The ISP Column
*A column on things Internet*

November 2023
Geoff Huston

# How We Measure: RPKI ROA Signing and Route Origination Validation

At APNIC Labs we publish a number of measurements of the deployment of various technologies that are being adopted on the Internet. Here we will look at how we measure the adoption of the signing of *Route Origination Attestations* (ROAs) as part of the framework for securing inter-domain routing on the Internet using the digital credential framework provided by the *Resource Public Key Infrastructure* (RPKI).

## RPKI and Securing Inter-Domain Routing

The Internet is an interconnected collection of component networks. Each of these component networks are separately managed as an *Autonomous System* (AS). Managing the *in-between* space between these networks is the role of Inter-Domain Routing. By its very nature, no network controls this inter-domain space. Each network advertises this network's reachability of address prefixes to other networks and learns from the advertisements of other networks.

To manage this routing space, we use the *Border Gateway Protocol* (BGP). This protocol is an instance of a Bellman-Ford distance vector routing algorithm. BGP allows a collection of connected devices (BGP speakers) to each learn the relative topology of the connecting network. The basic approach of this algorithm is very simple: each BGP speaker tells all its adjacent BGP neighbours about what it has learned using an *update*. If this new learned information alters the local view of the network, the local BGP view is updated, and all of this BGP speaker's neighbours are informed of the new information. This is a lot like a social rumour network, where every individual who hears a new rumour immediately informs all their friends. BGP works in a very similar fashion: each time a neighbour informs a BGP speaker about some updated reachability information about an IP address prefix, the BGP speaker compares this new reachability information against its stored knowledge that was gained from previous announcements from other neighbours. If this new information provides a shorter path to the prefix, then the local speaker moves this prefix and associated next-hop forwarding decision to the local forwarding table and informs all its immediate neighbours of a new path to a prefix, implicitly citing itself as the next hop. In addition, there is a withdrawal mechanism, where once a BGP speaker determines that it no longer has a viable path to a given prefix, it announces a *withdrawal* for this prefix to all its neighbours. When a BGP speaker receives a withdrawal, it stores the withdrawal against this neighbour. If the withdrawn neighbour path happened to be the currently preferred next hop for this prefix, then the BGP speaker will examine its per-neighbour data sets to determine which stored announcement represents the best path from those that are still extant. If it can find such an alternative path, it will copy this into its local forwarding table and announce this new preferred path to all its BGP neighbours. If there is no such alternative path, it will announce a withdrawal to its neighbours, indicating that it no longer can reach this prefix.

Like any rumour network, it is difficult to control the authenticity of information that is passed through the inter-domain routing space. Each BGP speaker receives information, performs a local process, and may or may not generate information to pass to its neighbours. This is a store/forward relay mode of information and there is no information that is passed through the routing space in an analogous manner as end-to-end TCP packets. Each BGP speaker operates in a manner that is opaque to all other BGP speakers. Individual BGP speakers may inject erroneous information into the routing domain, erroneously or deliberately, and absent of any other controls other BGP speakers will place the same level of trust in the authenticity of this routing information as it does to all other routing items that it learns. This has a number of impacts. Operational mishaps in routing configurations may have major impacts on the services that are provided over the Internet. Deliberate interference may result in misdirection of traffic, which also may result in disruption of services and potential compromise in the proper operation of other services.

The question is: How can we secure the operation of this inter-domain routing process?

There are a number of parts to securing the inter-domain routing process:

The first part concerns the *origination* of prefix information into the network. So, it's necessary to get some answers to some questions about origination. Is this address prefix able to be used in the global Internet? Some addresses are in a *reserved* state, some are *unallocated*, some are intended for use in *private networks*, and some are intended for use in *multicast* networking. Once an address is established as *routable* its necessary to confirm that the holder of this address prefix has granted an authority to the network operator to announce this address prefix into the routing environment. In the Internet environment the entity who holds a network prefix is not necessarily the same entity as a network operator, so it's necessary to refer to the address registry framework to confirm that the entity who is registered as the holder of the address prefix is the same entity that is granting his authority to advertise this prefix into the routing system. This aspect of secured routing is termed *Route Origination Validation*.

The second part concerns the *propagation* of the route advertisement through the inter-domain routing space. Has a routing advertisement been altered during propagation through the inter-domain space in ways that do not conform to correct operation of the BGP protocol? In particular, does the route propagation sequence, as represented by the AS PATH attribute of a route object, match the actual propagation of the route object? This aspect of secured routing is termed *AS Path Validation*.

The third part concerns the *conformance* of the route object to the routing policies of the sequence of networks that have propagated this route object. *Route Leaks* represent a common form of routing anomaly, where the route object has been propagated in a way that is contrary to local route policies. For example, a customer network may leak route objects learned from one transit provider to another.

The last objective here is to ensure that the actual *forwarding* path used by packets to reach their addressed destination is the same as the path described in the route object. This is a somewhat challenging objective in that BGP is intended to aid packet switches to perform local switching decisions in a stateless manner, while the abstract model that is used by the routing process relies on a coordinated state within the network to maintain a path.

Here we are looking at the level of progress in deployment of the first objective here, namely the generation of ROAs and *origination*.

## Measuring ROA Production

The first part of the measurement concerns the extent to which address prefix holders have taken up the option of publishing signed ROAs for all the address prefixes that they intend to have announced into the routing system. In many cases the address prefix holder and the network operator that manages the routing function are the same entity and the generation of these ROAs and the related advertisement of the address prefixes can be performed in a single process. However, a number of networks allow their customers to "bring your own addresses" and in such cases the generation of ROAs by the address hold is a distinct step. It is, in theory at any rate, a task that should not be outsources. The ROA is signed by the private key of the address holder, and if this function were to be outsourced to a third party, the address holder would in effect be passing effective control of the address block over to this third party.

So, this part of the measurement task is to measure the extent to which address holders have published signed ROAs for their routed addresses.

The way we perform this measurement is to use an RPKI client tool to collect the full set of validated ROA objects. The set will (hopefully) be the same for all RPKI clients at any given time so there is no need to perform this operation in multiple locations. A single data set is all we need as, at least in theory, the local RPKI cache maintenance code synchronises the local cache to the network-wide state of RPKI publication point repositories.

We then assemble as large a set of prefix and origin AS pairs that have been announced into the default free zone (DFZ) as we can. The most convenient way to gather this data is from a BGP-speaking router. It is probably best that this is performed in a number of places at much the same time. While the objective of the BGP protocol is to flood routing information across the entire network, this propagation function is moderated by local routing policies, so each individual view of the routed space is somewhat unique to that BGP speaker and its position in the inter-domain topology. The view of the routed space is also moderated by networks that discard route objects that are inconsistent with the information in valid ROAs. This means that a large collection of BGP perspectives should provide a better insight into the routing space and ROAs, and ROAs that invalidate routes in particular. We use this aggregated collection of BGP dumps from the route collectors operated by Route Views and by the RIPE RIS service.

For each address prefix we search through the list of validated ROAs to find a match against the address, prefix length and originating AS. If there is a match, we will label this prefix as *valid.* If there is no match, then we look to find an *invalid* match where there is a match for the address and prefix list but not the origin AS, or where the prefix length indicates a small prefix than that permitted by the ROA. Otherwise, the prefix is labelled as *unknown.*

We also calculate a *visibility index* for the prefix which is a value between 1 and 10 to indicate the proportion of peer networks of the Route Views and RIS collectors that observe this prefix. As the number of networks that perform dropping *ROV-invalid* prefixes increases, the propagation of such prefixes is reduced, and they tend to have much lower values of this visibility index.

We geolocate each prefix to a country and then we can assemble total counts for each country. We count the number of *valid, invalid* and *unknown* prefixes as a combined total, and separately for the IPv4 and IPv4 address families.

We also calculate the *exposed span* of each address prefix, which is the total count of the individual address in each prefix, less the total span of all more specific address prefixes. Because of the differences in the propagation of individual address prefixes, and the intentional dropping of *ROV-invalid* prefixes, this address span calculation is specific to each individual BGP peer for each route collector. We simplify this metric by assuming a situation where every announced prefix from every BGP peer is included in this address span calculation. We can use the prefix geolocation information to generate per-country totals of this metric as well, but here the differences between IPv4 and IPv6 address families mean that a combined total of the address span view is not generated.

As well as per-country views of ROA production, we generate per-network views, looking at the collection of routes originated by each network, again at both an address prefix level and an exposed address span level. There are some 75,000 networks within the Internet, and just a few hundred individual BGP peer of the route collectors. What that means is that when we construct a view of the routes originated by a network it is probably a view constructed at a distance, or a remote view. If any of the paths between a BGP route collector and the observed network perform actions such as invalid ROV dropping, then our remote view is then compromised to some extent. As the level of deployment of

ROV-invalid filters increases, particularly in transit networks, then the capability of this route collector-based measurement to "see" ROV-invalid routes is reduced.

The reports for the ROA production can be found at https://stats.labs.apnic.net/roas.

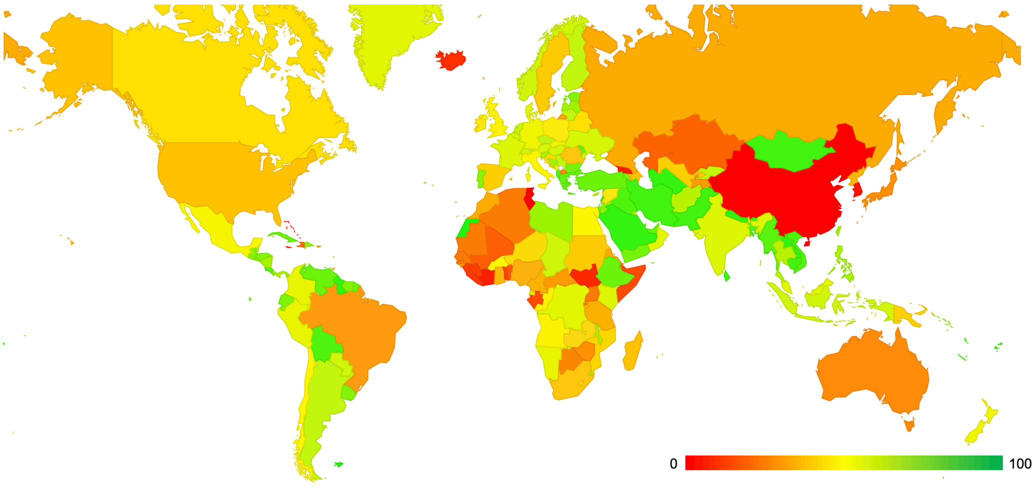

*Figure 1 – Map of the relative level of ROA generation per country*

## Measuring ROV-invalid Filtering

We also measure the deployment of filtering (discarding) of ROV-invalid routes. As already noted, a comprehensive measurement would require a view of the BGP state within every AS, but such a view is not readily available to us. We have chosen to adopt a user-centric measurement instead, leveraging the ad-based measurement that I've described in the previous measurement article (https://www.potaroo.net/ispcol/2023-10/measure-dnssec.html).

Here we use a total of four URL targets in order to provide both control and test measurements.

The first target is a time-variant target where we adjust the ROA at regular intervals such that the route to reach the server is valid or invalid depending on the day of the week and the time of day. The route to the web server has a valid ROA for twelve hours, then the ROA is replaced with one that contradicts the route to the web server. An ROV-aware network should pick up this change of the ROA status for this route and if it is performing ROV-invalid filtering it should drop the route to the server. We then cycle the ROA state again, but this time with a 36-hour period of valid and invalid ROV states. We then cycle the ROA state using a 24-hour interval for valid then invalid. This cycling of the ROV state between *valid* and *invalid* has a two-fold intention: to see which are the networks where the users cannot reach these beacons during their periods of invalid state, and secondly to look at the dynamic properties of the ROV system to see how quickly networks detect the change in the associated ROAs for the route, looking at both *valid-to-invalid* and *invalid-to-valid* transitions.

To undertake this measurement, we first used the VULTR anycast network, where we have control of the BGP routing state for the server. This was configured as a relatively modest four-node anycast platform. As ROV-invalid filtering is increasingly deployed in transit networks our ability to clearly see all the way to the edge network and determine their ROV filtering behaviour becomes clouded by the behaviour of the transit networks on the path.

To improve this situation, we now also use RPKI beacons operated by Cloudflare in their various points of presence. This has expanded the anycast network to some thousands of end points, but with Cloudflare we use permanently valid and invalid BGP routes. Here we have three target URLs in the measurement ad: a valid dual stack URL to determine reachability, a ROV-invalid IPv4-only URL and a ROV-invalid IPv6-only URL. This anycast service has reduced the number of transit networks between users and the beacon points and has provided a clearer (but by no means perfect) indication of whether a network is performing ROV-invalid filtering.

This is a *user-centric* measurement rather than a *network-centric* measurement, and it looks at the proportion of users that cannot reach a destination if the only viable path to that destination entails following a *ROV-invalid* route. The global view of this metric over time is shown in Figure 2.

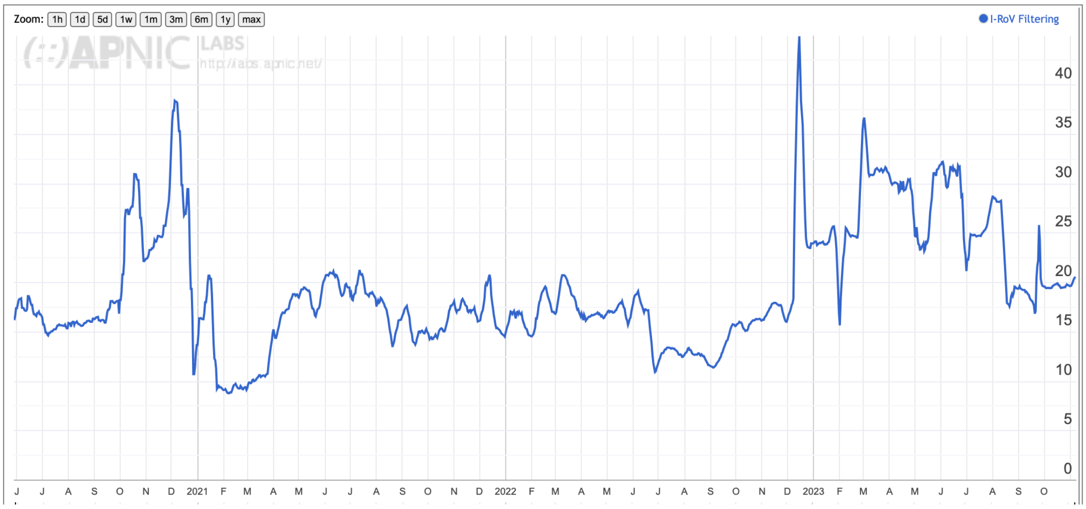

*Figure 2 – Proportion of Internet users who are behind ROV-Invalid Filtering*

It's an interesting measurement result, in that unlike ROA production, this is not an "up and to the right" data series, and the observation that some 20% of users who cannot connect to ROV-invalid route destinations is the same as it was in mid 2020! This seems a little counter-intuitive, but it's likely that as the density of interconnections increases over time, then the impact of a single transit network performing *ROV-invalid* route discard decreases.

We perform the same analysis on a country-by-country basis, and the resultant map of where *ROV-invalid* route discard is taking place is shown in Figure 3.

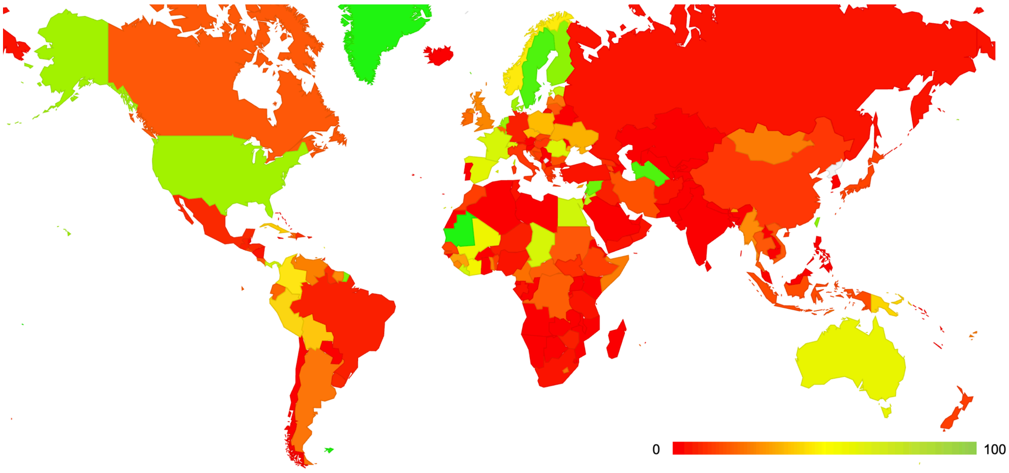

*Figure 3 – Per-Country map of proportion of national users who are behind ROV-invalid Filtering*

The analysis also extends to a per-network level of granularity, but this level of detail focus exposes an inherent issue in the measurement. While a network level measurement may purport to show whether or not the user's network performs *ROV-invalid* route discard, the same outcome is achieved if the network's transit network (or networks) perform *ROV-invalid* route discard. This is the case is any network in the path between the user's network and the network that contains the target point. So, this is not in fact a network measurement but a path measurement.

The task of measuring whether or not as network performs *ROV-invalid* route discard or not can use this same technique, but it requires the anycast cloud that contains the route targets to be located in the eBGP adjacent network to the network being measured, and this needs to be the case for all 75,000 networks! We will need to use an anycast configuration that contains service points in each of the 11,000 such immediately adjacent networks.

This is probably too big an ask. An ideal measurement scheme is to take a target address within the network being measured and create a specific ROA that makes the target address ROV-invalid. A traceroute from the target out to any external point would require the TTL-exceeded messages to enter the network. If the network is *ROV-invalid* route dropping no TTL-exceeded messages would be received by the target from any external points in the traceroute path.

The reports for user reachability of *ROV-invalid* destinations can be found at https://stats.labs.apnic.net/rpki.

## Acknowledgement

This measurement system would not be possible without the generous support of Google in running our advertising campaigns and supporting their unique configuration, for which we extend our sincere thanks.

## Author

*Geoff Huston* AM, B.Sc., M.Sc., is the Chief Scientist at APNIC, the Regional Internet Registry serving the Asia Pacific region.

*www.potaroo.net*