Observations on Anycast Topology and Performance

Steve Gibbard
Packet Clearing House

Introduction

- Lots of DNS systems using Anycast
- Lots of research being done on performance.
- Lots of topological variations between anycast systems.
- Lots of different research results.
- Why?
- What does this teach us about anycast design?

Past research

- Barber et. al: Life and Times of J Root (2004).
  - Showed slight geographic correlation for J Root.
- Colitti et. al: Effects of anycast on root name servers.
  - Cast doubt on effectiveness of K Root Delhi node.
- Liu, et. al: Two Days in the Life of DNS Root Servers.
  - Looked at three different systems, different results.
- Methodologies varied. Liu provides easiest point of comparison across multiple systems.

Behavioral differences (Liu…)

- Local nodes consistent; global nodes varied.
- C Root:
  - Four global nodes. All in the US. All on Cogent's backbone.
  - Queries generally went to closest node.
- K Root:
  - Five global nodes, spread around the world. Various transit arrangements.
  - Lots of queries to non-optimal locations.
- F Root's global nodes too close to matter.

C Root distribution details

- 92% of clients used nearby server
  - Chicago: Almost all traffic from Americas.
  - Los Angeles: Americas, Asia, Oceania.
  - New York/DC: Americas, Europe, Africa, (West?) Asia.

K root distribution details

- 29% of clients used optimal instance.
  - Figure includes local nodes, so probably worse for global nodes.
  - Miami: Almost all traffic from Americas.
  - Tokyo: Almost all traffic from Asia and Oceania.
  - Amsterdam: Most traffic from Europe.
  - London: 45% from Americas and 25% from Asia.
  - Delhi: 60% from Americas.

Why does optimal routing matter?

- Data moves at the speed of light.
  - Or slower if there's congestion.
- Internet routing (at least in the core) generally follows geography.
- Queries to far-away servers are slower.
- The longer a path is, the more things there are to break.

Packet Clearing House

Why the differences?

- Internet routing decision process:
  - Best to get paid, second best to not have to pay, worst to have to pay.
  - The shorter the distance the better.
- As implemented with BGP:
  - Local preference: Customer over peer. Peer over transit.
  - Best exit routing.

Routing and anycast

- For Unicast hosts, traffic flow is pretty optimal.
- Backbones get designed around this.
  - Requirements to peer in all areas of overlap.
  - Consistent transit.
- Anycast looks like a backbone.
  - Lots of entry points.
  - If transit and peering are inconsistent, closest nodes aren't the preferred path.

PCH
Packet Clearing House

Details

- Traffic that hit Level3 in US ended up in India.
  - K root was a customer of STPI only in Delhi.
  - STPI was a customer of Level3.
  - Routing for that node is different now.
- Traffic hitting AboveNet in US ended up in London.
  - K root is a customer of AboveNet only in London.
  - London gets more US traffic than Miami.
- Amsterdam and Tokyo try hard not to draw transit from outside via prepending.
  - Amsterdam gets less European traffic than London.
  - Tokyo gets 1/3 of London's Asia volume.

Other anycast systems

- I, J, and M roots all have routing that looks a lot like K Root's. Similar performance is expected.
- Indeed, J Root traces from Bay Area end up far away: Seoul, Toronto, Amsterdam.
- UltraDNS looks more optimal in some, but not all, clouds.

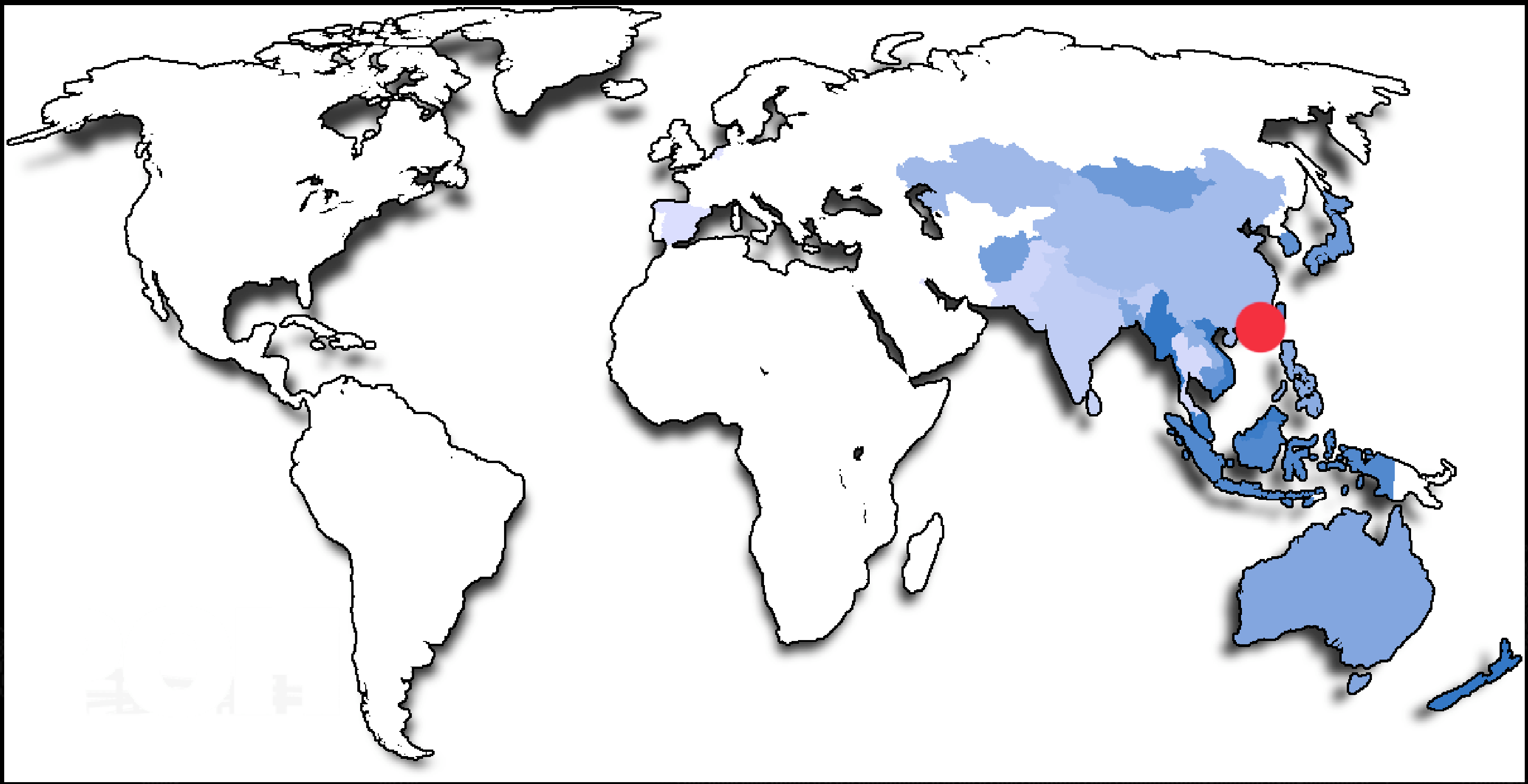
Testing my assumptions

- PCH anycast -- hosts 17 TLDs.
  - Four global nodes: Hong Kong, Palo Alto, Ashburn, London.
  - Consistent transit: NTT and Teleglobe.
  - Local nodes not included in analysis.
  - Refuse other, more regional, offers of transit.
  - Transit should hot-potato into closest nodes.
  - Hypothesis: Global node query distribution should be geographically optimal.

Methodology

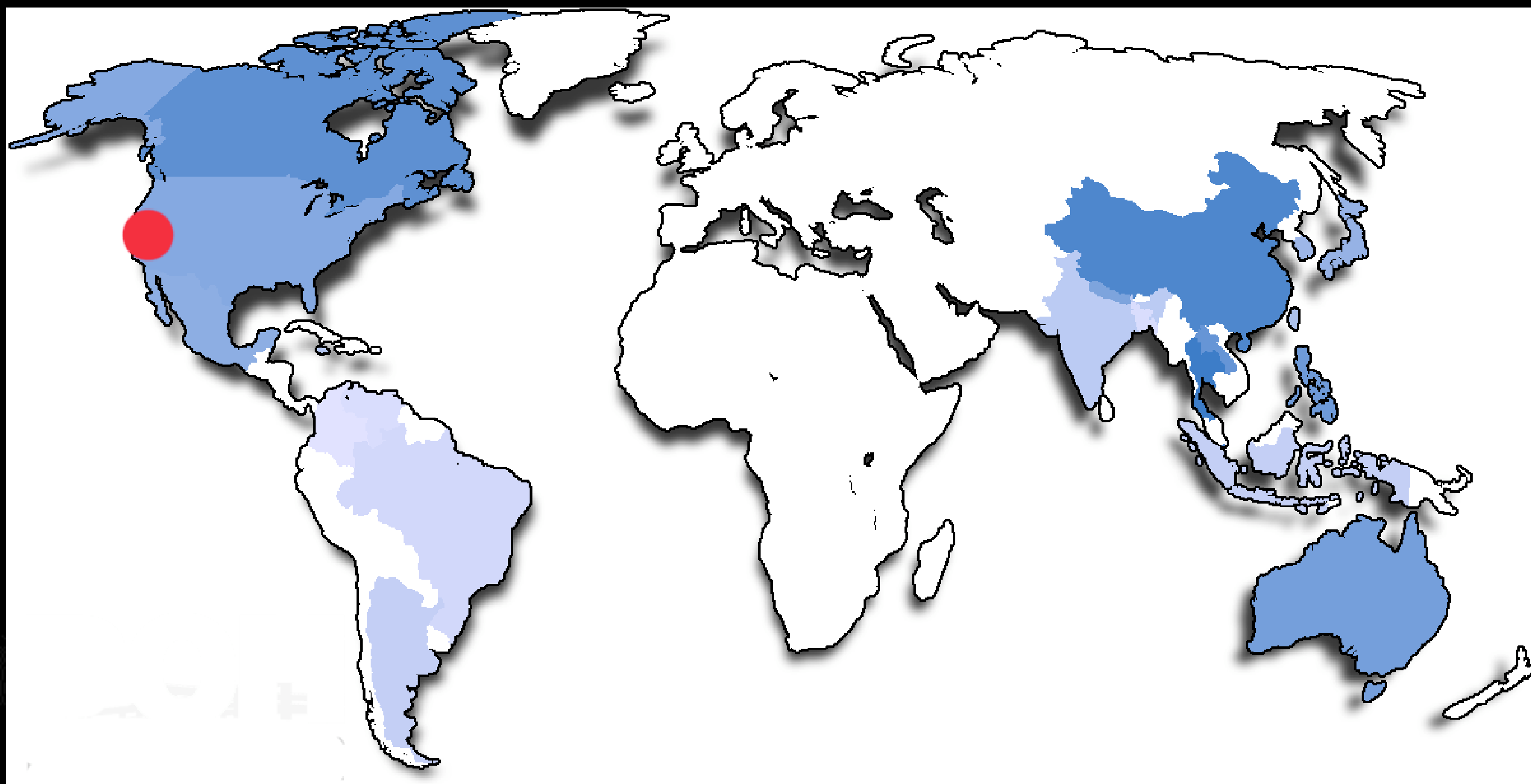
- Look at unique sources, not hit counts.
- Aggregate sources by /24
- Examined 24 hours of data from January, 2007.
- Some peering traffic included in data.

Hong Kong

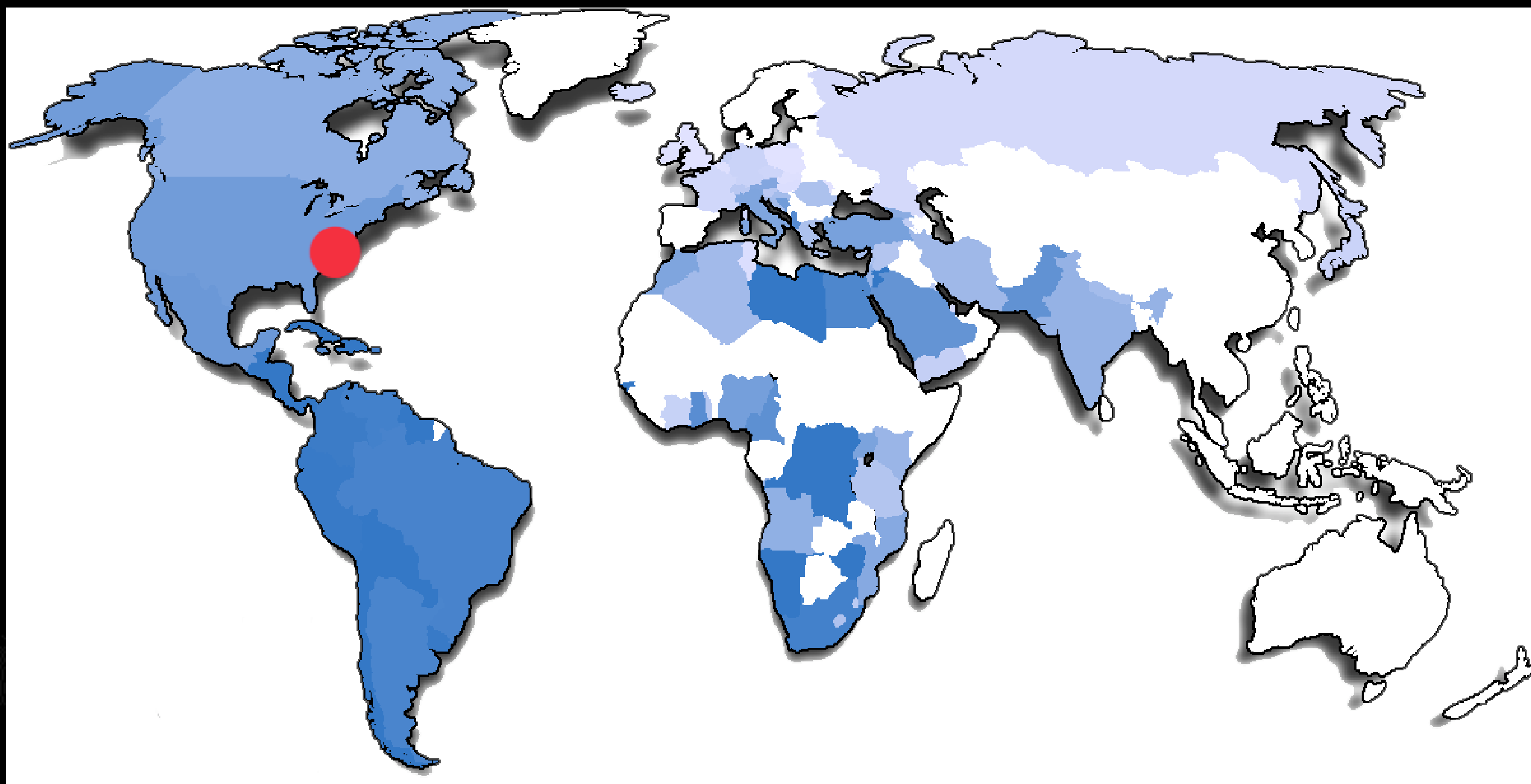


Packet Clearing House

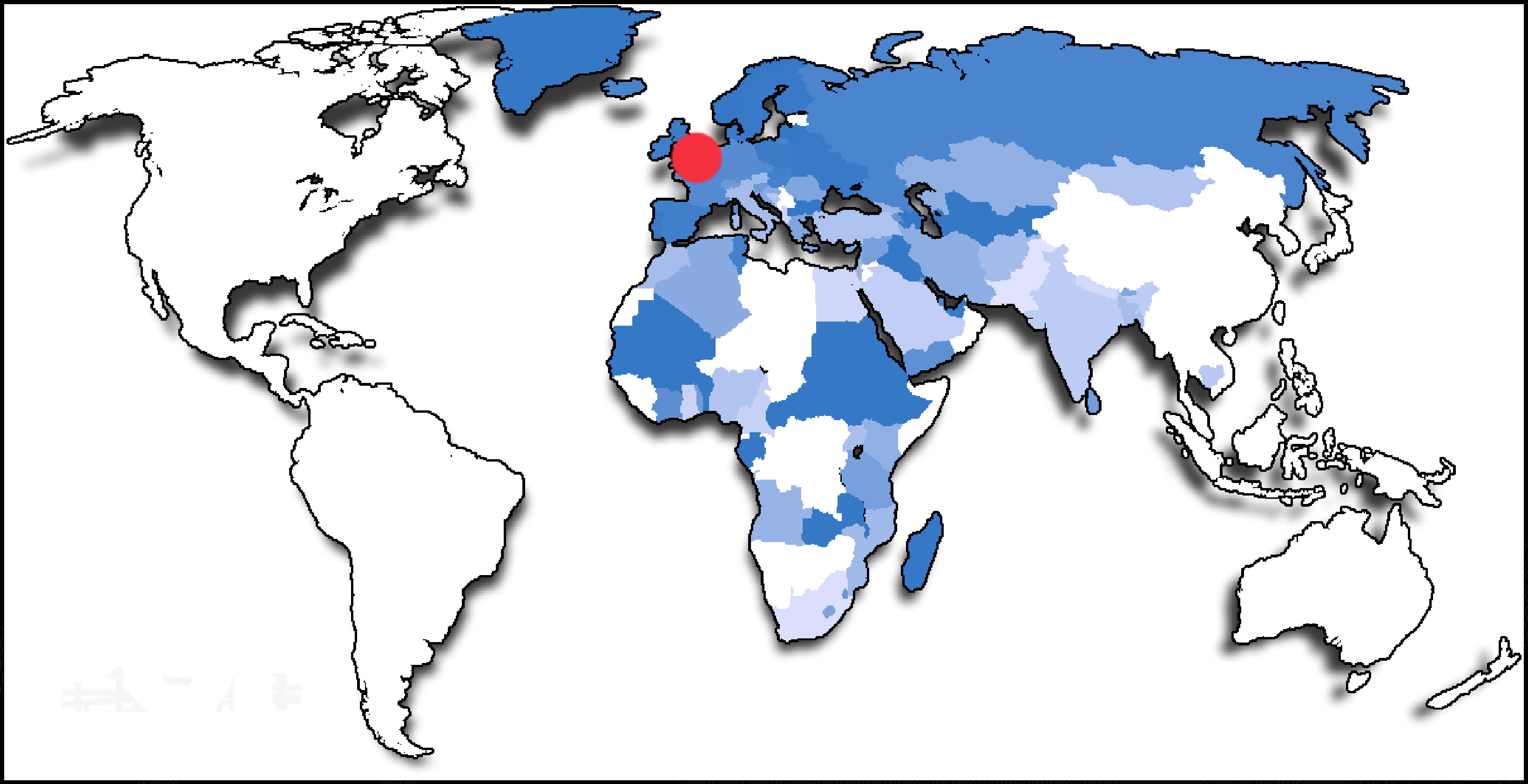
Palo Alto



Ashburn



London



Ashburn leaky, others about right

- Ashburn:
  - Africa
    - Satellite connectivity often to Eastern US.
  - Peering issues
    - Telecom Italia -- peers with us only in US.
    - KDDI not doing best exit (fixed).
    - Should there be policy changes?
- Hong Kong weak. Wrong location, or East Asia just too US-Centric?

Conclusions

- Performance improved by being consistent with transit arrangements.
- Backbone engineering principles seem to apply to anycast.
- Redundancy and diversity are good, but do it carefully.
  - Multiple distributed sets of global nodes, each with its own consistent transit?

Packet Clearing House

Thanks!

Paper at <insert URL here>

Steve Gibbard

Packet Clearing House

scg@pch.net

http://www.pch.net