An Introduction toRouting the Internet

Geoff Huston
APNIC
<table>
<thead>
<tr>
<th></th>
<th>IP Address</th>
<th>Domain Name</th>
<th>RTT1</th>
<th>RTT2</th>
<th>RTT3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>85.132.40.1</td>
<td>(85.132.40.1)</td>
<td>1.202 ms</td>
<td>1.272 ms</td>
<td>1.425 ms</td>
</tr>
<tr>
<td>2</td>
<td>94.20.50.141</td>
<td>(94.20.50.141)</td>
<td>2.394 ms</td>
<td>2.649 ms</td>
<td>2.044 ms</td>
</tr>
<tr>
<td>3</td>
<td>r02-greenxchange-r04.az-ix.net</td>
<td>(85.132.60.61)</td>
<td>2.634 ms</td>
<td>2.700 ms</td>
<td>2.292 ms</td>
</tr>
<tr>
<td>4</td>
<td>so-7-0-0-ycr1.skt.cw.net</td>
<td>(166.63.220.21)</td>
<td>71.707 ms</td>
<td>77.259 ms</td>
<td>78.499 ms</td>
</tr>
<tr>
<td>5</td>
<td>xe-0-3-1-xcr2.amd.cw.net</td>
<td>(195.2.9.217)</td>
<td>102.346 ms</td>
<td>103.459 ms</td>
<td>100.372 ms</td>
</tr>
<tr>
<td>6</td>
<td>ae3-xcr1.amd.cw.net</td>
<td>(195.2.30.105)</td>
<td>98.816 ms</td>
<td>99.660 ms</td>
<td>97.998 ms</td>
</tr>
<tr>
<td>7</td>
<td>xe-4-1-0-xcr1.lsw.cw.net</td>
<td>(195.2.25.93)</td>
<td>97.892 ms</td>
<td>97.840 ms</td>
<td>132.251 ms</td>
</tr>
<tr>
<td>8</td>
<td>xe-3-0-1-xcr1.lns.cw.net</td>
<td>(195.2.25.250)</td>
<td>101.963 ms</td>
<td>97.870 ms</td>
<td>97.707 ms</td>
</tr>
<tr>
<td>9</td>
<td>195.66.236.166</td>
<td>(195.66.236.166)</td>
<td>102.975 ms</td>
<td>101.481 ms</td>
<td>107.319 ms</td>
</tr>
<tr>
<td>10</td>
<td>i-2-0-ulco-core01.bi.telstraglobal.net</td>
<td>(202.40.148.218)</td>
<td>104.572 ms</td>
<td>98.277 ms</td>
<td>102.415 ms</td>
</tr>
<tr>
<td>11</td>
<td>i-6-0-eig-core01.bx.telstraglobal.net</td>
<td>(202.84.249.21)</td>
<td>241.369 ms</td>
<td>238.537 ms</td>
<td>240.648 ms</td>
</tr>
<tr>
<td>12</td>
<td>i-0-4-1-eqnx-core01.bi.telstraglobal.net</td>
<td>(202.84.249.34)</td>
<td>239.569 ms</td>
<td>239.580 ms</td>
<td>239.513 ms</td>
</tr>
<tr>
<td>13</td>
<td>i-0-4-0-sydo-core02.bx.telstraglobal.net</td>
<td>(202.84.144.186)</td>
<td>386.366 ms</td>
<td>386.084 ms</td>
<td>398.815 ms</td>
</tr>
<tr>
<td>14</td>
<td>tengige0-2-0-0.oxf-gw1.sydney.telstra.net</td>
<td>(203.50.13.29)</td>
<td>388.800 ms</td>
<td>399.029 ms</td>
<td>397.750 ms</td>
</tr>
<tr>
<td>15</td>
<td>bundle-ether1.ken-core4.sydney.telstra.net</td>
<td>(203.50.6.5)</td>
<td>394.591 ms</td>
<td>406.403 ms</td>
<td>396.072 ms</td>
</tr>
<tr>
<td>16</td>
<td>bundle-ether5.cha-core4.brisbane.telstra.net</td>
<td>(203.50.11.73)</td>
<td>410.061 ms</td>
<td>411.584 ms</td>
<td>413.307 ms</td>
</tr>
<tr>
<td>17</td>
<td>tengigabitethernet7-1.cha30.brisbane.telstra.net</td>
<td>(203.50.51.40)</td>
<td>432.973 ms</td>
<td>401.504 ms</td>
<td>400.854 ms</td>
</tr>
<tr>
<td>18</td>
<td>4608resolvers.brisbane.telstra.net</td>
<td>(139.130.130.194)</td>
<td>405.478 ms</td>
<td>407.010 ms</td>
<td>408.246 ms</td>
</tr>
<tr>
<td>19</td>
<td>wattle.rand.apnic.net</td>
<td>(203.133.248.2)</td>
<td>406.757 ms</td>
<td>410.261 ms</td>
<td>410.676 ms</td>
</tr>
</tbody>
</table>
The Objective of Routing

• To ensure that every packet gets delivered
  – To ensure that every packet gets delivered in the shortest possible time
• This implies that every router needs to understand the relative direction of all possible destinations
• So every router needs to know the current relative location of all possible destinations
• This is the task of the routing system
Routing Architecture

• The Internet uses a *decoupled* routing architecture
  – A routing protocol is designed to pass forwarding information between routers. It does not perform the forwarding function itself
  – Different networks may (and do) use different routing protocols
  – Routing protocols change over time
    • But need not change all at once
Routing Protocols

• Routing protocols are “self discovery” distributed protocols
  – Two flavours of algorithm:
    • Distance vector (rumours)
      – I tell you everything I know
      – You tell me everything you know
      – Repeat until nothing more to day!
    • Link State (map drawing)
      – I tell everyone about my links (link state messages)
      – I listen to all link state messages
      – I compute a optimal path map of the network
Routing Architecture

• The Internet uses a *two level* routing hierarchy:
  – Interior Routing Protocols, used by each network to determine how to reach all destinations that line within the network
  – Interior Routing protocols maintain the current topology of the network
Routing Architecture

• The Internet uses a *two level* routing hierarchy:
  – Exterior Routing Protocol, used to link each component network together into a single whole
  – Exterior protocols assume that each network is fully interconnected internally
Exterior Routing: BGP

- BGP is a large set of bilateral (1:1) routing sessions
  - A tells B all the destinations (prefixes) that A is capable of reaching
  - B tells A all the destinations that B is capable of reaching

```
10.0.0.0/24
10.1.0.0/16
10.2.0.0/18
192.2.200.0/24
```
BGP decisions

• BGP is an “offer and accept” protocol
  – A network “offers” to a neighboring network prefixes that it is prepared to act as a forwarder
  – The network need not offer all its prefixes to a routing neighbour
    • It may have internal routing policies to moderate what it is prepared to offer
  – The neighboring network is not obliged to accept the offer
    • It may have better offers from other routing peers that show a preferred path
    • It may have policies about redistribution of prefixes that determine what is offered and what is accepted
A view of a Routing Table

show ip bgp
BGP table version is 0, local router ID is 203.133.248.2
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
    r RIB-failure, S Stale, R Removed
Origin codes: i - IGP, e - EGP, ? - incomplete

<table>
<thead>
<tr>
<th>Network</th>
<th>Next Hop</th>
<th>Metric</th>
<th>LocPrf</th>
<th>Weight</th>
<th>Path</th>
</tr>
</thead>
<tbody>
<tr>
<td>* 1.0.0.0/24</td>
<td>202.12.28.1</td>
<td>0</td>
<td>4777</td>
<td>15169</td>
<td>i</td>
</tr>
<tr>
<td>*&gt; 1.0.4.0/22</td>
<td>203.119.76.3</td>
<td>0</td>
<td>4777</td>
<td>2516</td>
<td>6453 7545 7545 7545 56203 i</td>
</tr>
<tr>
<td>*&gt; 1.0.16.0/23</td>
<td>203.119.76.3</td>
<td>0</td>
<td>4777</td>
<td>15169</td>
<td>i</td>
</tr>
<tr>
<td>*&gt; 1.0.18.0/23</td>
<td>203.119.76.3</td>
<td>0</td>
<td>4777</td>
<td>15169</td>
<td>i</td>
</tr>
<tr>
<td>*&gt; 1.0.20.0/23</td>
<td>203.119.76.3</td>
<td>0</td>
<td>4777</td>
<td>15169</td>
<td>i</td>
</tr>
<tr>
<td>*&gt; 1.0.22.0/23</td>
<td>203.119.76.3</td>
<td>0</td>
<td>4777</td>
<td>15169</td>
<td>i</td>
</tr>
<tr>
<td>*&gt; 1.0.24.0/23</td>
<td>203.119.76.3</td>
<td>0</td>
<td>4777</td>
<td>15169</td>
<td>i</td>
</tr>
<tr>
<td>*&gt; 1.0.26.0/23</td>
<td>203.119.76.3</td>
<td>0</td>
<td>4777</td>
<td>15169</td>
<td>i</td>
</tr>
</tbody>
</table>
Routing Properties

• Routing is designed to be robust
  – Within a richly connected topology the interior routing protocol is designed to “self heal” in the even of breakage
  – Within the exterior network, connectivity will also “self heal” if there are alternative paths that route around the break

• Routing is efficient
  – Interior routing can be tuned to converge in milliseconds
  – Exterior routing will converge across the span of the Internet within 70 seconds on average following a change in local connectivity
Routing Policies

• Exterior routing systems are often used to express a number of objectives:
  – Business requirements (e.g.: customer, peer, upstream)
  – Traffic Engineering requirements (e.g.: traffic flow balancing)
  – Robustness requirements (e.g.: alternate connection arrangements and automated failover)
Routing Insights

- Routing is essentially a view of “all of the network” at once, and as such it provides a number of useful insights into the connected network
The Big Picture of the v4 Routing Table

- Address Exhaustion
- The GFC hits the Internet
- Broadband to the Masses
- The Great Internet Boom and Bust of 2000/2001
- Introduction of CIDR – March 1994
Routing Issues
The Big Picture of the v4 Routing Table

- Address Exhaustion
- The GFC hits the Internet
- Broadband to the Masses
- The Great Internet Boom and Bust of 2000/2001
- Introduction of CIDR – March 1994

Active BGP entries (FIB)

Date
Hot topics in Routing

Security

Why Google Went Offline Today and a Bit about How the Internet Works

November 6, 2012

Today, Google’s services experienced a limited outage for about 27 minutes over some portions of the Internet. The reason this happened dives into the deep, dark corners of networking. I’m a network engineer at CloudFlare and I played a small part in helping ensure Google came back online. Here’s a bit about what happened.

At around 6:24pm PST / 02:24 UTC (5 Nov. 2012 PST / 6 Nov. 2012 UTC), CloudFlare employees noticed that Google’s services were offline. We use Google Apps for things like email so when we can’t reach their servers the office notices quickly. I’m on the Network Engineering team so I jumped online to figure out if the problem was local to us or global.

Troubleshooting

I quickly realised that we were unable to resolve all of Google’s services — or even reach 8.8.8.8, Google’s public DNS server — so I started troubleshooting DNS.
Hot Topics in Routing

• Scaling
Open Routing Standards

• Interior Routing Protocols
  – Open Shortest Path First (OSPF)
  – Intermediate System – Intermediate System (IS-IS)

• Exterior Routing Protocols
  – Border Gateway Protocol (BGP)