path ID 0, Local Path ID 0, version 0

     Origin-AS validity: invalid
```

3. 192.168.122.1/32 Not Found

This is a private prefix and is not present in the ROA cache. BGP declared this prefix as ‘Not found’.

```
RP/0/RP0/CPU0:Cisco8000#show bgp 192.168.122.1/32

Thu Jan 21 05:44:39.861 UTC

BGP routing table entry for 192.168.122.1/32

Versions:

 Process           bRIB/RIB  SendTblVer

 Speaker                 33          33

Last Modified: Jan 21 00:03:33.344 for 05:41:06

Paths: (1 available, best #1)

 Not advertised to any peer

 Path #1: Received by speaker 0

 Not advertised to any peer

 8100

   10.0.12.2 from 10.0.12.2 (192.168.122.105)

     Origin incomplete, metric 0, localpref 100, valid, external, best, group-best

     Received Path ID 0, Local Path ID 1, version 33

     Origin-AS validity: not-found
```

Since RPKI is still gets adopted, 'not-found' prefixes are installed in the routing table. Otherwise will cause BGP to ignore these legitimate prefixes that are not registered in the RPKI database.

Allow Invalid Prefix

Although it is not recommended, the software does provide a knob to allow invalid prefixes to participate in the best path calculation algorithm.

```
router bgp 100

 address-family ipv4 unicast

 bgp bestpath origin-as allow invalid

 !
```

With this configuration, the router does consider invalid prefixes for best path calculation while This marked as ‘invalid’. This output shows ‘203.0.113.1/24’ marked as the best path.

```
RP/0/RP0/CPU0:Cisco8000#show bgp

Thu Jan 21 06:21:34.294 UTC

BGP router identifier 10.1.1.1, local AS number 100

BGP generic scan interval 60 secs
```

```
Non-stop routing is enabled

BGP table state: Active

Table ID: 0xe0000000   RD version: 34

BGP main routing table version 34

BGP NSR Initial initsync version 2 (Reached)

BGP NSR/ISSU Sync-Group versions 0/0

BGP scan interval 60 secs



Status codes: s suppressed, d damped, h history, * valid, > best

              i - internal, r RIB-failure, S stale, N Nexthop-discard

Origin codes: i - IGP, e - EGP, ? - incomplete

  Network            Next Hop            Metric LocPrf Weight Path

*> 203.0.113.0/24     10.0.12.2                0             0 8100 ?

*> 203.0.113.1/24     10.0.12.2                0             0 8100 ?

*> 192.168.122.1/32   10.0.12.2                0             0 8100 ?
```

As shown in this output, the prefix is marked as best despite kept invalid.

```
RP/0/RP0/CPU0:Cisco8000#show bgp 203.0.113.1/24

Thu Jan 21 06:23:26.994 UTC

BGP routing table entry for 203.0.113.1/24

Versions:

 Process           bRIB/RIB  SendTblVer

 Speaker                 34          34

Last Modified: Jan 21 06:05:31.344 for 00:17:55

Paths: (1 available, best #1)

 Not advertised to any peer

 Path #1: Received by speaker 0

 Not advertised to any peer

 8100

   10.0.12.2 from 10.0.12.2 (192.168.122.105)

     Origin incomplete, metric 0, localpref 100, valid, external, best, group-best

     Received Path ID 0, Local Path ID 1, version 34
```

```
     Origin-AS validity: invalid
```

It is to be noted that a router still treats invalid prefix as the last option and always prefers a valid prefix over an invalid prefix if it is available.

Manual ROA Configuration on Router

If for some reason, a ROA for a certain prefix is not yet created, received or is delayed, a manual ROA could be configured on the router. For example, the prefix ‘192.168.122.1/32’ is marked as ‘Not Found’ as shown here.

```
RP/0/RP0/CPU0:Cisco8000#show bgp origin-as validity

Thu Jan 21 06:36:31.041 UTC

BGP router identifier 10.1.1.1, local AS number 100

BGP generic scan interval 60 secs

Non-stop routing is enabled

BGP table state: Active

Table ID: 0xe0000000   RD version: 34

BGP main routing table version 34

BGP NSR Initial initsync version 2 (Reached)

BGP NSR/ISSU Sync-Group versions 0/0

BGP scan interval 60 secs



Status codes: s suppressed, d damped, h history, * valid, > best

              i - internal, r RIB-failure, S stale, N Nexthop-discard

Origin codes: i - IGP, e - EGP, ? - incomplete

Origin-AS validation codes: V valid, I invalid, N not-found, D disabled

   Network            Next Hop            Metric LocPrf Weight Path

V*> 203.0.113.0/24     10.0.12.2                0             0 8100 ?

I*> 203.0.113.1/24     10.0.12.2                0             0 8100 ?

N*> 192.168.122.1/32   10.0.12.2                0             0 8100 ?
```

A manual ROA could be configured as shown here. This command associates’ prefix ‘192.168.122.1/32’ with AS 8100.

```
router bgp 100

 rpki route 192.168.122.1/32 max 32 origin 8100
```

With this configuration, the state of the prefix changes from ‘N’ to ‘V’.

```
RP/0/RP0/CPU0:Cisco8000#show bgp origin-as validity

Thu Jan 21 06:36:34.151 UTC

BGP router identifier 10.1.1.1, local AS number 100

BGP generic scan interval 60 secs

Non-stop routing is enabled

BGP table state: Active

Table ID: 0xe0000000   RD version: 35

BGP main routing table version 35

BGP NSR Initial initsync version 2 (Reached)



Status codes: s suppressed, d damped, h history, * valid, > best

              i - internal, r RIB-failure, S stale, N Nexthop-discard

Origin codes: i - IGP, e - EGP, ? - incomplete

Origin-AS validation codes: V valid, I invalid, N not-found, D disabled

   Network            Next Hop            Metric LocPrf Weight Path

V*> 203.0.113.0/24     10.0.12.2                0             0 8100 ?

I*> 203.0.113.1/24     10.0.12.2                0             0 8100 ?

V*> 192.168.122.1/32   10.0.12.2                0             0 8100 ?
```

Route-Policy and Prefix Validation State

Prefix state result can be used to create route policies. These states can be used in a match statement and administrator desired actions can be taken. This example matches all the prefixes with an invalid state and sets a weight value of 12345 for them.

```
route-policy Invalid

 if validation-state is invalid then

   set weight 12345

 endif

end-policy

!
```

```
router bgp 100
 remote-as 8100
 address-family ipv4 unicast
  route-policy Invalid in
 !
 !
!
```

This output shows an invalid prefix applied weight of 12345.

```
RP/0/RP0/CPU0:Cisco8000#show bgp 203.0.113.1/24
Thu Jan 21 06:57:33.816 UTC
BGP routing table entry for 203.0.113.1/24
Versions:
 Process           bRIB/RIB  SendTblVer
 Speaker                 38          38
Last Modified: Jan 21 06:54:04.344 for 00:03:29
Paths: (1 available, best #1)
 Not advertised to any peer
 Path #1: Received by speaker 0
 Not advertised to any peer
 8100
   10.0.12.2 from 10.0.12.2 (192.168.122.105)
     Origin incomplete, metric 0, localpref 100, weight 12345, valid, external, best, group-best
     Received Path ID 0, Local Path ID 1, version 38
     Origin-AS validity: invalid
```

Share Prefix Validation Information through Extended Community

As BGP router can also share the prefix-validation state with other routers (with no local cache from validator) via BGP extended community. This saves the overhead of each and every router in the network with a session with the validator and downloading all the ROAs.

This is made possible by the BGP extended community.

This command enables the router to share ‘prefix-validation’ information with iBGP peers.

```
router bgp 100

 address-family ipv4 unicast

 bgp origin-as validation signal ibgp
```

Once Cisco 8000 router is configured as shown, BGP updates to peers enclose prefix validation information. In this case, the neighbor iBGP router is an IOS-XE router.

```
csr2#show ip bgp 203.0.113.1/24

BGP routing table entry for 203.0.113.1/24, version 14

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

 Not advertised to any peer

 Refresh Epoch 1

 8100

   10.0.12.2 from 10.0.13.1 (10.1.1.1)

     Origin IGP, metric 0, localpref 100, valid, internal, best

     Extended Community: 0x4300:0:2

     rx pathid: 0, tx pathid: 0x0

     Updated on Jan 21 2021 18:16:56 UTC
```

This extended community mapping can be understood with the use of 0x4300 0x0000 (4 bytes indicating state).

The four bytes indicating state are treated as a 32-bit unsigned integer having one of the values:

- 0 - Valid
- 1 - Not Found
- 2 - Invalid

Prefix 203.0.113.1/24's community is 0x4300:0:2 which maps to the 'Invalid' prefix. This way, the csr2 router despite no local cache of its own is still able to make decisions based on prefix-validation state.

Prefix validation state now can be used to match in a route-map or in BGP best path algorithm.

Recommendations for BGP RPKI Implementation

Good Practices for ROA Creation

These are some recommendations based on unreachable networks observed at RPKI-Observatory. The RPKI Observatory analyzes multiple aspects of the deployed RPKI landscape.

- If a ROA is created for any prefix, then it is recommended to announce that prefix in BGP. In absence of it, someone else can announce it by simply pretending to be ASN contained in

that ROA and use the prefix.

- If ROA is created with a maxlen greater than the prefix length, then it is equivalent to creating ROAs for all possible prefixes under the original prefix up to the maxlen. It is strongly recommended to announce all those prefixes in BGP.
- If a ROA is created for a prefix and the prefix owner announces a sub-prefix of the original prefix, then the ROA will invalidate that sub-prefix. A ROA for the sub-prefix as well or the maxlen of the original ROA must be extended to cover the sub-prefix.
- If an organization owns a prefix, but plan not to announce it in BGP, then a ROA for the prefix for AS0 must be created. This will invalidate any announcement of prefix because AS0 cannot appear in any AS path.
- If there are multiple ASNs originating the same prefix, then ROAs for that prefix must be created for each of the ASNs. Consequently, if a router has multiple ROAs for the same prefix, a BGP advertisement that matches any of them will be valid. Multiple ROAs for the same prefix do not conflict with each other.
- If ‘A’ is originating a prefix for its customer ‘B’ and create a ROA for that prefix on behalf of ‘B’, then ‘A’ must prepend ‘B’s’ ASN to the announcement or have the ‘B’ originate the prefix itself.

Performance Impact of RPKI on XR BGP Routers

Effect of ROA Update on CPU with Route-Policy

When ROAs are updated and if the router has a local ingress route-policy for a neighbor that contains a "validation-state is", then it becomes important to re-validate the status of prefixes based on new updated ROAs. This is achieved by the router sending a BGP REFRESH request to its peer.

When BGP neighbors receive this message as shown, neighbors send their prefixes again and the inbound route-policy can revalidate the incoming prefixes .

```
Jan 22 18:28:41.360: BGP: 10.0.12.1 rcv message type 5, length (excl. header) 4

Jan 22 18:28:41.360: BGP: 10.0.12.1 rcvd REFRESH_REQ for afi/safi: 1/1, refresh code is 0
```

The problem amplifies when a lot of neighbors refresh at the same time whenever ROAs are updated. If neighbor inbound route-policy are complex and require a lot of processing, then high CPU results for a few minutes after a ROA update. These REFRESH messages do not occur if the neighbor inbound route-policy does not contain a "validation-state is" command.

If “soft-reconfiguration inbound always” is configured for a neighbor, then BGP REFRESH messages will not be sent, but the same route policies will still be executed at the same rate and the same CPU usage can be expected.

It is recommended to prefer ‘bgp bestpath origin-as use validity’ approach over configuring a route- policy for the reasons explained in 6.2.2 below.

Minimize CPU Impact Caused by ROA Update

The best way to avoid the problem explained here is to use **bestpath origin-as use validity** without **validation-state is** in the policy.

```
router bgp 100

 address-family ipv4 unicast

 bgp bestpath origin-as use validity

 !
```

This command keeps a received invalid route on router but prevents it from becoming a best-path. It will not be installed or further advertised. It is as good as dropping it. If with the next ROA update it becomes valid, no REFRESH is required, and it will automatically become eligible for the best path with no policy execution necessary.

If the user prefers to allow ‘invalid’ prefixes and not use them, then in addition to **bestpath origin-as use validity**, use the configuration **best path origin-as allow invalid**.

In this case, when a ROA changes, the best path is automatically updated without requiring a REFRESH message. In order to de-prefer, a route means that during the BGP route selection the RPKI invalid path is considered less preferable than any other path to the same destination. It is similar to assigning it weight or local preference less than 0.

The number of RPKI invalids is relatively small and kept in the table does not result in a significant impact on resources.

> **Note**: In order to use "bestpath origin-as use validity", all paths of a route, including the IBGP paths, must have the correct RPKI validity. If not, then testing of validation-state in route-policy can still be used.

IBGP routes are not validated by the router against the ROA database. IBGP routes gain an RPKI validity from the RPKI extended community. If the IBGP route is received without this extended community, then its validation-state is set to not-found.

BGP RPKI Memory Footprint

Each ROA consumes memory for the index and the data. If two ROAs are for the same IP prefix, but have different max_len or are received from different RPKI servers, then they share the same index but have separate data. Memory requirements can vary because memory overhead is not constant. An overbudget of 10% is recommended. 64-bit platforms require more memory for each memory object than 32-bit platforms. IOS-XR memory usage in bytes for an index object and a data object is in the table. Some mostly constant overhead is included in the numbers.

	32-bit platform (bytes)	64-bit platform(bytes)
IPv4 index	74	111
IPv6 index	86	125
data	34	53

This section takes two scenarios to explain how ROAs consume memory.

Scenario 1. Three RPKI Servers Configured on Router

Consider a router using 3 RPKI servers, each providing 200,000 IPv4 ROAs and 20,000 IPv6 ROAs on a 64-bit route-processor will require this memory:

20000 * (125 + 3*53) + 200000 * (111 + 3*53) bytes = 59.68 million bytes

While calculating the memory, ROA for the same prefix from three different validators shared the same index value.

Scenario 2. Single RPKI Servers Configured on Router

BGP process memory without ROAs:

```
RP/0/RP0/CPU0:Cisco8000#show processes memory detail location 0/RP0/CPU0 | in $

Fri Jan 22 17:19:57.945 UTC

JID        Text       Data       Stack      Dynamic    Dyn-Limit  Shm-Tot    Phy-Tot
 Process

1069           2M        71M       132K        25M      7447M        50M        74M        bgp


RP/0/RP0/CPU0:Cisco8000#show bgp rpki server summary

Fri Jan 22 17:12:09.073 UTC

Hostname/Address    Transport     State         Time          ROAs (IPv4/IPv6)

192.168.122.120     TCP:3323      NONE          00:00:25      N/A
```

BGP process is seen consuming 25 MB memory without any ROAs.

BGP process memory with ROA:

```
RP/0/RP0/CPU0:Cisco8000#show bgp rpki server summary

Fri Jan 22 17:23:46.769 UTC


Hostname/Address    Transport     State         Time          ROAs (IPv4/IPv6)

192.168.122.120     TCP:3323      ESTAB         00:02:42      172796/28411


RP/0/RP0/CPU0:Cisco8000#show processes memory detail location 0/RP0/CPU0 | in $

Fri Jan 22 17:24:14.659 UTC

JID        Text       Data       Stack      Dynamic    Dyn-Limit  Shm-Tot    Phy-Tot     Process

1069           2M        99M       132K        53M      7447M        50M       102M       bgp
```

BGP process is seen consuming 25 MB memory without any ROAs.

BGP process memory with ROA:

```
RP/0/RP0/CPU0:Cisco8000#show bgp rpki server summary

Fri Jan 22 17:23:46.769 UTC


Hostname/Address    Transport     State         Time          ROAs (IPv4/IPv6)

192.168.122.120     TCP:3323      ESTAB         00:02:42      172796/28411

RP/0/RP0/CPU0:Cisco8000#show processes memory detail location 0/RP0/CPU0 | in $

Fri Jan 22 17:24:14.659 UTC

JID        Text       Data       Stack      Dynamic    Dyn-Limit  Shm-Tot    Phy-Tot     Process

1069          2M        99M       132K        53M      7447M        50M       102M       bgp
```

Cisco 8000 router runs 64-bit OS. It received 172796 IPv4 ROA and 28411 ROA.

Memory (Bytes) = 172,796 x [111 (index) + 53 (data)] + 28411 x [125 (index) + 53 (data)].

These calculations give ~27 MB which is approximately the increment noticed on the router's memory above.