Engineering a Content Delivery Network

Bruce Maggs

Network Deployment

20000+ Servers	1200+ Networks	72+ Countries

Current Installations

Akamai

Part I: Services

http://www.yahoo.com

http://www.amazon.com

http://windowsupdate.microsoft.com

http://www.apple.com/quicktime/whatson

http://www.fbi.gov

Design Themes

- Redundancy
- Self-assessment
- Fail-over at multiple levels
- Robust algorithms

FirstPoint – DNS (e.g., Yahoo!)

- Selects from among several mirror sites operated by content provider

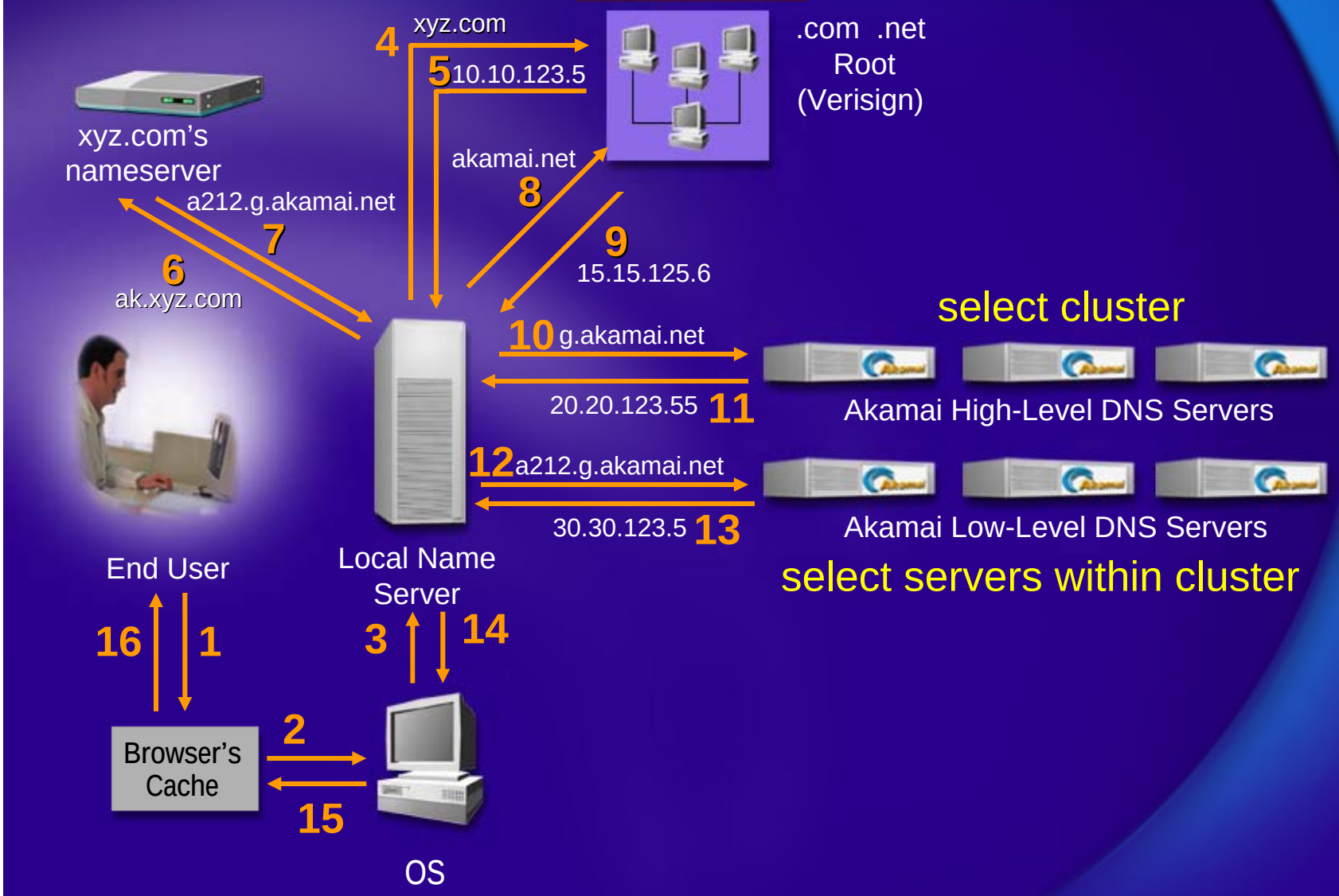
Embedded Image Delivery (e.g., Amazon)

Embedded URLs are Converted to ARLs

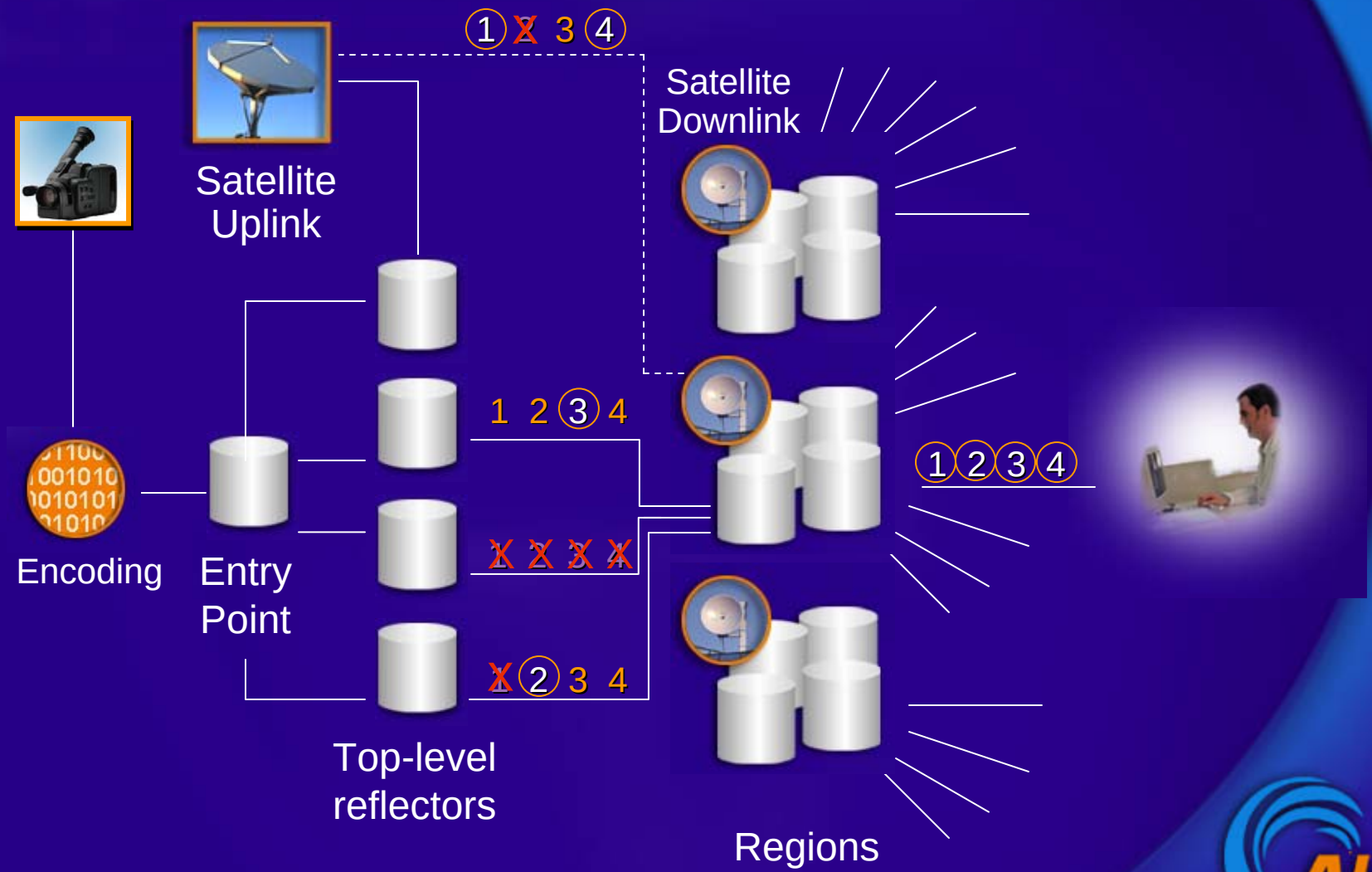
```
<html>
<head>
<title>Welcome to xyz.com!</title>
</head>
<body>
<img src="http://www.xyz.com/logos/logo.gif">
<img src="http://www.xyz.com/jpgs/navbar1.jpg">
<h1>Welcome to our Web site!</h1>
<a href="page2.html">Click here to enter</a>
</body>
</html>
```

ak

Akamai DNS Resolution

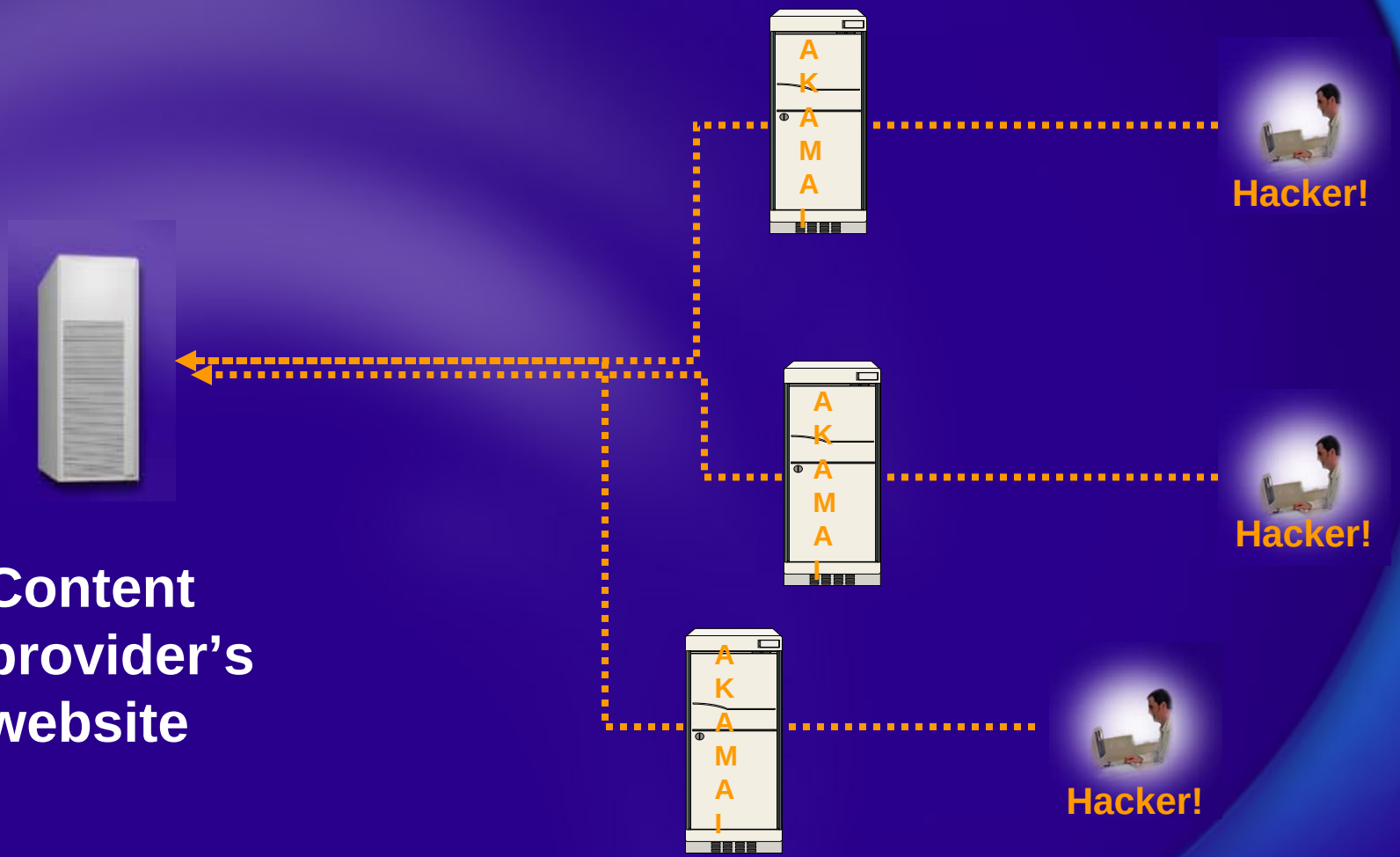


Live Streaming Architecture



Akamai

SiteShield (www.fbi.gov)



Part II: Failures

1. Hardware
2. Network
3. Software
4. Configuration
5. Misperceptions
6. Attacks

Hardware / Server Failures

Linux boxes with large RAM and disk capacity, Windows servers

Sample Failures:

1. Memory SIMMS jumping out of their sockets
2. Network cards screwed down but not in slot
3. Etc.

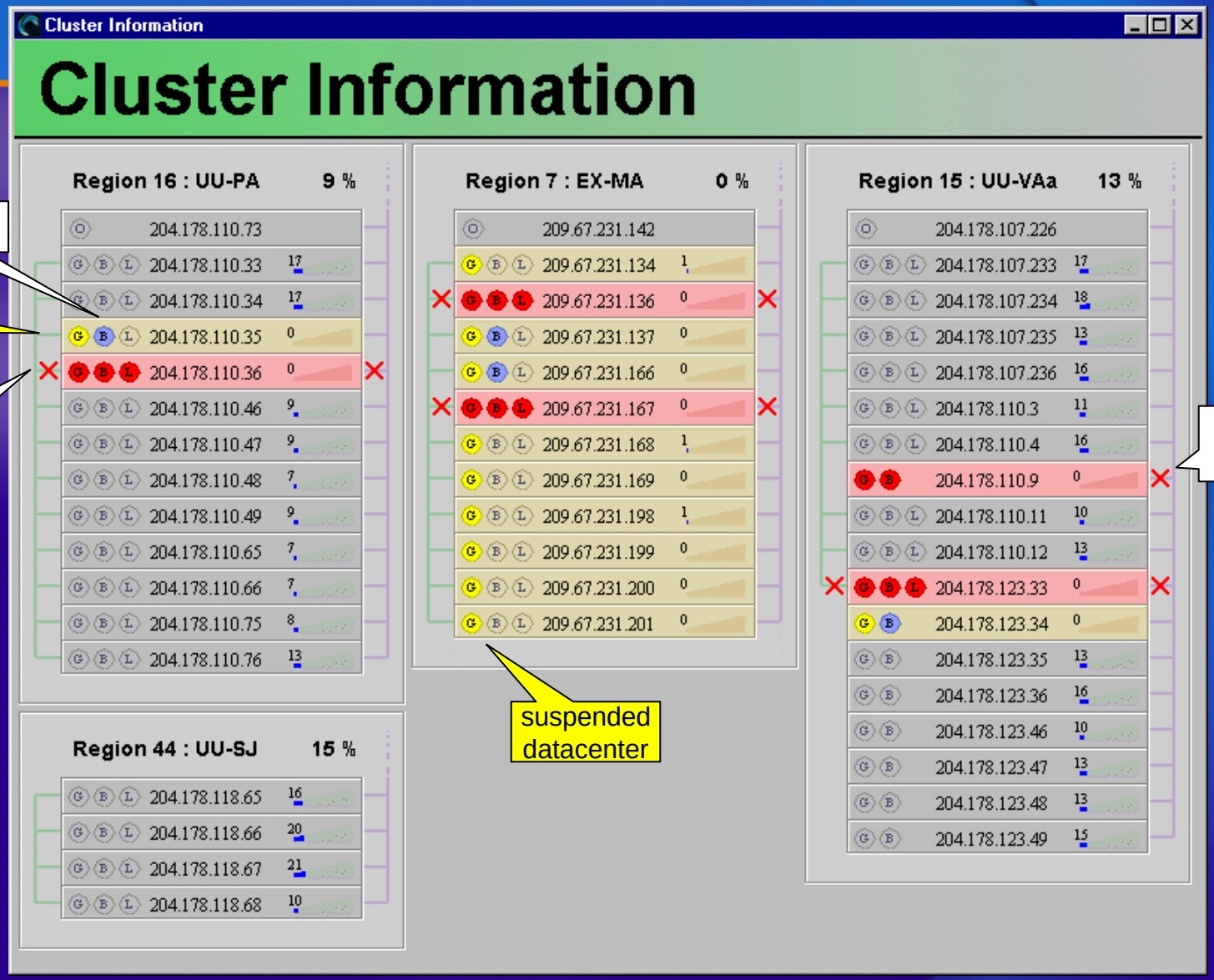
Akamai Cluster

Servers pool resources

- RAM
- Disk
- Throughput


Akamai

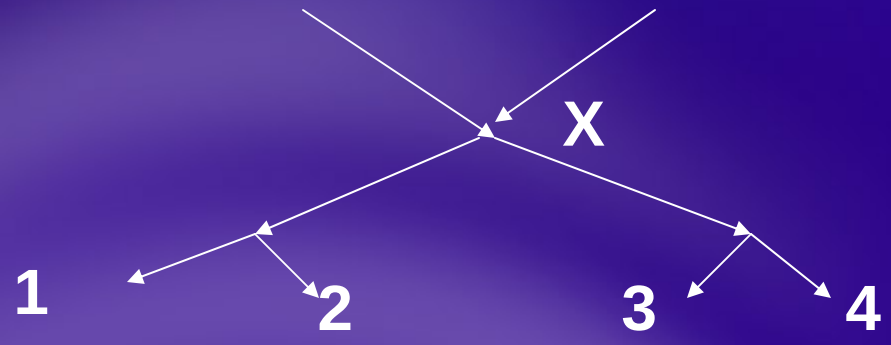
View of Clusters



Network Failures

E.g., congestion at public and private peering points, misconfigured routers, inaccessible networks, etc., etc., etc.

Core Points



- Core point X is the first router at which all paths to nameservers 1, 2, 3, and 4 intersect.
- X can be viewed as the straddling the core and the edge of the network.

Core Points

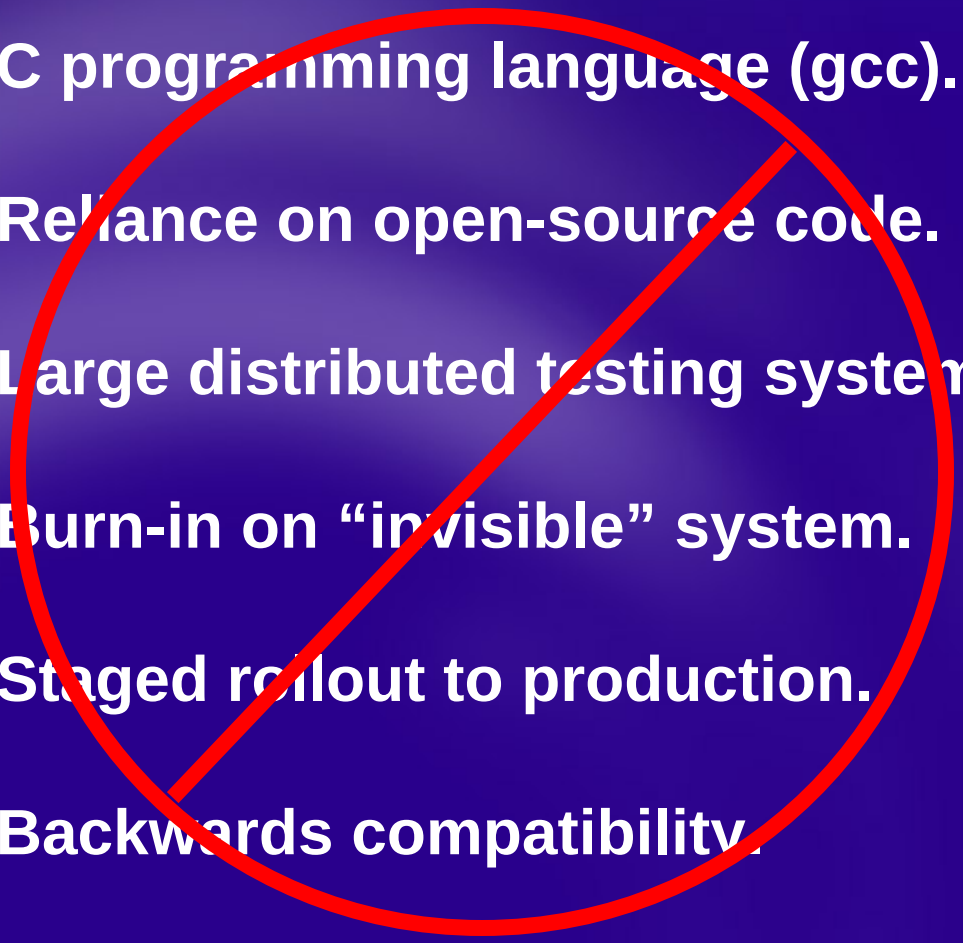
500,000 nameservers
reduced to

90,000 core points

7,000 account for 95% end-user load

Engineering Methodology

- C programming language (gcc).
- Reliance on open-source code.
- Large distributed testing systems.
- Burn-in on “invisible” system.
- Staged rollout to production.
- Backwards compatibility.



Perceived Failures

Examples

1. Personal firewalls
2. Reporting tools
3. Customer-side problems
4. Third-party measurements

Cascading Failures

MTU adjustment problem in Linux 2.0.38 kernel

Linux 2.0.38 crashes when TCP connection forces it to reduce MTU to approximately 570 bytes.

Someone in Malaysia configured a router to use this value as its MTU.

Client connecting through the router caused a cascade of Akamai servers to fail.

Attacks

- 8Gb/s attack inflicted on Akamai customer, October 2003
- Attack on Akamai FirstPoint DNS system, July 2004

Lost in Space

The most worrisome “attack” we faced:

One of our servers started receiving properly authenticated control messages from an unknown host.

Fortunately, the messages were not formatted correctly and were discarded by our server.

After two days of investigation, we discovered that the “attacker” was an old server we had lost track of, trying to rejoin the system.

It had been sending these messages for months before we noticed!