RIPE Labs

RIPE NCC and Duke University BGP Experiment

On 27 August 2010, the RIPE NCC's Routing Information Service (RIS) was involved in an experiment using optional attributes in the Border Gateway Protocol (BGP). As a result of this experiment, a small, but significant percentage of global Internet traffic was disrupted for a period of about 30 minutes. The following article provides some background information on the experiment itself and its effect on the network.

Background on RIS Experiments

As part of its mission, the RIPE NCC works with other members of the Internet technical community to contribute towards the secure and stable operation of the network. The RIPE NCC Routing Information Service (RIS) has a long tradition of supporting Internet researchers.

Since 2002, the RIS has announced a set of beacon prefixes [http://www.ripe.net/ris/docs/beacon.html] . These prefixes are announced and withdrawn at predictable times, to assist in propagation and flap dampening research. In 2007, the RIS was the second network in the world to start announcing a prefix from a 4-byte AS Number. This helped operators test their 4-byte AS capabilities and allowed us to measure the effectiveness of the transition mechanisms for 4-byte AS Numbers.

The announcements made by RIS are also a vital part of the De-bogon Project [http://ris.ripe.net/debogon/] , with RIS measuring the visibility of former bogon prefixes. We have also done measurements on traffic attracted after announcing 1/8 [../franz/content-pollution-18] , work later extended by APNIC.

The Experiment

A research group at Duke University in the United States approached the RIPE NCC for help with experimental research. This group is working on a secure Border Gateway Protocol (BGP) design, in which optional transitive attributes are used to propagate some of the certification information. In order to estimate the feasibility of such a design, they asked the RIPE NCC to announce a route resembling their design from the RIS network.

The design of BGP allows routes to have an attribute that is not recognised by the BGP implementation. If this attribute is set as transitive, it is passed to other routers, without intermediate routers understanding what it actually means. This aspect of the protocol has been key for the transition to 4-byte AS Numbers.

This ability of the BGP protocol allows some implementations to support a new feature, while others do not yet understand the contents of the attribute. In the design proposed by the team from Duke University, upgraded routers add certification information and verify certificates from other routers, without affecting the rest of the Internet.

As the researchers did not have their own AS Number or address space, they provided the RIPE NCC with a patch to Quagga [http://quagga.net] , the BGP software used by RIS. This allowed us to run the experiment from our infrastructure. We checked the patch for security or protocol problems.

In addition, all announcements were sent through another Quagga instance, so that any protocol violation would be noticed before the announcement went to the Internet.

Issues Encountered During the Experiment

To run the experiment, we installed a custom Quagga instance announcing the route through the RIS collector connected to the Amsterdam Internet Exchange (AMS-IX [http://ams-ix.net]) and Groningen Internet Exchange (GN-IX [http://www.gn-ix.net]). We started the announcement at 08:41 (UTC) on Friday, 27 August 2010. It was originated from AS12654, using the prefix 93.175.144.0/24.

The attribute used by the RIS had never been announced on the Internet before, although it was in accordance with the BGP specification.

The announcement was withdrawn, as planned, at 09:08 (UTC). Shortly after, we discovered that the experiment had caused a negative impact on Internet operations that lasted for approximately 30 minutes.

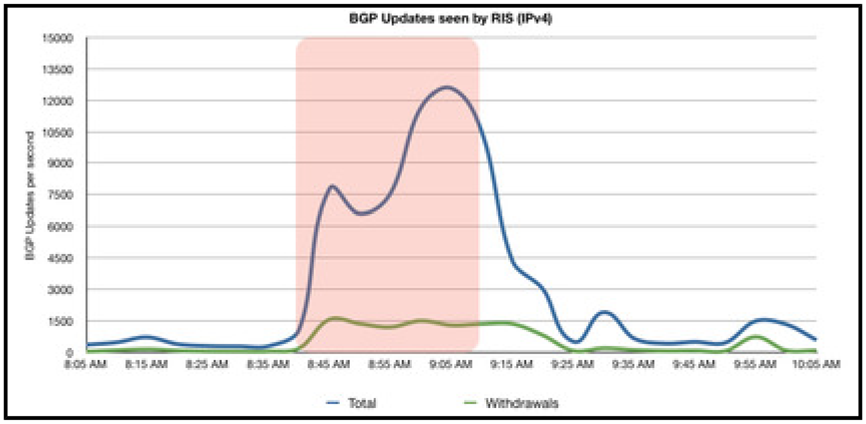
We immediately started an investigation, using input from the affected operators. The investigation indicated that the attribute had triggered a bug in some Cisco router models, which corrupted the announcement and sent this on to other routers. Their peers recognised the corruption, and dropped the peering session.

We provided Cisco with all of the information that we had collected and they released a security advisory [http://www.cisco.com/en/US/products/products_security_advisory09186a0080b4411f.shtml] the same day. The data collected during the announcement was preserved for processing by the researchers from Duke University.

Impact of the Experiment on the Internet

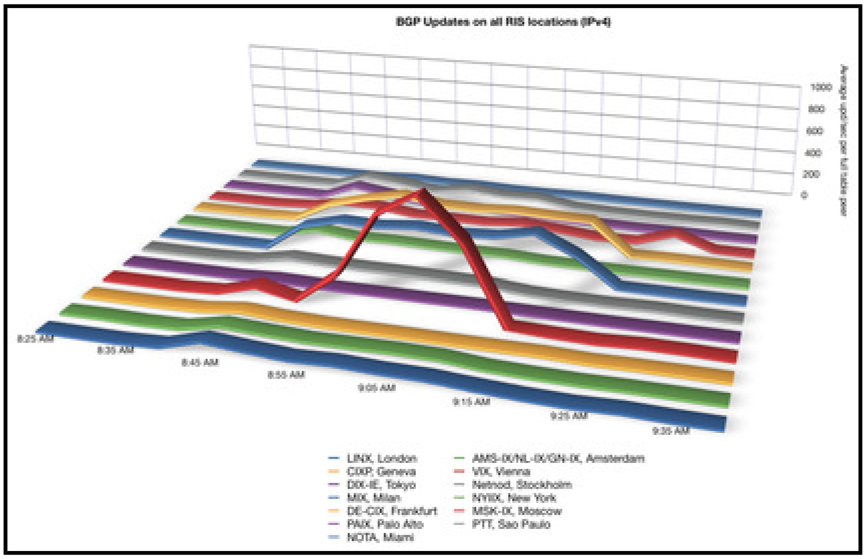
The following is an analysis of the impact of the experiment, using the data provided by the RIS and other RIPE NCC services.

The graph below shows the rate of updates (changes in routing) seen by RIS around the time of the experiment. We can see up to 20 times as many updates, indicating massive instability in the routing system.



[TotalUpdates.png]

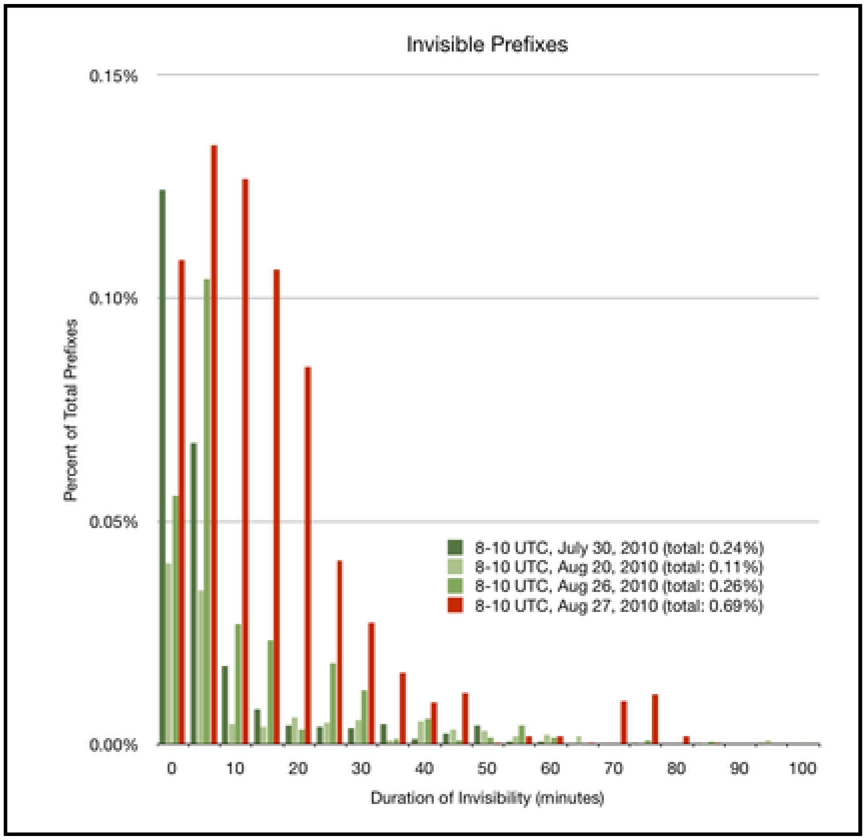
Looking at the data for each Remote Route Collector (RRC), we can see that the effects of the experiments were much stronger in some specific locations. The collector in Vienna registered many times more updates per peer than all other collectors. This may indicate that this region had a higher amount of affected routers.



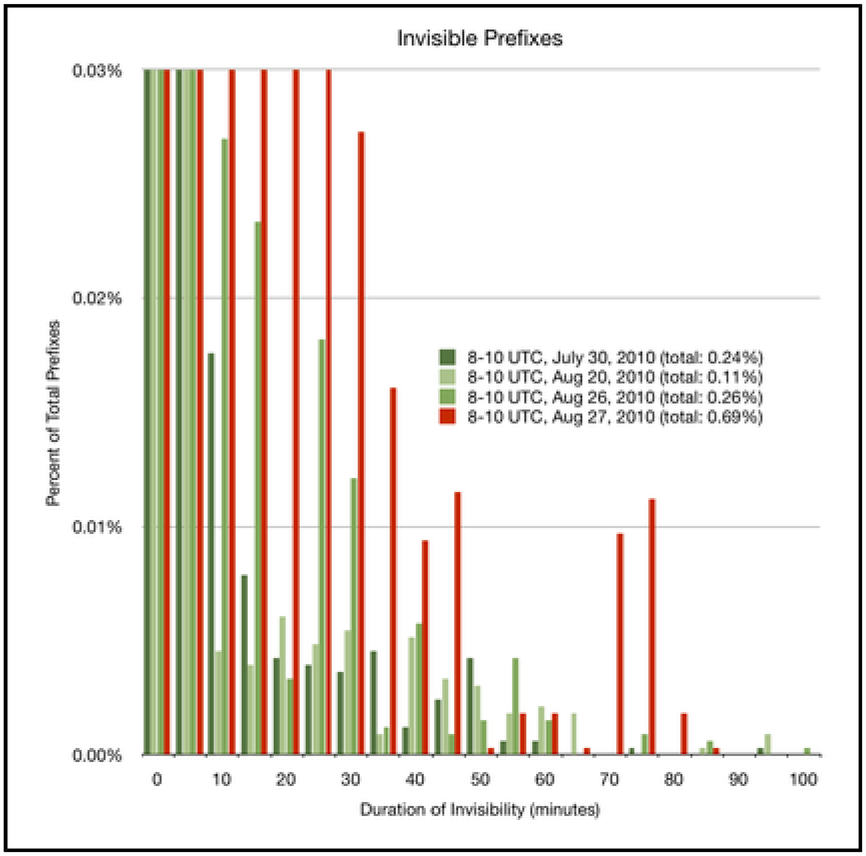
[Updatesperrrc.png]

Knowing that the experiment had a significant effect on the routing system as a whole, we've attempted to look at how much of the Internet was actually affected. A first step is to look at prefixes being withdrawn from the Internet. We have measured this around the time of the experiment and used three reference sets for comparison.

The graph shows the percentage of prefixes on the Internet that became invisible for a certain period around the time of the experiment. There is a large variance in the dataset, with the values for very short outages in the reference sets affecting between 0.04% and 0.13% of all prefixes on the Internet. Overall though, and especially looking at outages longer than 30 minutes, the values during the experiment were up to three times higher than usual. We conclude that the experiment caused an additional 0.5% of the prefixes to become completely unreachable, and to be unreachable for a longer period than they would have under normal conditions.



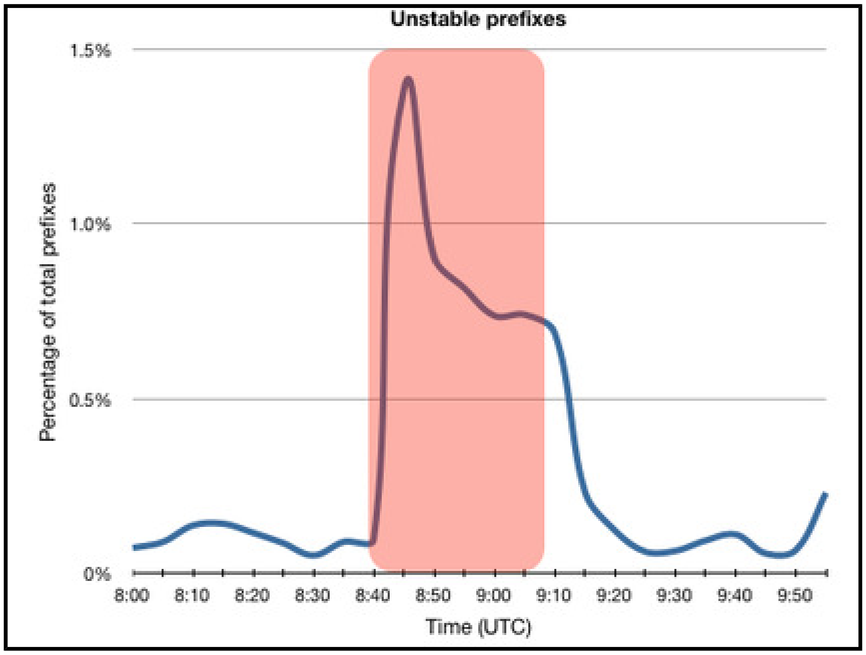
[invisibleprefixes.png]



[invisbleprefixeszoom.png]

Another way of looking at how much of the Internet was impacted is to look at the number of unstable prefixes. For this measurement, we consider a prefix unstable if we see more than 100 updates in a 5-minute period.

The graph shows that under normal conditions, less than 0.1% of the prefixes on the Internet are unstable. The experiment caused this to hit a peak of 1.4%, which amounts to almost 4500 prefixes, about nine times more than usual. For reasons unknown to us, this spike quickly fell to about 0.8%, and stayed there for the remainder of the experiment. About 20 minutes after the experiment, most prefixes returned to normal.

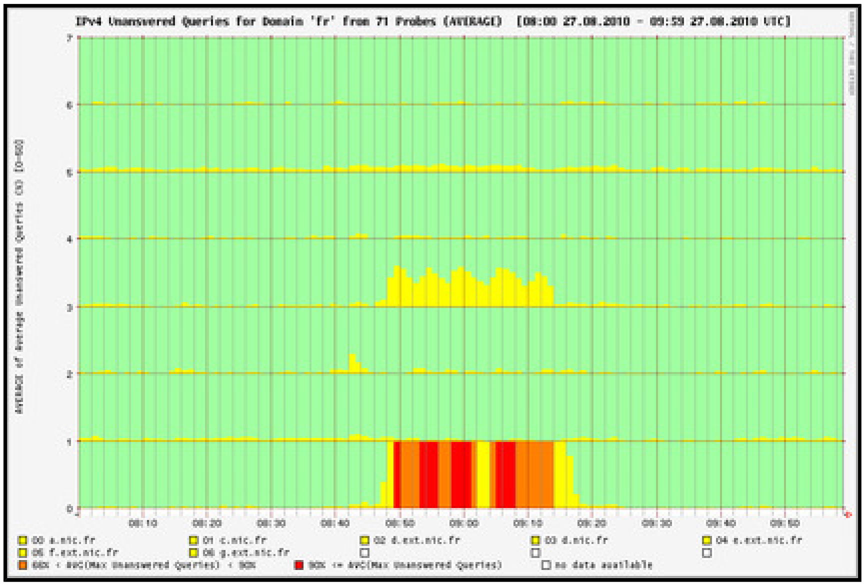


[unstableprefixes.png]

The effect of the experiment on major DNS servers was very limited. The RIPE NCC DNS Monitoring Service (DNSMON [http://dnsmon.ripe.net/]) monitors DNS servers for the root and many Top Level Domains (TLDs) from probes worldwide.

None of the root servers were affected. Minor problems, like a few dropped queries for a few of the probes on just one or two of the DNS servers, were observed in about 15 monitored domains, including the .com domain. We believe that users would not have noticed this. For 63% of the domains monitored by DNSMON, no extra queries were lost.

Noticeable problems were seen for the Slovenian and French TLDs, .si and .fr. In the case of .fr, two DNS servers became almost completely unreachable. However, the other five name servers for the TLD showed no effects, so this will not have caused anything more than some additional delays for users.



[dnsmonfr.png]

Some Conclusions

The experiment caused a massive increase in routing instability, but with different strength in different locations. It caused about three times more prefixes to have periods of invisibility, for longer periods. In total, up to 1.4% of the Internet was affected by instability around the time of the experiment.

The DNS servers for vital Internet infrastructure, such as the root and TLDs were not widely affected.

Final Results

Disruption to the routing system was limited to a relatively small subset of Internet traffic, and the event drew attention to a software bug for which the vendor has now issued a patch. Through a coordinated effort, the situation was quickly recognised and corrected by network operators and those conducting the experiment.

The disruption caused is regrettable, and future experiments conducted with the cooperation of the RIPE NCC will need to meet far stricter internal guidelines, including comprehensive impact assessments, prior announcements with sufficient lead time for Internet operators, and the responsible handling of detected vulnerabilities.

21 Comments

Experiment

Posted by Jared Mauch at 31 August 2010 15:39

Thanks for the detailed analysis. While the impact is regrettable I feel ripe bears little to no blame for a well formed bgp announcement.

Scope of impact?

Posted by Zoe OConnell at 31 August 2010 15:56

I haven't really spent time dissecting the bug so I have little to add there, but I suspect you may not have access to full figures showing the scale of the problem. Rather than looking at unstable prefixes, have you looked at traffic levels at various IXes showing, in the case of LINX and AMS-IX, a 10-20% drop in traffic levels?

It seems the bug may have had more of an impact on iBGP sessions than eBGP ones due to the nature of service providers networks plus it also had an apparent major impact on the network of BT, the local ex-monopoly telco in the UK and affected service to a large number of DSL (And oddly, 3G) services.

Scope of impact?

Posted by Erik Romijn at 31 August 2010 16:35

We have looked at the traffic levels of IXPs. Particularly AMS-IX showed a noticeable drop, about 100 Gigabit/s. This was mostly caused by a specific network operator with very high traffic volumes. However, we chose not to include them, as we can not determine whether this traffic was stopped, or rerouted over a different network.

Cisco IOS XR Bug

Posted by Andre Oppermann at 31 August 2010 15:56

I agree with Jared. RIPE is not to blame. This bug in Cisco IOS XR should have been detected within Cisco's QA department and the buggy release should never have shipped. It's a particular embarassing bug affecting a core feature of BGP.

The patch

Posted by Daniil Baturin at 31 August 2010 16:19

Thanks for the report, very interesting to read.
Will the patch used in the experiment be public available? I would like to do some lab testing with my systems and check how do they deal with those attributes.

The patch

Posted by Erik Romijn at 31 August 2010 16:36

The patch will not be available at this time, as it can pose a security risk.

RIPE do need stricter guidelines for experiment

Posted by leon at 31 August 2010 16:21

Something should be tested in some internal or isloated environment before being pushed to the Internet

No.

Posted by Richard Hartmann at 31 August 2010 16:42

This does not make any sense. It is literally impossible for RIPE to lab test every possible outcome. It's clearly the vendor's liability to test their stuff properly. Randomized input testing has been around for decades and Cisco should use it.

Hardening against random input...

Posted by Richard Hartmann at 31 August 2010 16:40

I agree with Jared; this is not RIPE's fault at all.

I appreciate the openness and detailed analysis. This is an important showcase of why vendors must harden their devices against garbage input. As randomized testing is hardly a new field, this bug does not shed good light on Cisco. Fortunately, we were not affected in our backbone even though some of our upstreams had massive problems.

Hardening against random input...

Posted by sspies at 31 August 2010 17:42

It's no garbage input. The path attribute was well-formed, but was treated incorrectly.

who's fault?

Posted by Kiall Mac Innes at 31 August 2010 16:56

Really - I see this as cisco's fault rather than RIPE. Experiments like this are necessary for the ongoing advancement of the internet!

On the other hand.. Cisco and friends could setup a research network with groups like RIPE having access to it..

Thanks for the experiment and the report

Posted by Stéphane Bortzmeyer at 31 August 2010 17:03

I agree that RIPE and Duke U. are not at fault. Testing is a necessary and very important part of the security and the stability of the Internet. People who criticizes RIPE-NCC or Duke U. just want to shoot the messenger, while we should read the message instead.

And thanks for the very good and detailed report.

Thanks for the experiment and the report

Posted by Jan Marius Evang at 31 August 2010 20:28

I agree that this is not RIPE's fault and RIPE has done everything correct after the incident. However, I feel that RIPE should have warned us before conducting this experiment.

Thanks for the experiment and the report

Posted by Daniel Karrenberg at 01 September 2010 16:30

Indeed "originating unusual BGP routes" needs to be announced in advance so that operators are aware of a proosible cause for any problems that might arise. This is standard procedure but unfortunately it was not followed in this particular instance; mistakes do happen. We learn from them.

29 august ?

Posted by Marc Roger at 31 August 2010 19:06

Isn't it the 27th ?

29 august ?

Posted by Franz Schwarzinger at 31 August 2010 20:21

Thanks for letting us know! We have corrected the date now.

Cisco market share

Posted by Pavel Stan at 31 August 2010 20:22

This incident gives us an insight of XR products market share in the IP Carriers, thats CRS, ASR, and GSR-PRP.

RIPE is blameless

Posted by Daniel Golding at 31 August 2010 20:44

This was a perfectly reasonable experiment with a properly formed BGP announcement. The problems were not foreseeable - withdrawal of routes was not a desired or anticipated impact.

RIPE is not at fault, nor is Duke. Research on Internet routing is vital to security and scalability and must continue.

Experiments & Upgradations are for Betterment

Posted by Muhammad Younas at 01 September 2010 08:38

Unless you keep on experiment and upgradation of product and services for interoperatibility, betterment can not be expected.

Blame?!

Posted by Matt at 01 September 2010 14:40

By definition it was an experiment, there was a concept in the process of being proved and sadly the outcome was marginally regrettable.
Whilst there are lessons to be learnt I don't feel that it's appropriate to blame a party for what we saw.
Is this not why we test services before roll out?!

No need for blames

Posted by Kostas Zorbadelos at 01 September 2010 16:21

Perhaps better communication next time, but definitely continue experimenting and openly announce the results like you did!

Add Comment

You can add a comment by filling out the form below. Plain text formatting.

Name (Required)
Please enter your name.

Subject (Required)

Comment (Required)

Enter the word below (Required)
loud