Next Stop, the Cloud: Understanding Modern Web Service Deployment in EC2 and Azure

Keqiang He, Alexis Fisher, Liang Wang, Aaron Gember, Aditya Akella, Thomas Ristenpart

University of Wisconsin-Madison
Move to IaaS clouds

Private Datacenters

Dropbox

Pinterest

NETFLIX

LinkedIn

Windows Azure

Rackspace

MSN
Incomplete view of cloud use

1/3 of daily users
One third of all internet users will access an Amazon AWS cloud site on average at least once a day.

1% of Internet traffic
One percent of all internet consumer traffic on average is coming or going to Amazon managed infrastructure.

Want a global, in-depth understanding of IaaS cloud usage patterns
Fundamental questions

1. **Who is using public clouds?**
   → Percentage of cloud-using domains, traffic profile

2. **How are Web services using the cloud?**
   → Impact of failures; ways to improve availability
   → Design of new systems/services
Datasets

• University packet capture
  – Deep, but local perspective
  
  IP Ranges
  Campus Traffic

  TCPDUMP

  1.4TB Capture

• Alexa subdomains DNS records
  – Broad, high-level perspective

  Top 1 million
  Zone transfer

  Subdomains

  DNS records
Fundamental questions

1. *Who is using public clouds?*

   → Percentage of cloud-using domains, traffic profile
How many popular web services use the cloud?

Cloud availability and performance issues may impact over 4% of popular web services.
Which (locally) popular web services use the cloud?

Traffic Volume (Bytes)

1. Dropbox 68.2%
2. Netflix 1.7%
3. TRUSTe CERTIFIED PRIVACY 1.1%
4. CHANNEL3000.com 0.7%
5. Pinterest 0.6%

1. atdmt.com 3.1%
2. msn 2.4%
3. msecnd.net 2.3%
4. Microsoft 1.5%
5. s-msn.com 1.4%
Fundamental questions

2. How are Web services using the cloud?
   → Impact of failures; ways to improve availability
   → Design of new systems/services
Which front ends are used?

<table>
<thead>
<tr>
<th>DNS Server</th>
</tr>
</thead>
<tbody>
<tr>
<td>sub.abc.com</td>
</tr>
<tr>
<td>abc.elb.aws.com</td>
</tr>
<tr>
<td>proxy.heroku.com</td>
</tr>
<tr>
<td>sub.abc.com</td>
</tr>
<tr>
<td>proxy.heroku.com</td>
</tr>
</tbody>
</table>

String matching on CNAME
Which front ends are used?

There is limited uptake of value-added features.
How many and which regions are used?
How many and which regions are used?

521,681
(74%)

116,366
(16%)

40,548

Virginia
Ireland
California
Oregon
Singapore
Tokyo
Sao Paulo
How many and which regions are used?

- Virginia: 1,035
- California: 1,205
- Illinois: 632
- Texas: 502
- Ireland: 862
- Netherlands: 558
- Singapore: 2,071
- Hong Kong: 1,395
How many and which regions are used?

Single region failures can take down a large fraction of cloud-using subdomains.
How many and which availability zones are used?

- Latency measurements
  - Low latency => instances are in same zone
  - High coverage, low accuracy (noisy)

- IP address correlation
  - IPs are in the same /16 subnet => instances are in the same zone
  - Higher accuracy, low coverage
How many and which availability zones are used?

- 3+ Zones: 22.3%
- 2 Zones: 33.2%
- Single zone: 44.5%

Single zone failures can take down 16% of subdomains and affect many others.
Highlights

• The first comprehensive study of Web service deployment in public clouds
• 4% of popular web services use EC2/Azure
• VMs are the most popular EC2 front end, but value-added features are used by top domains
• Limited region and zone usage (and diversity) makes web services vulnerable to failures