

Measurement Study on the Internet reachability

Internet reachability can be assessed using control-plane and data-plane measurements. However, there are biases in the results of these two measurement methods that are caused by factors such as the use of default routing. Here, we examine reasons for the biases between control-plane and data-plane measurement results, and explain a dual probing methodology that enables more accurate measurements of reachability.

3.1 Introduction

First, let us briefly review the AS (Autonomous System) and BGP (Border Gateway Protocol) terms that appear frequently in topics related to routing. An AS is a network or group of networks under a common administration and with common routing policies. As shown in Figure 1, an AS typically represents a single ISP. However, in some cases an AS may belong to more than one ISP, or conversely, a single ISP may have more than one AS. ASes are allocated a 32-bit value called an AS number, and ISPs are sometimes referred to using this AS number. For example, IJ's AS number is 2497, so IJ is sometimes referred to as AS 2497. A protocol called BGP (Border Gateway Protocol) is used for exchanging routing information between ASes. Each AS is allocated a block of addresses that share the same N bits from the leftmost digit of the IP address. This is called an address prefix, or simply a prefix. BGP is used to exchange information regarding the reachability of the address prefixes belonging to each AS. Additionally, the initial N bits that each AS shares are called the prefix length. When discussing the prefix length of an address it is sometimes referred to as a /N prefix.

The most fundamental service of the Internet is provision of reachability between two given points. However, we still have a great deal to learn about the basic reachability service of the Internet. Researchers and operators rely on two views of reachability to assess it: examining BGP routing information (control-plane measurements), and testing actual reachability using tools such as ping and traceroute (data-plane measurements).

Here, we show that both of these methods are insufficient for understanding reachability of the Internet as a whole, and present methods for better understanding Internet reachability through supplemental measurements. This paper is based on the results of tests carried out by IJ senior researcher Randy Bush together with O. Maennel, M. Roughan, and S. Uhlig. For details regarding the tests covered here, see reference 1 that was presented at the ACM SIGCOMM ICM (Internet Measurement Conference) in November 2009.

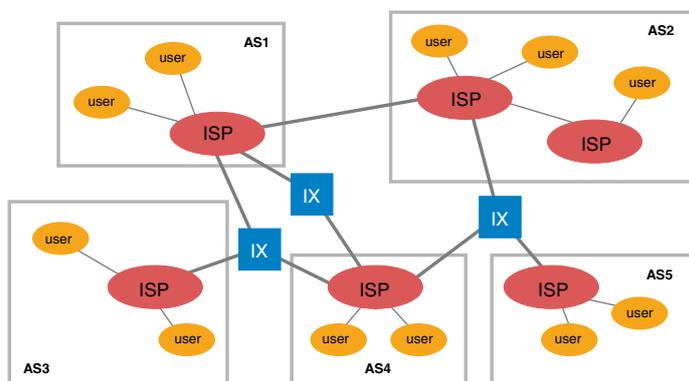


Figure 1: Overview of AS

3.2 How Far does a /25 Propagate?

Anecdotally, most providers filter prefixes more specific than /24 to bound the number of routes in the global Internet in order to reduce utilization of common resources necessary for handling announcements of overly specific prefixes, or to prevent route hijacking. As the starting point for this paper, we sought to test whether such filters are as prevalent as commonly presupposed.

We advertised a /25 prefix from AS 3130 on June 22 2008, and measured its reachability across the Internet via both control-plane and data-plane measurements. At this point, no other routing information containing this /25 existed. The results were so inconsistent that it highlighted the fact that control-plane measurements are insufficient to measure data-plane reachability. We confirmed reachability via control plane by referencing BGP monitors such as RouteViews and RIPE/RIS. As a result, we confirmed that the /25 prefix had been propagated to 11 AS locations. This matched our expectation that a /25 would be severely filtered and would not propagate far.

Our data-plane measurements were based on pinging a large set of IP addresses spread widely across the Internet, using an IP address from the /25 as the source of the ping packets. Receiving a ping-response therefore indicates that the ping target can reach our prefix. No response could mean that the IP might be down, or the pinged IP might not have a path towards the /25-address space, and so we only draw conclusions from the positive responses.

To our surprise, we found 1,024 ASes that had usable connectivity back to our /25. This represented nearly 5% of all the ASes visible in this experiment. This is not significant in proportion to the Internet as a whole, but it is an extremely large figure compared to the results of control-plane BGP routing information.

Furthermore, according to BGP monitors, all ASes containing routing information for the /25 prefix were within 2 AS-hops of AS 3130^{*1}. In other words it did not propagate further than 2 ASes from the source. The solid line in Figure 2 shows AS number distribution for each number of hops. The prefix was announced by AS 3130, which has two tier-1 upstream providers. The /25 did not propagate further than one hop beyond those tier-1s, only reaching the “core” of the Internet.

Moreover, the results of using traceroute toward the pingable target IP addresses to measure the number of AS hops are shown as the blue dashed line in Figure 2. This was not much different from previous results of measuring reachability for the /20 prefix (the red dashed line in Figure 2). The data-plane measurement results indicate that the /25 is reachable from ASes that are further away (up to 4 hops) compared to the BGP monitor results (up to 2 hops).

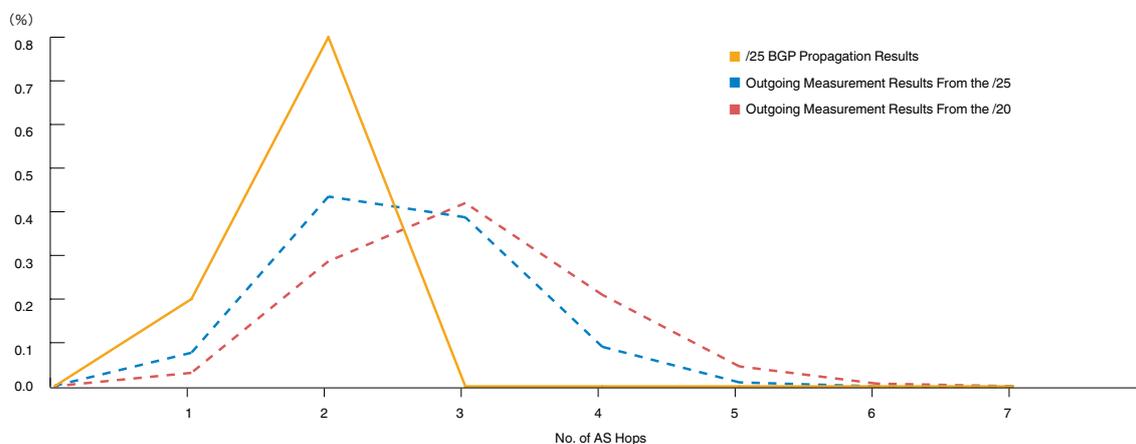


Figure 2: Distribution of the Number of AS-hops to the /25

*1 In this report, we have unified the method of measuring the number of hops with section 3.3.2, instead of using the method in reference 1.

These results show a clear difference between control- and data-plane measurements of reachability. However, data-plane measurements show real reachability, so they need to take precedence. There are two likely reasons for the difference:

- The prefix propagated further than expected on the control plane to sites which were not visible from the standard BGP monitors
- The default routes provided effective connectivity to some ASes despite the fact that they never learned of our prefix

Over 75% of those ASes with data-plane reachability were stub ASes*². Default routing is presumed to be more common in stubs, so we sought to investigate this cause further.

3.3 Utilization of Default Routing in the Internet

Here, we use AS-path poisoning to measure the extent to which default routing is used. As illustrated in Figure 3, our test box in AS 3130 announced a set of test-prefixes to its upstream tier-1 provider. We announced these prefixes with paths containing the AS number of an AS we want to test for reachability to AS 3130, so we can be sure this AS does not have this prefix in its BGP routing tables.

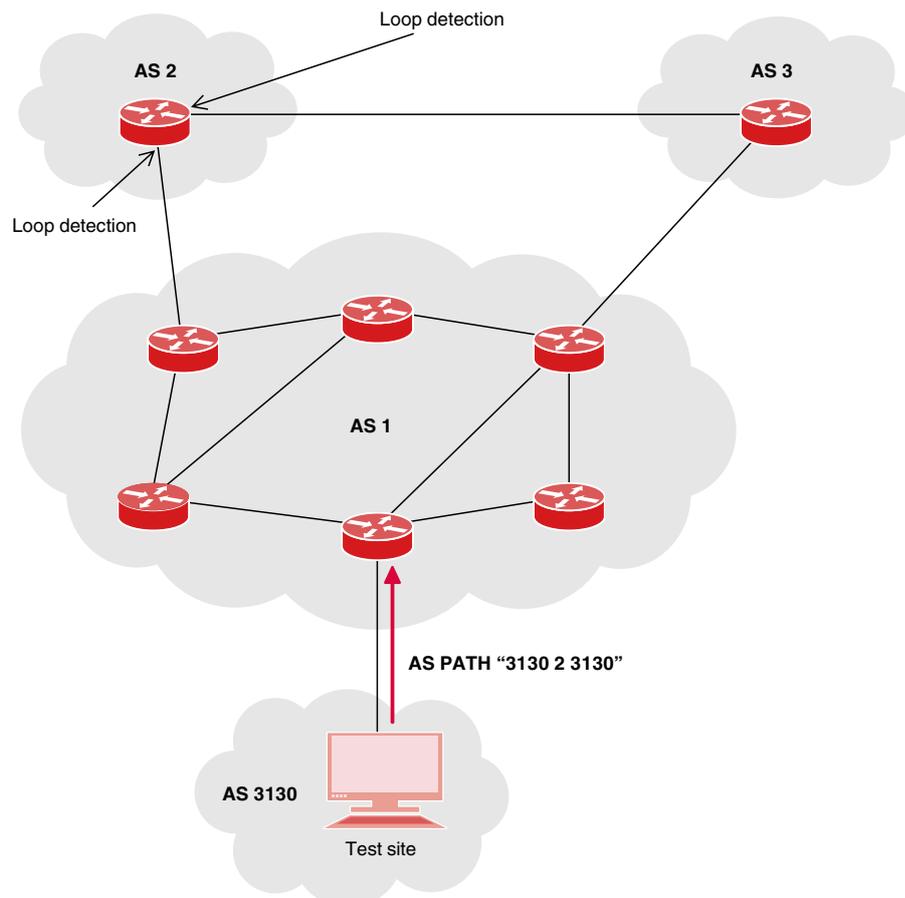


Figure 3: AS Path Poisoning

*2 ASes that do not relay communications from other ASes are called stub ASes. Meanwhile, ASes that relay communications from other ASes are called transit ASes.

For example, if AS 2 is to be measured, we announce routing information containing the “3130 2 3130” AS path. When AS 2 receives this routing information, it sees its own AS number in the AS path, and drops the announcement because of BGP loop prevention. So, as long as AS 2 has no default route, it is possible to create an environment where it cannot reach an IP address inside this test-prefix. We call this method AS path poisoning. AS path poisoning tests were conducted from April 18 2009 until May 1 2009. During that 13 day period, we tested 25,780 ASes for their use of defaults. For the experiment we used the address space 98.128.0.0/16 sliced into /24's, and carried out concurrent tests using the following steps covering a large proportion of the Internet.

1. We check in advance that the target AS is reachable from an address with an unpoisoned prefix.
2. Next, we withdraw the test-prefix, and wait for 1.5 hours to reduce the possible influence of route flap damping.
3. We then announce the test-prefix containing the AS number of the target AS, and wait for 20 minutes to allow it to propagate.
4. We then start testing the target AS by sending pings from the test-address space to our set of target IP addresses in the relevant AS.

A typical test run takes between 2 and 3 hours. We tested a large number of ASes by conducting the tests in parallel. We also compared test probes of the target IP addresses from unpoisoned address space to check that these addresses were consistently available over the course of the whole experiment. The results showed that the vast majority of ASes (99.2%) were consistently reachable.

The results showed that 64% of all IP addresses tested were still reachable after their AS path was poisoned. We tested multiple IP addresses per AS, and found that 74.8% of ASes (19,291) answered consistently despite the poisoning. In other words, the majority of ASes have a default route configured.

Of the remaining ASes 20.9% (5,381) did not ever answer, and 4.3% (1,108) answered for some IP addresses, but not others. We noticed that some probes to the test address space failed, but this was a very small percentage (0.7%). This is thought to be due to bogon-filters.

We interpret a non-response as the AS being default-free, though this interpretation is less certain as we cannot say that no IP address in the AS ever uses default. The mixed result category reveals the complexity of network management practices for the target AS. For example, one AS apparently does not use BGP default routing, but has a default route manually configured to some routers for IP-TV and VoIP services. This illustrates that some ASes are not operated under a unified policy. Interestingly, there seems to be cultural differences in the use of default routing. In one test, results indicated that 60% of Japan ASes did not use default routing, while 36% did, and 4% had a mixed configuration.

The results of these tests were published on our website, and we took feedback from the ASes that were tested. Of the 191 ASes that replied, 94% confirmed that the results of these tests were correct. Additionally, some of the addresses in the IP address list that we pinged belonged to an address block delegated by that AS to another AS. Surprisingly, some AS administrators were not aware that they were using default routing. This can happen in cases such as when default routing from an upstream provider is accepted without applying a filter.

3.3.1 The Impact of AS Type

Intuitively, we might expect that ASes that provide transit to other networks will be less likely to use default routes than stub ASes. We tested this by breaking down our previous results by AS type. Here, we used the AS classifications provided by reference 2.

Table 1: Default Utilization Distribution by AS Category

	# Tested	Default	Default-free	Mixed
Stub	24,224	77.1%	19.3%	3.6%
Small ISP	1,307	44.5%	42.2%	13.3%
Large ISP	246	17.1%	60.6%	22.3%

As Table 1 indicates, utilization of default routing declines when going from stubs, to small ISPs, to large ISPs. The number of ASes with mixed results for default routing increases from stubs to larger ISPs. This suggests that the operation of larger ISPs is more complex. However, during our tests pings were sent to more IP addresses for larger ASes, so this must be taken into consideration when interpreting the results.

Figure 4 shows a breakdown of the test results against the number of peers with other ASes. This demonstrates that use of default routing declines until having at least 100 AS peers. Additionally, 80% of ASes with 20 or fewer peers rely on default routing, and ASes with 300 or more peers use default routing in less than 15% of the cases.

The uneven popularity of default routing in different types of ASes shown in the test results is extremely interesting. For example, when using a traceroute from a stub AS, it may be able to travel for the first few hops without explicit routing information using a default route, but when it reaches a large ISP the default route drops off, and it stops there. However, this does not necessarily mean that there is a problem with the location the traceroute has stopped at. The fact that the traceroute made it that far differs from the reachability data acquired through control-plane information, suggesting that neither data-plane nor control-plane measurements are adequate by themselves.

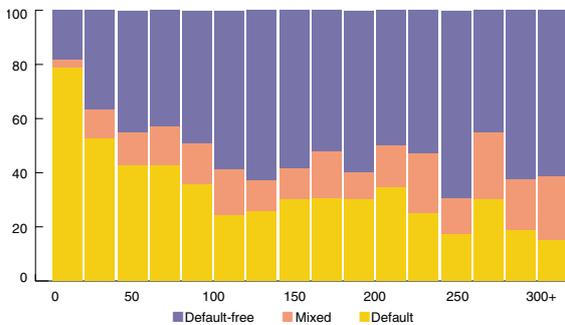


Figure 4: Default Utilization by AS Peer Numbers

3.3.2 The Impact of Defaults

In order to determine the impact that default routing has on the measurements of the Internet, we carried out simulations using the AS topology data from reference 2. In our simulations, we used the default routing utilization ratios from our test results, and assigned default routing to 77.1% of stub ASes, 44.5% of small ISPs, and 17.1% of large ISPs within the topology data. We included mixed cases with the default-free results. We adopted two methods for determining where the default route points for ASes with defaults assigned. The first was a random allocation scheme in which we chose randomly from an AS's upstream providers, and the second was a max allocation scheme in which we chose the AS's upstream provider with the maximum number of customers.

For each simulated topology we chose a random set of 1,000 sources. For each source, we then calculated how many potential destination ASes could be reached from this source using default routes.

The results showed that if we use only defaults, then we can reach only a very few ASes. The Internet hierarchy is relatively flat, so even from a small stub AS, we need only go up through a few layers of providers before we reach a large, or tier-1 provider that doesn't use default. We found that typically only 1-3 ASes could be reached in this way, and that the maximum was 5.

Here, we look at cases where routing information advertisements for a prefix are only propagated to the upstream provider one hop away, and not propagated to ASes beyond that.

Figure 5 shows the Complementary Cumulative Distribution Function (CCDF) of the number of ASes that can be reached from a random source. This graph shows that for the max default allocation we can now reach 1,000 ASes from approximately 50% of sources, and over 2,000 ASes from around 1/3 of sources. For the random allocation of default routes, we can reach somewhat fewer destinations, but the number is still substantial.

Figure 5 also shows the results of the assumption that routing information advertisements propagate two-hops. We can see that this has a dramatic impact on the number of ASes that are reachable, with 6,000 reachable about 50% of the time. There were also cases where up to 19,000 ASes were reachable.

As we observed with the propagation for the /25 prefix from 3.2, it does not actually follow a simple "hop-count" mechanism, and local filters at each AS must also be considered. However, the simulations do provide us with some valuable intuition. It is quite possible that once the /25 reaches our provider, and perhaps a few other ASes, that it will be reachable from a significant proportion of the Internet despite the limited propagation of its routing announcements. We believe these results also shed much light on the phenomena of reachability using data-plane measurements when there is no reachability using control-plane measurements.

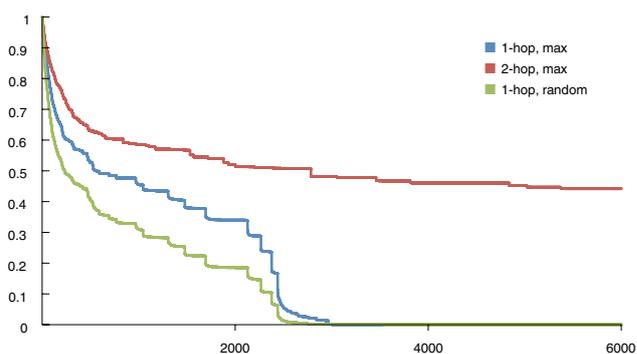


Figure 5: Distribution of Reachable ASes

3.4 Testing Reachability with Dual Probing

The existence of default routing indicates that there is a limit to the predictions that can be made through control-plane observations. Such limitations should be kept in mind before making claims based on control-plane observations alone. On the other hand, the types of data-plane measurements we have used so far are limited as well. It is easy to find situations where it is hard to interpret the results of ping probes simply because end-host (or something in the middle) behavior is so varied. Reachability is hard to measure because the following two viewpoints must be considered.

- How do I see the world?
- How does the world see me?

The first is based on the information a router receives from routing protocols. The second question is something operators would often like to know in order to debug reachability problems. In other words, how does a network appear from other parts of the Internet? Unfortunately, this information is not directly available from the network layer.

There is data available to see how the world sees us. Services such as BGP monitors, looking glasses, and traceroute servers provide public views of the Internet. However, only a sample of ASes operate these as a public service, therefore it is hard to get direct data from the world viewpoint. What we see when we combine data from the available viewpoints is actually a sampled world viewpoint.

A significant problem with this sampled view is that the operators with the sophistication and resources to operate public viewpoints tend to be larger ISPs, nearer the “core” of the Internet, so the bias in the viewpoints could mislead. For instance, we might hypothesize that these large, densely connected ISPs have fewer reachability problems than stub ISPs, so they are not useful for evaluating stub ISP reachability. There is therefore a need for methods that allow reachability to be tested from a variety of viewpoints encompassing the entire Internet.

Here, we advocate a data-plane testing method called dual probing that can be applied to a wider range of situations. Suppose a network administrator wants to check that external hosts can reach their network. A simple test would be to ping from a machine towards a large set of external IP addresses covering much of the Internet. If those IP addresses answer the probes, this indicates that the source machine is reachable from the probe’s destination. We call these “out-probes.”

Figure 6 illustrates the out-probe concept. In the “traditional” case (black solid arrow), a probe is in-bound from a public looking glass towards the test-site. In the out-probe case (green dashed arrows), a probe is sent from the network for which reachability is being tested towards many destinations in the Internet. Note that the address space under investigation must be the source address of the outgoing IP packet. In this case, it is the return traffic towards the test address space that reveals the reachability of the test IP address space.

When the ping probes are not answered, it could be due to one of the following reasons.

- The IP address simply does not answer pings.
- The ping packets are dropped by firewalls on the way towards the probed IP addresses.
- The IP addresses answer the ping probes but the answers are dropped somewhere on the path back towards the probing host.
- There is no route from the IP address in question back to the probing host or vice versa.

Only the latter two cases concern reachability of our test network. However, even the third cause may be a poor indication of unreachability, because ICMP probes are often given lower priority and may be dropped preferentially over TCP traffic. At any rate, negative responses provide little information by themselves.

If we can calibrate our expectations we will know how to interpret the responses of probes. If actual tests are carried out after first conducting tests in advance similar to the methods used in “3.3 Utilization of Default Routing in the Internet,” it is possible to gain useful information by comparing the expected results with the actual ones. In other words, using two probes separated in time allows for better interpretation of the second probe. This is also possible through using multiple IP addresses for the ping target. We call this approach dual probing. Although “dual” is used in the name, in some cases more than two probes may be involved.

Using dual probing, we can compare probe answers against probes from another prefix, called here an anchor-prefix. The anchor prefix is an old, well-established prefix known to have very good reachability. This comparison reveals far more information than a single probe from the test-prefix. Lacking a reply from the anchor probe to a particular IP address we know there is a problem probing this IP address, and so we can discount test measurements as not useful. With a reply to both, we can infer successful reachability. When we receive a ping reply to an anchor probe, but no reply to a test probe, we have evidence that there is a reachability problem somewhere between the IP address of the ping target and the test site. The evidence is not conclusive, as ICMP packets may be dropped, but over a series of such measurements we can build confidence in the results.

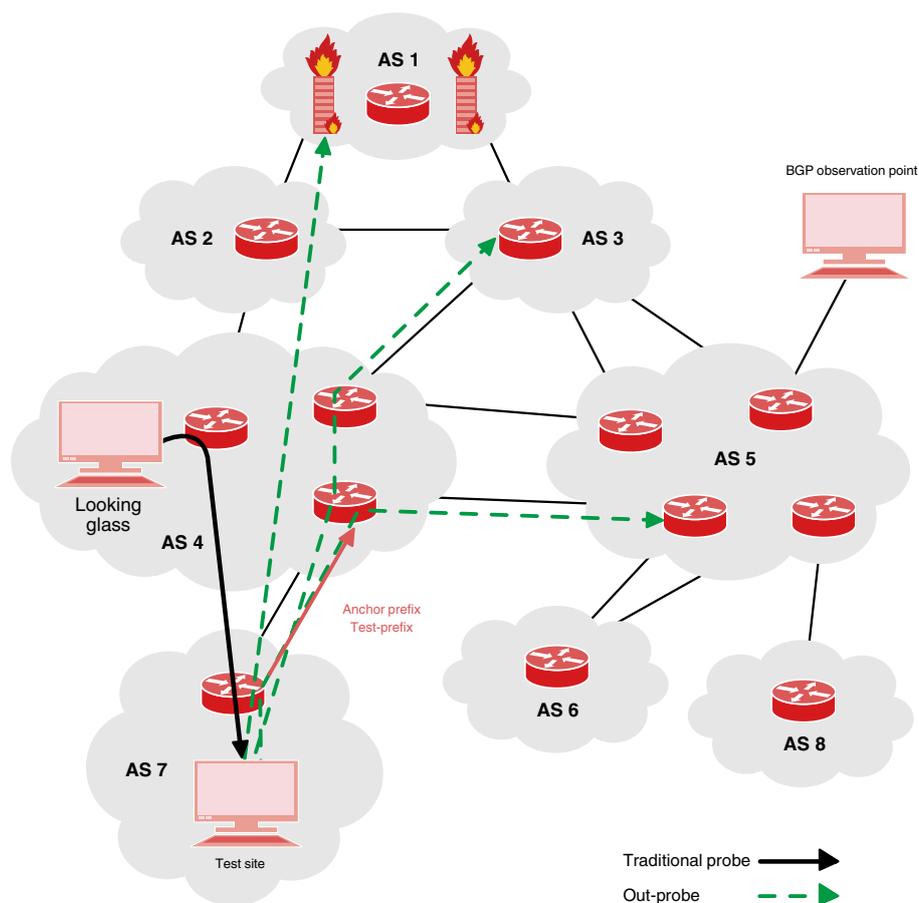


Figure 6: The Dual Probing Concept

3.4.1 Bogus Bogon Filter Detection

A bogon refers to a bogus routing announcement. These are sent either accidentally, or deliberately to hijack address space. Consequently, ISPs commonly configure either control- or data-plane filters to prevent traffic to/from obviously bogus addresses. For instance, it is common to set a filter to refuse traffic or announcements from prefixes that have not yet been allocated to ISPs by the registry. However, the configuration of these filters does not always keep up as prefixes are allocated, so reachability problems to certain prefixes can sometimes occur. In the past it was difficult to detect the location of these bogus bogon filters. We conducted tests to detect bogus bogon filters using dual probing.

ARIN allocated two large segments of new address space (173.0.0.0/16 and 174.128.0.0/16) for our experiment, and we used five smaller segments of this address space. We announced those prefixes from five different locations that volunteered to participate in our experiment: PSNet in Seattle (USA), Verio in Ashburn (USA), SpaceNet in Munich (Germany), CityLink in Wellington (New Zealand), and IJ in Tokyo (Japan). The anchor-IP was the normal address of a machine inside the ISP that ran the experiment. A test IP address was selected from the test-prefixes and configured as a secondary IP address on the same interface.

We ran three different measurement campaigns: the first starting in April 14 2008, the second starting on May 27 2008, and the third starting on June 12 2008. About a week was necessary to run all our probes. The first measurement campaign occurred before ARIN announced that this address space had been issued. The goal of these measurements was to understand how prevalent legitimate bogon filters were. After the first campaign, ARIN announced that the address space had been issued, and that it should be removed from bogon black-lists. In addition, we sent emails to the operators of ASes that we identified as having a filter set in the first campaign, and asked them to remove the filter. Thus, reachability problems identified in the second set of measurements are genuine problems, which needed to be fixed. The third set of measurements was used to assess how the reachability problems were changing over time.

For an AS to be identified conclusively as having a reachability problem, we required zero returns back to the test-IP, and at least five returns to the anchor-IP. If there were zero returns to the test address space, and less than five returns to the anchor IP, we considered this an indication that the AS might have a problem.

Figure 7 shows the results of these tests. The black solid line represents all ASes. The green dashed line represents ASes that had no problems. ASes for which a bogon filter was detected are indicated by the red dashed line, totaling approximately 500. The nearly 2,000 ASes represented by the yellow dashed line also potentially had a filter set.

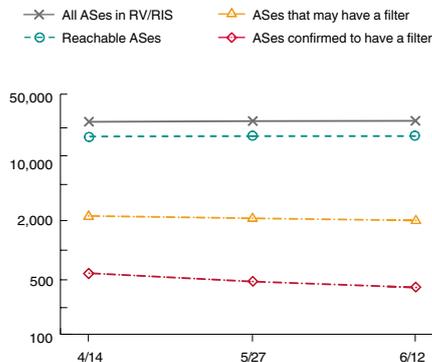


Figure 7: Bogon Filter Detection Tests

These results mean that between 2% and 7% of ASes across the whole Internet could not see the newly allocated addresses. We can also see that the problem barely changes between our second and third tests. This is a serious problem.

Figure 8 shows the distribution of ASes with a filter confirmed or with a filter potentially set by AS type. We can see that the vast majority are stub ASes. This suggests that most problems occur at the edge of the Internet. However, it is possible to incorrectly identify a problem as related to a stub AS when it is actually related to a transit AS, so this must be considered.

3.5 Impact of Methodological Issues on Measurement Confidence

So far we have shown how observations from the control plane do not indicate reachability correctly, and that the data plane can offer a different perspective. However, data-plane observations also have their own limitations. In this section we briefly discuss three methodological issues that should be taken into account for data-plane measurements.

3.5.1 Topological Coverage

The motivation for out-probes is to cover areas of the Internet where no BGP monitors or looking glasses exist — in other words, to look at reachability at the edge, rather than the core of the Internet. To achieve this, it is necessary to create an IP address list that covers the entire Internet, such as that used in “3.4 Testing Reachability with Dual Probing.” This address list must have wide coverage, and be able to probe inside ASes for non-homogeneously configured parameters, in addition to limiting the number of probes that have to be sent. The quality of the IP address list that is created affects the quality of actual measurements.

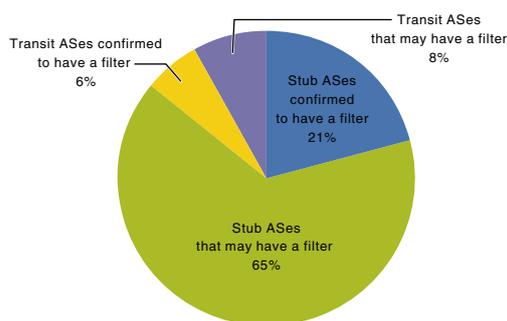


Figure 8: Distribution by AS Type

3.5.2 IP to AS Number Mapping

A general issue we encountered during all experiments is the mapping of IP to AS number. BGP routing table look-ups are used to achieve this. However, when a transit provider connects a customer, an IP address may be provided from an address block allocated to the provider. When this happens, if the router at the border of the customer AS has no reachability to certain prefixes due to a bogon filter, etc., there is a possibility that it may be detected as having no reachability from the upstream provider.

After creating correct IP address to AS number mappings, it is also important to maintain them. For example, comparing mappings we created in 2007 to those we created in 2009, only 88% of the prefixes were still mapped to the same AS. When there is an error in IP address and AS number mapping, observations may be interpreted incorrectly.

3.5.3 What Type of Probe to Use?

The decision of whether to use ping or traceroute for data-plane measurements is also crucial. Moreover, when relying on ping probes, the type of packet (ICMP, UDP, or TCP) used can have significant impact. When conducting our tests we found that for ICMP, roughly 70% of the IP addresses were reachable. For UDP, the responsiveness was only 30%. This is because it is filtered by firewalls and NATs. For TCP the responsiveness was even worse, with around 5% reachability.

3.6 Conclusion

In this paper we showed that assessing actual reachability on the Internet is more complex than it appears from publicly available BGP server data, and that data-plane reachability is different from control-plane visibility. We also demonstrated that packet reachability is provided through default routing even when routing information does not propagate. Additionally, we proposed new methods for verifying reachability using AS path poisoning and dual probing. IIJ will continue to strive for the stable operation of our Internet backbone, in addition to carrying out tests and disseminating information related to the stable operation of the Internet as a whole such as covered in this paper, so the Internet can function as a safe and stable social infrastructure.

References

1. R. Bush, O. Maennel, M. Roughan, S. Uhlig, "Internet Optometry: Assessing the Broken Glasses in Internet Reachability," ACM SIGCOMM IMC, 2009.
2. R. Oliveria, B. Zhang, "IRL - Internet Topology Collection," 2009.

Author:

Toshiya Asaba

President and CEO, IIJ Innovation Institute Inc. Mr. Asaba joined IIJ in its inaugural year of 1992, becoming involved in backbone construction, route control, and interconnectivity with domestic and foreign ISPs. Asaba was named IIJ director in 1999, and as executive vice president in charge of technical development in 2004. Mr. Asaba founded the IIJ Innovation Institute Inc. in June 2008, and became president and CEO of that organization.