Being Open:
How Facebook Got It’s Edge

James Quinn
Network Engineer, facebook
1.71 Billion Users

1.1+ Billion Daily Users

2 Billion Photos daily

60 Billion Messages daily

15% of global egress is already IPv6
Our edge connects to the world

1.71 Billion People
What is the service we support?

- **dynamic**
  - newsfeed
  - likes
  - status updates

- **static**
  - images
  - css
  - js
Asia -> Oregon

TCP Connect: 150ms
HTTPS Asia -> Oregon

TCP connection established: 150 ms
SSL session established: 450 ms
Response received: 600 ms

TCP SYN: 150 ms
ACK: 300 ms
ClientHello: 450 ms
ServerHello: 540 ms
ChangeCipherSpec: 630 ms
ChangeCipherSpec: 720 ms
GET: 810 ms
HTTP 1.1: 900 ms

75 ms

DC
Asia -> Oregon

TCP Connect: 30ms
SSL Session: ??
HTTP Response: ??
HTTPS Asia -> POP -> Oregon

Sessions established: 90 ms (vs 450 ms)
Response Received: 240 ms
Asia -> Oregon

TCP Connect: 150ms
SSL Session: 450ms
HTTP Response: 600ms

PoP

DC

30ms 90ms 240ms
edge routers -> edge clusters
edge routers -> edge clusters
edge routers -> edge clusters
Video Is Growing
edge metro topology

100G Everywhere!
Evolving beyond BGP
Global BGP routing

AS 32934
Which POP is best?

Considerations:
- Closest Edge to user
Sonar: Measuring “closeness”
Global controller architecture

Internet

ISP DNS resolver

Facebook Infrastructure

capacity
resolvers/latency
routing
health

global controller

DNS map

DNS Server

offline processing
Global controller architecture

- **ISP A DNS resolver**
- **ISP B DNS resolver**
- **ISP C DNS resolver**
- **ISP D DNS resolver**

Internet

- A123.4
- A123.5
- A123.6

Global controller architecture

- Global Controller
- DNS map
- DNS Server
Global controller in action

User Demand

Controlled Utilization

morning  mid-day  evening  sleep
Regional load shedding
Global load shedding
But what about BGP & targeting?
Metro-level BGP peering islands

AS 32934

global controller

DNS Targeting

Ingress

Egress
Even on an island, BGP is limited

<table>
<thead>
<tr>
<th>BGP cares about...</th>
<th>We care about...</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. longest prefix</td>
<td>1. capacity</td>
</tr>
<tr>
<td>2. local preference</td>
<td>2. packet loss</td>
</tr>
<tr>
<td>3. AS path</td>
<td>3. latency</td>
</tr>
<tr>
<td>4. MED</td>
<td>4. service performance</td>
</tr>
</tbody>
</table>
Local network controllers
Host routing!

- Looks great in slides
- Not so simple in practice...
- Synchronization across hosts
- A *lot* of hosts
- Non-local traffic sources
- And which encapsulation?
How do the servers encapsulate!?

- MPLS – Kernel support problems, router features
- PBR - limits platform diversity, router configuration state
  - DSCP – very coarse range
  - GRE key – repeated router bugs

What we learned.... keep it simple
Evolving beyond BGP ... with BGP

- Peer
- Transit
- overloaded link
- happy links!
- Peering Router
- BMP, NetFlow, Counters
- BGP
- Local Controller

Where should this traffic go?
Metro Traffic Engineering

- Peer
- Peering Router
- Switch
- Local Controller
Awesome! So it’s all solved?
Huge Prefixes

- Peer
- 3ffe::/24 IPv6 prefix
- 20 Gbps egress
- BGP Peering Router
- Injected /26 sub-prefixes
- Local Controller
Oscillation between controllers
Local and Global Controller interaction

link utilization

overloaded

Local Controller

Global Controller

good

Local Controller

Global Controller

underloaded
Clear roles so they don’t oscillate
Oscillation from imprecise data

Lack of Precise Bandwidth Estimation

20 Gbps

10 Gbps

w/o precision
Oscillation from imprecise data

Lack of Precise Bandwidth Estimation

20 Gbps

10 Gbps

w/o precision

w/ precision
All together...

Global Controller

POP Cluster
POP Cluster
POP Cluster
All together...
Increasing:

- Modularity
- Programmability
- Performance
Questions?
Photo Credits

http://www.flickr.com/photos/27587002@N07/5170590074
http://www.flickr.com/photos/yaketyyakyak/7001664846
http://www.flickr.com/photos/hinnosaar/3778903507
http://www.flickr.com/photos/eamoncurry/8698726494
http://www.flickr.com/photos/43158397@N02/4514113429
http://www.flickr.com/photos/nobusue/6876280595
http://www.flickr.com/photos/29487672@N07/14760573314
http://www.flickr.com/photos/joyosity/3595242078
http://www.flickr.com/photos/kyntharyn74/3262089319
http://www.flickr.com/photos/rexipe/826987087
http://www.flickr.com/photos/lablasco/6815671096