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

GTNetS Simulation

- 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

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

*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

*	192.168.1.1/32	13.221.0.1		20	0 3549 8220 2516 8928 i
*		13.163.0.1		20	0 3491 6539 8928 i
*		13.28.0.1		20	0 3356 1273 8928 i
*		32.28.0.1		30	0 8220 2516 8928 i
*		51.181.0.1		20	0 13237 6539 8928 i
*		21.24.0.1		20	0 5400 5511 8928 i
*		12.185.0.1		20	0 3257 1273 8928 i
*		5.19.0.1		20	0 1299 8928 i
*		18.29.0.1		30	0 4637 5511 8928 i
*>		4.249.0.1		30	0 1273 8928 i
*		11.98.0.1		20	0 2914 5511 8928 i
*		25.61.0.1		20	0 6461 5511 8928 i
*		25.53.0.1		20	0 6453 5511 8928 i
*		13.233.0.1		20	0 3561 8928 i
*		1.30.0.1		20	0 286 1273 8928 i
*		12.248.0.1		20	0 3320 5511 8928 i
*		21.135.0.1		20	0 5511 8928 i
*		2.191.0.1	0	20	0 703 i
*		12.231.0.1		10	0 3303 8928 i

- after withdrawing adveterisement from AS 8928:
  *> 192.168.1.1/32 2.191.0.1 0 20 0 703 i

BGP Table Snapshots

- * 192.168.1.1/32 18.158.0.1 30 0 4766 2516 4637 1257 i
- * 21.24.0.1 30 0 5400 5511 13237 1257 i
- * 0.209.0.1 20 0 209 286 4637 1257 i
- * 13.221.0.1 20 0 3549 8220 3257 1257 i
- * 2.189.0.1 20 0 701 4637 1257 i
- * 16.227.0.1 20 0 4323 2828 3257 1257 i
- * > 21.135.0.1 30 0 5511 13237 1257 i
- * 5.19.0.1 20 0 1299 1257 i
- * 11.12.0.1 20 0 2828 3257 1257 i
- * 13.28.0.1 20 0 3356 6517 i
- * 24.251.0.1 20 0 6395 3257 1257 i
- * 12.248.0.1 20 0 3320 1257 i
- * 30.231.0.1 20 0 7911 6517 i
- * 13.233.0.1 20 0 3561 6453 1257 i
- * 32.28.0.1 30 0 8220 3257 1257 i
- * 11.98.0.1 20 0 2914 3257 1257 i
- * 12.185.0.1 20 0 3257 1257 i
- * 25.61.0.1 20 0 6461 6453 1257 i
- * 0.174.0.1 20 0 174 6453 1257 i
- * 25.53.0.1 20 0 6453 1257 i
- * 4.215.0.1 20 0 1239 1257 i

- * 192.168.1.1/32 18.158.0.1 30 0 4766 2516 4637 1257 i
- * 21.24.0.1 30 0 5400 5511 13237 3320 1257 i
- * 0.209.0.1 20 0 209 286 4637 1257 i
- * 13.221.0.1 20 0 3549 8220 3257 1257 i
- * 2.189.0.1 20 0 701 4637 1257 i
- * 16.227.0.1 20 0 4323 2828 3257 1257 i
- * 21.135.0.1 30 0 5511 13237 3320 1257 i
- * 5.19.0.1 20 0 1299 1257 i
- * 11.12.0.1 20 0 2828 3257 1257 i
- * 13.28.0.1 20 0 3356 6517 i
- * 24.251.0.1 20 0 6395 3257 1257 i
- * 12.248.0.1 20 0 3320 1257 i
- * 30.231.0.1 20 0 7911 6517 i
- * 13.233.0.1 20 0 3561 6453 1257 i
- *> 32.28.0.1 30 0 8220 3257 1257 i
- * 11.98.0.1 20 0 2914 3257 1257 i
- * 12.185.0.1 20 0 3257 1257 i
- * 25.61.0.1 20 0 6461 6453 1257 i
- * 0.174.0.1 20 0 174 6453 1257 i
- * 25.53.0.1 20 0 6453 1257 i
- * 4.215.0.1 20 0 1239 1257 i

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

Inferences

- Simulation highlights problems with BGP path exploration
- Link failures ←→e-bgp link failures (assuming there are no mechanisms to explictly detect link down)
  - Longer to converge but lesser network overhead (updates)
- Explicit Withdraws ←→ 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 => 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

Future Work

- 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 http://www.isc.org/pubs/tn/index.pl?tn=isc-tn-2003-1.html

Joe Abley. A Software Approach to Distributing Requests for DNS Service using GNU Zebra, ISC BIND 9 and FreeBSD http://www.isc.org/pubs/tn/index.pl?tn=isc-tn-2004-1.html

Lorenzo Colitti – Effect of anycast on K-root. Some early results. http://www.ripe.net/ripe/meetings/ripe-51/presentations/pdf/ripe51-anycast_k-root.pdf

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

Dina Katabi and John Wroclawski , A Framework for Scalable Global IP-Anycast (GIA). Sigcomm 2000