Delivering the World’s Most Popular Web Content

Villanova University Colloquium
January 28, 2008
Jonathan Anderson, Akamai Technologies
Agenda

- A brief history of the *commercial* Internet
- The Akamai story
- The Internet with and without Akamai
- Why being *highly* distributed matters
- Akamai network statistics
- Q & A
Terminology: The Internet vs. The Web

• “The Internet”
  • IP (layer 3)
  • TCP and UDP (layer 4, over IP)
  • DNS (layer 5, over UDP and TCP)

• “The Web”
  • HTTP and HTTPS (layer 5, over TCP)
  • URIs / URLs
How the Internet became commercially managed

- 1986: Publicly funded NSFNet goes online, linking five major NSF supercomputing centers and the ARPANET via TCP/IP
  - Flow of traffic so great that Merit Network, Inc was awarded a contract to design an upgrade
- 1988: 1st backbone upgrade, 170 networks connected
- 1990: Sir Tim Berners-Lee proposes the Web at CERN
- 1991: 2nd backbone upgrade to 45 Mbps, 3500 networks
- 1991: Berners-Lee posts a summary of the WWW project on the alt.hypertext newsgroup
  - Web servers start to appear on NSFNet
- 1993: NCSA releases Mosaic Web browser
- 1994: Berners-Lee founds the W3C at MIT
- 1995: 100,000 networks connected
The commercial Internet is born

- April, 1995: NSFNet backbone decommissioned
  - NSF awards four contracts to maintain four separate Network Access Points (NAPs) that replace the NSFNet backbone
    - MFS Datanet (now Verizon)
    - Sprint
    - Ameritech (now AT&T)
    - Pacific Bell (now AT&T)
  - The Internet is effectively privatized

- The monkeys are now running the zoo
  - Law of the jungle = free market economy
Agenda

- A brief history of the *commercial* Internet
- The Akamai story
- The Internet with and without Akamai
- Why being *highly* distributed matters
- Akamai network statistics
- Q & A
The Akamai story begins in 1995...

...during a water-cooler conversation at MIT

Dr. Tom Leighton
- Co-founder of Akamai
- Professor in MIT’s Math Department
- Head of Algorithms group in MIT’s lab for Comp Sci

Dr. Tim Berners-Lee
- Director of the W3C
- Recognized father of the World Wide Web

The problem: flash traffic can crash Web servers
1998: Akamai is founded in Cambridge, MA

- Danny Lewin was a student of Dr. Leighton’s
  - His master’s thesis revolved around the theory of consistent hashing
  - Provided a way of distributing requests among a changing population of web servers

- Leighton and Lewin enter the MIT 50K business plan competition... and lose

- Thinking that they were actually onto something, they incorporated Akamai Technologies in 1998
Akamai’s Big Idea

- The nature of the commercial Internet will *significantly* impede a content provider’s ability to deliver Web content from a single location in a scalable, reliable manner

- Deliver content from the edge of the Internet, as close to end-users as possible

- Map the Internet to dynamically direct end-users to the most optimal edge-server available
Agenda

- A brief history of the commercial Internet
- The Akamai story
- The Internet with and without Akamai
- Why being highly distributed matters
- Akamai network statistics
- Q & A
The Internet
The Internet: Simple on the outside...
...but problematic on the inside

The Problems with Peering

- Economic considerations limit peering capacity – results in congestion and poor performance.
- Routing algorithms (BGP) ignore congestion!
- Data used to determine routes is subject to intentional inaccuracies and human error.
- Routing algorithms are subject to accidental loss of routes, filtering, and intentional theft of routes.
The Problems with Peering

- Economic considerations limit peering capacity – results in congestion and poor performance.
- Routing algorithms (BGP) ignore congestion!
- Data used to determine routes is subject to intentional inaccuracies and human error.
- Routing algorithms are subject to accidental loss of routes, filtering, and intentional theft of routes.
Implications

- **Slow downloads and applications**
  - Content and transactions must traverse multiple backbones and long distances

- **Unreliable**
  - Content and transactions may be blocked by congestion or backbone peering problems

- **Not scalable**
  - Usage limited by bandwidth and CPU power available at master site

- **Expensive**
  - Provisioning for peak usage and redundancy takes time and is cost prohibitive

- **Insecure**
  - Traffic can be stolen or blocked
Implications

- **Slow downloads and applications**
  - Content and transactions must traverse multiple backbones and long distances

- **Unreliable**
  - Content and transactions may be blocked by congestion or backbone peering problems

- **Not scalable**
  - Usage limited by bandwidth and CPU power available at master site

- **Expensive**
  - Provisioning for peak usage and redundancy takes time and is cost prohibitive

- **Insecure**
  - Traffic can be stolen or blocked

- **Broadband doesn’t help**
  - As broadband becomes ubiquitous, the problems with centralized solutions become more obvious—not better
1) User enters www.xyz.com
   Browser requests IP address for www.xyz.com which is CNAMEd to Akamai
   DNS returns IP address of optimal Akamai server

2) Browser requests HTML
   Akamai server assembles page, contacting the origin server only if necessary
   Akamai server returns HTML

3) Browser requests additional objects from Akamai servers, and is served directly from cache whenever possible
• Internet routing uses Border Gateway Protocol
• BGP ignores congestion and latency, slow to route around failures
• SureRoute finds alternate routes by using intermediate Akamai servers
Benefits provided to Content Providers

- **Performance**
  - Content is served from locations near to end users
  - Significant origin-offload possible through caching or tiered distribution

- **Reliability**
  - No single point of failure
  - Automatic failover
  - Multiple routes to a customer’s origin infrastructure

- **Scalability**
  - Global capacity on demand
  - Distributed by design

- **Cost efficiency**
  - No over-provisioning
  - No redundant datacenters
  - Simple to manage

- **Security**
  - Traffic and routes much harder to steal
  - Defends against Distributed Denial of Service attacks
  - Defense in depth protects central infrastructure
  - Highly configurable SSL functionality
Agenda

• A brief history of the commercial Internet
• The Akamai story
• The Internet with and without Akamai
• Why being highly distributed matters
• Akamai network statistics
• Q & A
The Akamai Platform

The world’s largest on-demand, distributed computing platform delivers all forms of Web content and applications for over 2,400 customers.

**The Akamai EdgePlatform:**

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>28,000+</td>
<td>1,200+</td>
<td>1000+</td>
<td>650+</td>
<td>72</td>
</tr>
<tr>
<td>Servers</td>
<td>PoPs</td>
<td>Networks</td>
<td>Cities</td>
<td>Countries</td>
</tr>
</tbody>
</table>

Typical daily traffic:
170+ billion hits
100+ million streams
3,000+ terabytes delivered
Why *highly* distributed computing?

**The Answer in a Nutshell:**

- Because Internet users are *highly* distributed.
- Greater distribution means greater performance and reliability.
- Greater distribution means greater scalability.
- Greater distribution *lowers* Akamai’s cost.
The Internet is highly distributed

- No one network has more than about 8% of the access traffic.
- The top 30 networks combined add up to only 50%.

8.1% Network A
5.1% Network B
5.0% Network C
3.1% Network D
2.5% Network E
2.4% Network F
2.0% Network G
1.9% Network H

% of Access Traffic

ISPs (13,000+)
Some economic truths of the Internet

- Capacity is \textit{not} free nor is it infinite.

- Dollars flow in at the edge, so there is little incentive to build capacity between network peers.

- Routing is determined by business rules (who pays who for what), not by performance or reliability.
How much capacity is needed?

Consider an audience of 50 million (Nielsen viewership for Grey’s Anatomy and CSI Miami)

<table>
<thead>
<tr>
<th>Quality</th>
<th>Encoding Rate</th>
<th>Aggregate Demand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recreational</td>
<td>700 Kbps</td>
<td>35 Tbps</td>
</tr>
<tr>
<td>Television</td>
<td>2 Mbps</td>
<td>100 Tbps</td>
</tr>
</tbody>
</table>
Akamai Edge Cost

- Additional locations at low incremental cost
  - Scale enabled by our software automation
  - Main cost is commodity servers

- Reduced or free bandwidth and colocation
  - Better performance & reliability for ISP customers
  - Significant savings on upstream bandwidth for ISP
Agenda

- A brief history of the *commercial* Internet
- The Akamai story
- The Internet with and without Akamai
- Why being *highly* distributed matters
- Akamai network statistics
- Q & A
Live Akamai NOCC screens

• ...browser demo
Agenda

- A brief history of the *commercial* Internet
- The Akamai story
- The Internet with and without Akamai
- Why being *highly* distributed matters
- Akamai network statistics
- Q & A
Questions?

- Thanks for coming!