Skynet: A Cloud-Based Data Transfer Architecture

Aleksandar Kuzmanovic

http://networks.cs.northwestern.edu
The Rise of “Hyper Giants”

Consolidation of Content

- Top 150 content providers account for 50% of traffic
- Top 30 content providers account for 35% of traffic
Hyper Giants Morphing into CDNs
Hyper Giants’ CDN Examples

- More than 60% of traffic flows directly between Google and consumer networks
  - Good for Google and consumer networks
- Facebook applies the same approach
  - More than 25% of Facebook runs through direct peerings with last-mile providers
- Microsoft started building its own CDN

Google Global Cache Server
Hyper Giants CDNs’ Properties

- Hyper giants’ CDNs are different from traditional CDNs (e.g., Akamai’s):
  - Support bi-directional data transfers, i.e., both uploads and downloads
    - Needed for various Web 2.0 apps
  - Agile replication of content towards anticipated receivers

Research question:
- Can we build a generic data transfer architecture on top of clouds?
Legacy Data Transfer
Cloud-Based Data Transfer
Why Should This Work?

- Cut the e2e Internet path into several shorter-RTT path chunks
- Avoid Internet bottlenecks
- Benefit from agile cloud replication
Deployment Scenarios

- **The bad**
  - “Free-riding” not likely to be liked by clouds:
    - Private resources used by 3rd parties
      - The problem is a transparent data transfer service provided by clouds
    - Hard to detect because:
      - Large amounts of data
        - we were able to push 100 Gbytes with no problems
      - Opening multiple accounts is straightforward

- **The good**
  - Potential for selling data acceleration services to 3rd parties
    - such that the performance of host applications is not degraded
Cloud-Based Data Transfer Services

Multicast

Multi-cloud hopping

Multi-cloud homing
Remaining Outline

- Cloud selection and properties
- Per-cloud data transfer performance
- Decomposing cloud-based data transfers
- Multi-cloud homing
- Cloud-based multicast
- Cloud path consistency
- Cloud-hopping overlays
- Cloud-supported overlays
Investigated “Carrier Applications”

<table>
<thead>
<tr>
<th>Region</th>
<th>Gmail</th>
<th>Hotmail</th>
<th>YouTube</th>
<th>Flickr</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D</td>
<td>U</td>
<td>D</td>
<td>U</td>
</tr>
<tr>
<td>N. America</td>
<td>17</td>
<td>17</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Europe</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Asia</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>S. America</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Oceania</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Unknown</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td>23</td>
<td>6</td>
<td>2</td>
</tr>
</tbody>
</table>

| Region     | |         | |         | | |
|------------|---------|---------|---------|---------|---------|
|            | D      | U       | D       | U       | D       | U       | D | U |
| N. America | 171917 | 1917    | 1512    | 121    |
| Europe     | 62591  | 2591    | 80      | 01     |
| Asia       | 01186  | 1186    | 40      | 00     |
| S. America | 0154   | 154     | 10      | 00     |
| Oceania    | 0285   | 285     | 00      | 00     |
| Unknown    | 099    | 99      | 012     | 12     |
| Total      | 6232   | 6232    | 281     | 12     |
Per-Cloud Performance

Cloud-based paths can often, consistently and significantly outperform regular Internet paths.

Transcoding effects

- YouTube: 42.3%
- Flickr: 63.5%

CDF

Transfer time [second]
Cloud-based paths can *often, consistently* and *significantly* outperform regular Internet paths.

Fully transparent data transfer

![Graph showing CDF of transfer time for different services](image)

- Gmail: 41%
- Hotmail: 19.4%
- YouTube: xx%
- Flickr: xx%
- Direct Transfer: xx%
Transfer-Time Variance

Variance of transfer times is much smaller for cloud-based transfers then for Internet paths.

![Graph showing variance of transfer times between cloud and Internet for different services like Gmail, Hotmail, Youtube, Flickr, and Direct Transfer.](image-url)
The Role of RTTs

Longer-RTT paths are more likely to be improved via clouds, but other factors play the role as well.

"Cloud" curves are shifted to the right.
The Role of RTTs

Longer-RTT paths are more likely to be improved via clouds, but other factors play the role as well.

- Short-RTT paths can be improved via clouds.
- Long-RTT paths may not be improved by clouds.
Remaining Outline

- Cloud selection and properties
- Per-cloud data transfer performance
- Decomposing cloud-based data transfers
- Multi-cloud homing
- Cloud-based multicast
- Cloud path consistency
- Cloud-hopping overlays
- Cloud-supported overlays
Upload latency depends upon the scale and distribution of upload servers.
Skynet: A Cloud-Based Data Transfer Architecture

Replication Latency

Different replication policies

CDF

Replication Latency [second]

No replication

Replication

Gmail
Hotmail
Youtube
Flickr
Download Latency

Pushing data closer to end users pays off

![CDF Graph]

- Gmail: 14 sec
- Hotmail: 73 sec
- YouTube: rate limiting
Cloud-Based Multicast

For multicast group sizes > 25, cloud-based multicast outperforms direct multicast for all clouds.
Remaining Outline

- Cloud selection and properties
- Per-cloud data transfer performance
- Decomposing cloud-based data transfers
- Multi-cloud homing
- Cloud-based multicast
- Cloud path consistency
- Cloud-hopping overlays
- Cloud-supported overlays
Cloud-Supported Overlay

Cloud-based paths bring additional benefits in overlay scenarios

35% improvement for files that experience longest transfer times.
Remaining Outline

- Cloud selection and properties
- Per-cloud data transfer performance
- Decomposing cloud-based data transfers
- Multi-cloud homing
- Cloud-based multicast
- Cloud path consistency
- Cloud-hopping overlays
- Cloud-supported overlays
Conclusions

- The rise of hyper giants and their transparency enable cloud-based data transfers

- We demonstrated the feasibility of building data transfer services on top of clouds and their superior performance
  - Opens novel security challenges and may lead to an “arms race” in this area
  - Shows significant potentials for building explicit ties between legacy- and the new Internet
Other Projects

- Monitoring net neutrality (NSF and Google)
  - Design auditing tools to enable ISPs’ transparency
  - Measurement Lab
- Auditing Internet content (NSF)
  - How do we know that the information on the Web is not biased?
- Analyzing human mobility with applications in networked systems (Narus Inc.)
- Sound fusion project
Sound Fusion Project

- A huge amount of live concert recordings is available online
  - But the quality can be very poor
- Can we fuse these recordings to generate a high-quality signal?
- Complications:
  - No reference that defines which part is music and which noise
  - SNR unknown
  - No pilot signal
Thank You!

Questions?

http://networks.cs.northwestern.edu