the Availability Digest

www.availabilitydigest.com
@availabilitydig

The Internet Hits a Capacity Limit

August 2014

On the morning of Tuesday, August 12, 2014 (Eastern Standard Time), the Internet slowed to a crawl. For some web sites, the speed was simply pathetic. Others became inaccessible. Among the hardest hit was eBay. British users were unable to log onto eBay for most of the day, leading many traders to demand compensation for lost sales.

The reason was not a datacenter failure or a network outage. Rather, a long-known limit of the Internet had been exceeded. The number of routes needed to link the major Internet domains surpassed a default limit in many of the routers that provide this function. These routers crashed or could not provide full routing functions, bringing major portions of the Internet to a halt.

The problem was in the Border Gateway Protocol that interconnects the major Internet domains.

The Border Gateway Protocol

The Border Gateway Protocol (BGP) is how traffic finds its way across the globe. It maintains a current view of network connectivity. Changes to the network in one part of the world propagate in seconds throughout the world.

Internet Addressing

The delivery of messages over the Internet uses a three-level hierarchy - networks that comprise subnetworks (subnets) that serve hosts. Each host has a unique IP (Internet Protocol) address that in IP Version 4 (IPv4) is a 32-bit address comprising four 8-bit octets (an *octet* is Internet-speak for bytes). IPv4 addresses are usually expressed in *dotted decimal notation*, such as 161.35.1.19.

In an IPv4 address, the two higher-order octets specify a network. The two lower-order octets specify a subnet in the higher-order bits and a host in the lower-order bits. Often, the third octet is a subnet address. The forth octet is the host address, but the number of bits used by each is determined by the network administrator.

The network and subnetwork addresses are called the IP address *prefix*.

The 32-bit IPv4 address space provides for four billion users. However, this address space is nearing exhaustion. The new IPv6 address specification extends the address space to a 128-bit address – 16 octets – that provides 10^{38} addresses. IPv6 is still in the early stages of adoption.

Autonomous Systems

The Internet is more highly structured than simply a set of interconnected networks. It comprises a network of interconnected *autonomous systems* (AS). An AS is a collection of networks controlled by a single (or in some cases, more than one) entity, such as an ISP or a large corporation. More specifically, it is a collection of IP routing prefixes under the control of one or more network operators.

Each AS is assigned a unique autonomous system number (ASN). Currently, the ASN is a 32-bit number. However, this address space is about to be exhausted; and the ASN is being extended to 64 bits.

BGP Routers

Autonomous systems are interconnected by routers. They use the Border Gateway Protocol and are called BGP routers. BGP routers are in every major Internet hub such as those run by ISPs, large businesses, and universities.

Routers that share a direct connection (a single hop) are *BGP neighbors*. A message propagates through the Internet by being relayed from one BGP router to its neighbor until it arrives at the destination network. There, the destination network takes over and routes the message to its appropriate subnet and host.

It is the connectivity between the external BGP routers that describes the relationships of the various ASs and therefore the topology of the Internet.

To maintain a view of the current Internet topology, BGP routers exchange with their neighbors messages that advertise new routes and that withdraw unfeasible routes. These changes quickly propagate through the Internet so that all routers have a reasonably current view of the Internet topology.

The BGP Protocol

The BGP protocol is the core routing protocol of the Internet. It provides the mechanism for BGP routers to maintain routing tables that designate network reachability among the ASs.

A router's routing table may contain several paths advertised by different neighbors to a particular IP prefix. When choosing a route, the BGP protocol selects the path that routes a prefix most directly to the destination address. For instance, assume that the message is to be delivered to IP address 161.45.210.10. The router finds one entry that will route to IP prefix 161.45 and another that will route to 161.45.210. It will choose the latter route as being the more direct route to the destination.

BGP routers advertise their routes by periodically sending Update messages to all of their neighbors. The messages can be sent whenever a router experiences a routing table change, or they may be sent periodically – such as every thirty seconds.

An AS path is built by the routers as a new path propagates through the Internet via router advertisements. When a router receives a new route, it adds the ASN of its autonomous system to the path before it advertises the path to its neighbors. For instance, if a router in AS 22 receives a new path that is [345, 2078], it will advertise a path of [22, 345, 2078] to its neighbors. A neighbor then knows that it can route a message to IP prefix 2078 by sending the message to autonomous system 22, from where the message will traverse AS 345 and AS 2078 before reaching its destination.

If a router ends up with more than one route to a specific prefix, it will choose the route that goes through the least number of ASs. If there are ties, it uses additional conditions, such as route loading, to choose a route.

BGP Address Creep

The Internet limitation that was reached on August 12nd is that the BGP routing tables have grown too large for many routers. Routes are stored in routers in memory called TCAM (Tertiary Content Addressable Memory). In early routers, the default TCAM allocation held 512K (524,288) BGP addresses. In the early days of the Internet, this was a number so large that no one thought it would ever be exceeded (like the IPv4 address exhaustion).

However, the Internet has grown at a tremendous rate. In May, 2014, Cisco issued an advisory that the number of stored addresses in an average router was approaching the default limit. The typical distribution of BGP addresses today ranges from 497,000 to 511,000 and varies from minute to minute. In six years, the number of routing table entries has doubled from 256K to 512K.

It would take just a small percentage increase to overflow TCAM space in thousands of early routers still in use. That is what happened on August 12. A mundane maintenance procedure by Verizon generated a burst of new addresses. The 512K limit was exceeded in BGP routers around the world, and these routers came to their knees. They either crashed or lost routes. In any event, Internet traffic being handled by major ISPs such as Verizon, AT&T, Sprint, Cogent, and Level 3 slowed to a crawl and frustrated millions of Internet users.

Verizon Delivers the Final Blow

Verizon wasn't doing anything wrong. It was reconfiguring its network to free up the limited IPv4 addresses as it and many other companies have been doing for a while. Basically, it was deaggreagating its routes, fragmenting its network into smaller ones to make more effective use of IPv4 addresses. For instance, if a subnet had a large number of end-point IPv4 addresses assigned to it but was using only a few, breaking the unused end-points into smaller subnets freed up many unused addresses.

Of course, deaggreagation adds BGP routes to incorporate the new subnets. In this particlular case, Verizon released 15,000 new routes into the worldwide BGP network. This was enough to overload the TCAM memories of thousands of routers, and the Internet slowed to an instant crawl.

What Can Be Done?

According to Cisco, most routers have more than enough space to support larger routing tables; but the default configuration must be abandoned and more memory allocated to TCAM so that its route capacity can be significantly expanded.

Cisco has released a list of affected products. They include Cisco 6500 switches, 7600 series routers, and ASR 9000 and 1000 Series Aggregation Services Routers

Another option has to do with IPv6 addressing. Many system managers have reduced their IPv4 address capacity by reconfiguring memory to provide IPv6 address space at the expense of IPv4 address space. However, the adoption of IPv6 addressing is proceeding very slowly. Therefore, some of the allocated IPv6 memory can be returned to IPv4 use.

However, rebooting old routers is a risky business. Many system administrators will be reluctant to do so.

Summary

The BGP addressing limitation has been known for a long time, and the fact that the Internet was nearing this limit has also been known. However, system administrators have been slow to take corrective action. The problem finally caught up with them.

Future slowdowns will only be avoided if the major ISPs, corporations, and universities expend the resources necessary to upgrade old routers to significantly expand the BGP address capability of their routers. If the BGP address space is doubling every six years, it will not be long before an address capability of 1,024K will not be enough.

Acknowledgements

Information for this article was taken from the following sources:

Internet hiccups today? You're not alone. Here's why, *ZDNet*; August 12, 2014.
Internet Touches Half Million routes: Outages Possible Next Week, *Renesys*; August 13, 2014.
Internet outages expected to abate as routers are modified, rebooted, *Computerworld*; August 13, 2014.
Here's why your Internet might have been slow on Tuesday, *Washington Post*; August 13, 2014.
Slow Internet connection? The Web might have a bigger problem, *Tech Times*; August 14, 2014.
The day the Internet broke, *Telegraph*; August 15, 2014.