BGP ANYCAST SIMULATIONS
Using GTNetS

Sunitha Beeram
Talal Jaafar
Agenda

- Introduction
- GTNetS Simulations
- Topology
- Some early results
- Future work
- Q&A
IP Anycast

- Hierarchical
  - Local and global nodes
- Flat
  - Ultra DNS – Multiple anycast addresses associated with servers

- Current measurements
  - Planet Lab measurements (Sarat, Terzis et al)
    - Decrease in latency with anycast
    - Clients don’t always hit nearest server
    - Relatively small number of outages, but lasted for long time (>100s)
    - Hint more global nodes might cause instability
  - Ripe K-root measurements (Lorenzo)
    - Good latency
    - Local nodes take load off global nodes, but not by a great factor
    - Quite stable – few switches
  - Ballani and Francis with their Planet Lab measurements conclude that anycast nodes are quite stable – hardly any flips observed
Why Simulations?

- Study the impact of BGP on Anycast
  - BGP Convergence – Path Exploration!
  - Flap Dampening!
  - ....

- Study the impact of Anycasting on BGP
  - BGP Table sizes growth
  - Large # of global nodes → Convergence impact?

- Simulations Might be useful in analyzing different options for good placements of future anycast servers
Discrete-Event Packet Level simulations
BGP : BGP++ implementation ported from ns-2 (zebra based)
Simulation handles actual routing – longest prefix match based FIB which is populated by BGP
Anycast servers supported using /32 prefix address advertisement
Milnet Topology

- Realistic network model deduced from:
  - Internet topology maps
    - RocketFuel map (6 ISPs that operates in the US)
    - Scan Project map (2 Tier-1 ISPs)
  - Other data sets
    - BGP Routing Table Dump (RouteViews Project)
    - NetGeo (Internet Geographical Database)
    - Mapnet (PoP connections)
Milnet Topology (contd.)

- U.S. Internet Backbone inferred from Milnet
  - 8 national-level ISP networks
  - Total of 9000 routers (793 BGP speakers only)
    - Simulation of 9000 routers
      - Intermediate routers running OSPF
    - Simulation of BGP routers only
      - Abstracting intermediate routers
    - Decision on Intra-AS routing policies
    - Need to infer Inter-AS routing policies
F-root Topology

- Representative f-root Internet backbone connections inferred from routeviews.
  - Simulate 1 BGP speaker per AS
  - Peer-Peer, Customer and Provider relations inferred and appropriate policies applied.
  - Local and global Anycast nodes using no-export policy
  - Total of 44 large ISPs with 467 interconnecting links simulated
Some early results

<table>
<thead>
<tr>
<th>Case</th>
<th>Downtime</th>
<th># of updates</th>
<th>Change in latency</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple redundant links to destn; best path 1 hop away – 1 withdraw</td>
<td>~0</td>
<td>40,190</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>18 redundant links to Destn – 2 hops away -1 withdraw</td>
<td>75s</td>
<td>224,396</td>
<td>-0.43 s</td>
<td>Chooses different router – happens to be closer (overridden by policy earlier)</td>
</tr>
<tr>
<td>18 redundant links to Destn - 2 hops away– 1 link down</td>
<td>140s</td>
<td>34,501</td>
<td>-</td>
<td>Chooses same router – different link</td>
</tr>
<tr>
<td>19 redundant links to Destn – 3 hops away -1 withdraw</td>
<td>90s</td>
<td>53,402</td>
<td>-0.65 s</td>
<td>Again, chooses a closer router (overridden by policy earlier)</td>
</tr>
<tr>
<td>19 redundant links to Destn – 3 hops away – 1 link down</td>
<td>~180s</td>
<td>34,163</td>
<td>-</td>
<td>Chooses same router- different link</td>
</tr>
</tbody>
</table>

*2 Anycast Servers and 1 Client ; Client sends requests at constant rate of 1/s
## BGP Table snapshots

- 18 redundant links – Failure by explicit withdraw

### before withdraw

<table>
<thead>
<tr>
<th>Prefix</th>
<th>Next Hop</th>
<th>AS_Path</th>
<th>Med</th>
<th>Loc_Pfx</th>
<th>Loc_Rt</th>
<th>NexthopPref</th>
<th>Exit_Prop</th>
<th>Exit_Pf</th>
<th>Exit_Rt</th>
</tr>
</thead>
<tbody>
<tr>
<td>* 192.168.1.1/32</td>
<td>13.221.0.1</td>
<td>20</td>
<td>0</td>
<td>3549</td>
<td>8220</td>
<td>2516 8928</td>
<td>i</td>
<td></td>
<td></td>
</tr>
<tr>
<td>* 13.163.0.1</td>
<td>13.28.0.1</td>
<td>20</td>
<td>0</td>
<td>3356</td>
<td>1273</td>
<td>8928 i</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* 32.28.0.1</td>
<td>51.181.0.1</td>
<td>20</td>
<td>0</td>
<td>13237</td>
<td>6539</td>
<td>8928 i</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* 21.24.0.1</td>
<td>12.185.0.1</td>
<td>20</td>
<td>0</td>
<td>3257</td>
<td>1273</td>
<td>8928 i</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* 5.19.0.1</td>
<td>18.29.0.1</td>
<td>30</td>
<td>0</td>
<td>4637</td>
<td>5511</td>
<td>8928 i</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* 4.249.0.1</td>
<td>11.98.0.1</td>
<td>30</td>
<td>0</td>
<td>1273</td>
<td>8928</td>
<td>i</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* 25.61.0.1</td>
<td>25.53.0.1</td>
<td>20</td>
<td>0</td>
<td>6461</td>
<td>5511</td>
<td>8928 i</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* 25.30.0.1</td>
<td>13.233.0.1</td>
<td>20</td>
<td>0</td>
<td>3561</td>
<td>8928</td>
<td>i</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* 1.30.0.1</td>
<td>12.248.0.1</td>
<td>20</td>
<td>0</td>
<td>320</td>
<td>5511</td>
<td>8928 i</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* 21.135.0.1</td>
<td>2.191.0.1</td>
<td>0</td>
<td>20</td>
<td>703</td>
<td></td>
<td>i</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* 12.231.0.1</td>
<td></td>
<td>10</td>
<td></td>
<td>3003</td>
<td>8928</td>
<td>i</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### after withdrawing advertisement from AS 8928:

<table>
<thead>
<tr>
<th>Prefix</th>
<th>Next Hop</th>
<th>AS_Path</th>
<th>Med</th>
<th>Loc_Pfx</th>
<th>Loc_Rt</th>
<th>NexthopPref</th>
<th>Exit_Prop</th>
<th>Exit_Pf</th>
<th>Exit_Rt</th>
</tr>
</thead>
<tbody>
<tr>
<td>*&gt; 192.168.1.1/32</td>
<td>2.191.0.1</td>
<td>0</td>
<td>20</td>
<td>703</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### BGP Table Snapshots

<table>
<thead>
<tr>
<th>Prefix</th>
<th>Address</th>
<th>AS</th>
<th>Route</th>
<th>Metric</th>
<th>Next Hop</th>
<th>IGP Cost</th>
<th>Exit</th>
</tr>
</thead>
<tbody>
<tr>
<td>192.168.1.1/32</td>
<td>18.158.0.1</td>
<td>30</td>
<td>0 4766 2516</td>
<td>4637</td>
<td>1257 i</td>
<td>0 4766</td>
<td>2516</td>
</tr>
<tr>
<td>1257 i</td>
<td>21.24.0.1</td>
<td>30</td>
<td>0 5400 5511 13237</td>
<td>4637</td>
<td>1257 i</td>
<td>0 5400</td>
<td>5511</td>
</tr>
<tr>
<td>1257 i</td>
<td>0.209.0.1</td>
<td>20</td>
<td>0 209 286 4637</td>
<td>21.24.0.1</td>
<td>1257 i</td>
<td>0 209</td>
<td>286</td>
</tr>
<tr>
<td>1257 i</td>
<td>21.135.0.1</td>
<td>30</td>
<td>0 1299 1257 i</td>
<td>5.19.0.1</td>
<td>1257 i</td>
<td>0 1299</td>
<td>1257 i</td>
</tr>
<tr>
<td>1257 i</td>
<td>11.12.0.1</td>
<td>20</td>
<td>0 2828 3257 1257 i</td>
<td>11.12.0.1</td>
<td>1257 i</td>
<td>0 2828</td>
<td>3257 1257 i</td>
</tr>
<tr>
<td>1257 i</td>
<td>13.28.0.1</td>
<td>20</td>
<td>0 3356 6517 i</td>
<td>24.251.0.1</td>
<td>1257 i</td>
<td>0 3356</td>
<td>6517 i</td>
</tr>
<tr>
<td>1257 i</td>
<td>24.251.0.1</td>
<td>20</td>
<td>0 6395 3257 1257 i</td>
<td>12.248.0.1</td>
<td>1257 i</td>
<td>0 6395</td>
<td>3257 1257 i</td>
</tr>
<tr>
<td>1257 i</td>
<td>12.248.0.1</td>
<td>20</td>
<td>0 3320 1257 i</td>
<td>30.231.0.1</td>
<td>1257 i</td>
<td>0 3320</td>
<td>1257 i</td>
</tr>
<tr>
<td>1257 i</td>
<td>30.231.0.1</td>
<td>20</td>
<td>0 7911 6517 i</td>
<td>13.233.0.1</td>
<td>1257 i</td>
<td>0 7911</td>
<td>6517 i</td>
</tr>
<tr>
<td>1257 i</td>
<td>13.233.0.1</td>
<td>30</td>
<td>0 8220 3257 1257 i</td>
<td>32.28.0.1</td>
<td>1257 i</td>
<td>0 8220</td>
<td>3257 1257 i</td>
</tr>
<tr>
<td>1257 i</td>
<td>11.98.0.1</td>
<td>20</td>
<td>0 2914 3257 1257 i</td>
<td>25.61.0.1</td>
<td>1257 i</td>
<td>0 2914</td>
<td>3257 1257 i</td>
</tr>
<tr>
<td>1257 i</td>
<td>25.61.0.1</td>
<td>20</td>
<td>0 6461 6453 1257 i</td>
<td>11.98.0.1</td>
<td>1257 i</td>
<td>0 6461</td>
<td>6453 1257 i</td>
</tr>
<tr>
<td>1257 i</td>
<td>0.174.0.1</td>
<td>20</td>
<td>0 174 6453 1257 i</td>
<td>11.98.0.1</td>
<td>1257 i</td>
<td>0 174</td>
<td>6453 1257 i</td>
</tr>
<tr>
<td>1257 i</td>
<td>25.53.0.1</td>
<td>20</td>
<td>0 6453 1257 i</td>
<td>25.61.0.1</td>
<td>1257 i</td>
<td>0 6453</td>
<td>1257 i</td>
</tr>
<tr>
<td>1257 i</td>
<td>4.215.0.1</td>
<td>20</td>
<td>0 1239 1257 i</td>
<td>0.174.0.1</td>
<td>1257 i</td>
<td>0 1239</td>
<td>1257 i</td>
</tr>
</tbody>
</table>

19 redundant links - Link between 1257 and 13237 down after convergence
Inferences

- Simulation highlights problems with BGP path exploration
- Link failures $\leftrightarrow$ e-bgp link failures (assuming there are no mechanisms to explicitly detect link down)
  - Longer to converge but lesser network overhead (updates)
- Explicit Withdraws $\leftrightarrow$ i-bgp failures or end server failures resulting in a withdraw
  - Faster but at the cost of lot more updates
- Policies can effectively model the relations – before failures, in the simulated case, longer path was chosen because of local preference metrics $\Rightarrow$ Reinforces the fact that “nearest” is not in necessarily in terms of latency
Simulation Caveats

- Router Ids introduce randomness! Used as tie breakers when all else equal => might impact chosen routes and hence convergence
- Topology data might be insufficient
  Need a more global view to infer topology and AS relations correctly
Future Work

Use simulations to

- Compare Unicast Vs Anycast
  - Load Balancing properties
  - Latency
  - Impact on Convergence
- Study Impact of Flap Dampening
- Study impact of large number of global nodes
- Viability of Anycast for session oriented protocols
DDoS attacks
- Earlier measurements show local nodes don’t really take load off global nodes. Would this hold in case of DDoS attacks?

Include i-BGP routers and i-BGP topology
- Will it have any impact?
- With I-BGP we can model Hot Potato Routing which can also largely impact routing decision!
- Use PoP topology for i-BGP?

Impact of growth of Internet (=> increase in path length?) on anycast stability
Is Anycasting the right way to go?

- IP level anycasting should be good for low level services like DNS
- IP anycasting: Load balancing – really? Definitely does not take server loads into account
- Others should use application level anycasting – Eg. PIAS
Questions
References

Joe Abley. Hierarchical Anycast for Global Service Distribution


Sandeep Sarat et al – On the use of anycast in DNS.
http://www.cs.jhu.edu/~sarat/Anycast-TR.pdf

Hitesh Ballani, Paul Francis – Towards a global IP anycast service. Sigcomm 2005