# The Akamai Network : A Platform for High-Performance Internet applications

Erik Nygren, Ramesh K. Sitaraman, Jennifer Sun
Proc. of ACM SIGOPS '10, 44(3):2-19, Jul. 2010

Presented by: Zaina and Shuang

## Overview

- Introduction
  - Internet Application Requirements
  - Internet Delivery Challenges
- What is a CDN? And its overview
- CDN for content, streaming media, and application delivery
- Platform Components
- Example
- Overall benefits and Results
- Assumptions, Challenges, and followup
- Q&A

## The Internet is expanding every second

- Internet's magnanimity brings with it challenges of performance, reliability, and scalability
- Any outage can cost millions of dollars & bad reputation
- Internet's architecture just as-is is incapable to provide desired performance & reliability due to bottlenecks
  - Latency, packet loss, network outages, inefficient protocols & inter network friction
- Scalability : Online video, Live streaming, HD to global audience?
- We need to therefore bridge the gap between capabilities & future requirements

## Internet Delivery Challenges

- Peering point congestion: The middle mile - high cost & zero revenue points where networks interact with each other, due to lack of investment cause **packet loss & increase latency**

- Inefficient routing protocols: BGP is based only on AS hop count, & is inefficient in times of failover, human errors & foul play leading to r**oute flapping, bloated paths & outages**.

- Unreliable networks: Due to natural & accidental reasons, & fragile peering there may be outages.
  - Eg: Southeast asia & Middle East 75% reduction
  - Sprint & cogent depeering, 3500 networks affected
  - BGP hijacking - global YT blackout by Pak

# Internet Delivery Challenges (contd.)

- Inefficient communications protocols
  - TCP designed for reliability & congestion avoidance, BUT it carries significant overhead, & has suboptimal performance.
  - TCP requires an ACK for every packet being sent, bottleneck in videos and large files
  - HTTP requests require multiple round trips
  - Limit on # parallel connections
  - Throughput ∝ 1/RTT

**Table 1: Effect of Distance on Throughput and Download Time**

| Distance (Server to User) | Network RTT | Typical Packet Loss | Throughput | 4GB DVD Download Time |
|---|---|---|---|---|
| Local: <100 mi. | 1.6 ms | 0.6% | 44 Mbps (high quality HDTV) | 12 min. |
| Regional: 500–1,000 mi. | 16 ms | 0.7% | 4 Mbps (basic HDTV) | 2.2 hrs. |
| Cross-continent: ~3,000 mi. | 48 ms | 1.0% | 1 Mbps (SD TV) | 8.2 hrs. |
| Multi-continent: ~6,000 mi. | 96 ms | 1.4% | 0.4 Mbps (poor) | 20 hrs |

# Internet Delivery Challenges (contd.)

- Scalability
  - Internet application owners/providers need to have exactly enough resources.
  - Underprovisioning → Potential Business loss; Overprovisioning → Waste of $$ on unused infrastructure
  - Ensuring adequate n/w b/w across **all** points between end users & the application
- Application limits and slow adoption to change
  - End users' software restrictions
  - IE 6,7,8 etc

# Solution to our Problems??? - **CDN!!**

- Started by caching static site content at the edge of the internet, close to end users
- Today: web & IP based applications, media delivery networks for live HD & on demand media, and Edge computing networks
- Additionally: maintain visibility & control for enterprises
- Also robust security, logging, SLAs, diagnostics, reporting & analytics, management tools

# Delivery Network: Overview

- Delivery Network = Virtual Network
  - s/w layer over the Internet
  - Deployed on widely distributed h/w
- Aimed at providing reliability, performance, security & scalability
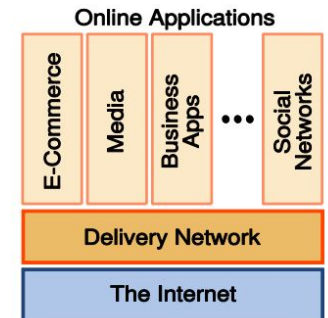- Advantage: Works over Internet as-is with no client software or network changes



Figure 2: A delivery network is a virtual network built as a software layer over the Internet that is deployed on widely distributed hardware.

## Anatomy of a delivery network

1. Mapping system: URL → IP address of an edge server
2. Edge servers ∈ Edge server platform serves requested content (large distributed system with tens of thousands globally deployed servers)
3. Responding to the request
   a. Either cached content
   b. To be fetched from the origin via the **transport system**
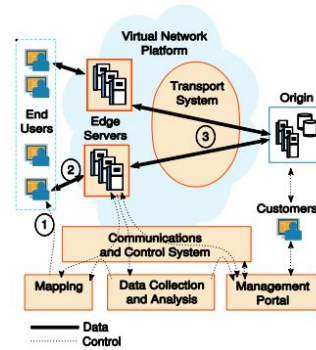


## Anatomy of a delivery network (contd.)

4. Communications & control system
5. Data collection & analysis system: collecting & processing data
6. Management portal
   a. Configuration management portal
   b. Users interaction visibility

** Specific design for each delivery system changes based on the purpose it serves
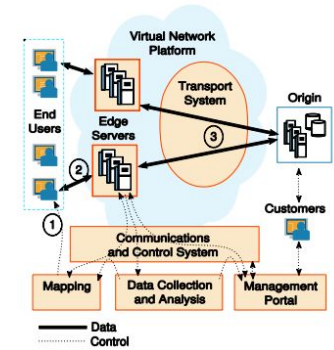


## System Design Principles

● Design for reliability: *≈100% E2E availability, no SPOF.*

● Design for scalability: *More data, traffic, content & customers*

● Limit human management necessity: *ensuring easy scalability, and automatic fault tolerance, deployment & configuration*

● Design for performance: *better response times, cache hitrate & network resource utilization*

## High performance Streaming & Content delivery networks

Key: Minimize long haul communication through middle bottleneck layers, Hence large number of distributed servers are as "close" as possible to end users

● Video-grade scalability
   ○ Speed, reliability, E2E scalability.
   ○ HD video has reached unimaginable orders of magnitude, live content being accessed in parallel, and continuous up & download.
● Throughput across the encoders to servers to end users is important.

- Streaming Performance
  - Streaming availability
  - Startup time
  - Frequency & duration of interruptions
  - Effective bandwidth
  - Optimization of packet loss, jitter, frame loss, RTT & E2E delay

- Transport System
  - Tiered Distribution: Cold/Infrequently accessed content
    - Edge clusters are connected to several parent clusters, and retrieves content from the parent cluster when it doesn't have the data ready
    - Reduces the load on the origin server
    - Origin only connected to a few dozen parent clusters that are categorized

- Transport System (contd.)

  - Overlay Network: Live Streaming
    - Live Stream is captured, encoded & sent to clusters of servers: Entrypoints
    - Entrypoints transport stream packets to edge servers in a publish subscribe model
    - Reflectors are intermediaries between entrypoints & edge clusters, making multiple copies of each received stream enabling rapid replication.
    - Reflectors also provide multiple paths between edge clusters and entrypoints.
    - Transport system chooses the best path or multiple link disjoint paths that are most efficient.
    - Prebursting: reducing startup time

# High Performance Application Delivery Networks

- Transport system for Application Acceleration: non cacheable customer content, retrieving content from the origin server
  - Using Akamai's **highly distributed edge servers** as a high performance overlay network
  - End user mapped to an edge server connects to an Akamai m/c near the origin server

  - Performance Improvement Techniques
    - Path Optimization: alternate paths after analyzing topology of the overlay
    - Packet loss reduction: multipath and FEC techniques
    - Transport protocol optimizations: Proprietary transport layer protocols
    - Application optimizations: Parse & prefetch embedded content; Content compression

# High Performance Application Delivery Networks (contd.)

- Distributing Applications to the Edge: EdgeComputing services
  - Cloud computing: resources are allocated on demand and near the end user
  - Challenges:
    - Session management
    - Security sandboxing
    - fault management
    - distributed load balancing
    - Resource monitoring and management
    - Testing and deployment

## High Performance Application Delivery Networks (contd.)

- Applications that can run in this fashion are based on
  - Content aggregation/transformation
  - Static Databases
  - Data collection
  - Complex applications: that run the presentation layer on the edge

- High performance, scalability, fault tolerance
- Quick development, and deployment

## Platform Components

- Edge Server Platform

- Mapping System

- Communications and Control System

- Data Collection and Analysis System

- Additional Systems and Services
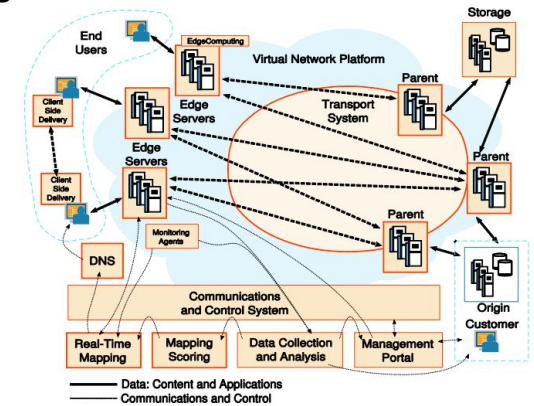


## Platform Components

- Edge Server Platform
  - Process end user requests and serve the requested content
  - Configurable via *metadata configuration*
    - EdgeComputing
    - Cache control & Cache indexing
    - Response to origin server failure
    - …
  - Make it simple to evolve platform to meet customers' changing needs
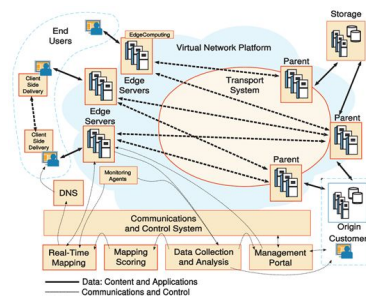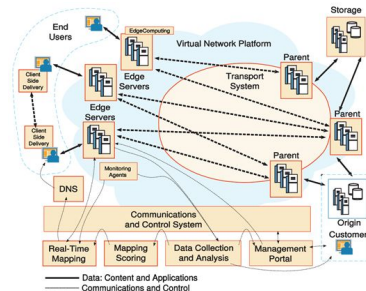


## Platform Components

- Mapping System
  - Scoring mapping
    - Create a current topological map capturing the connectivity of the entire Internet
    - Enable immediate response to Internet faults

  - Real-time mapping
    - Create the actual maps used by the Akamai
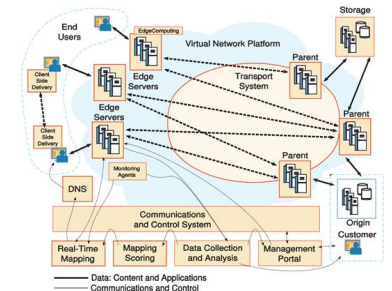    - Mapping to cluster & Mapping to server
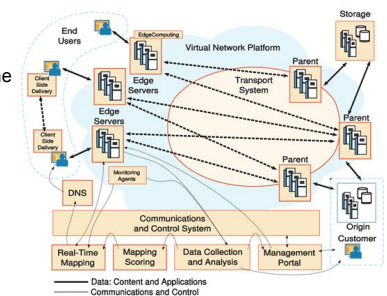
## Platform Components

- Communications and Control System

  - Real-time distribution of status and control information

  - Dynamic configuration updates
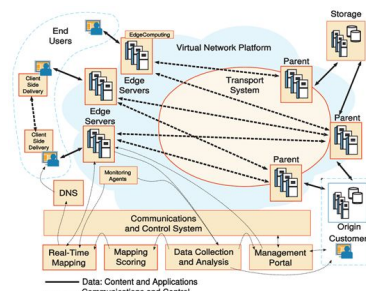
  - Key Management Infrastructure

  - ...



## Platform Components

- Data Collection and Analysis System

  - Log collection

  - Real-time data collection and monitoring

  - Analytics and Reporting



## Platform Components

- Additional Systems and Services
  - DNS
  - Monitoring Agents
  - Global Traffic Manager
  - Storage
  - Client Side Delivery
  - Management Portal



## Example: Multi-level Failover

- Machine failure
  - Machine is down within an edge cluster

  - Another machine starts responding to the IP address of the failed machine

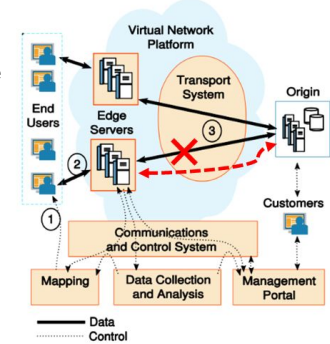  - Mapping is updated and redirects new requests
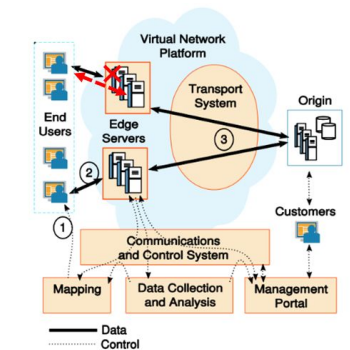
## Example: Multi-level Failover

- Cluster failure
  - An entire cluster fails

  - Mapping is rapidly updated and do a new assignment
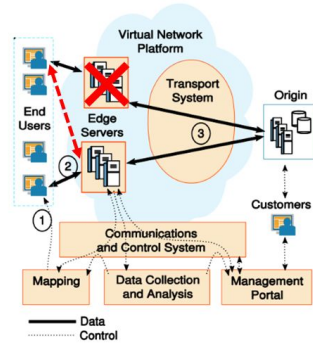


## Example: Multi-level Failover

- Connectivity failure
  - Connectivity between the origin server and the edge degrades

  - Use path optimization to find alternate paths



## Overall Benefits and Results

- Content and Streaming Delivery (3 examples)
  - New York Post
    - Handle its flash crowd
    - 20X ↑ faster for home page downloading

  - U.S. Government
    - Protection against DDoS attack (8 years' worth of traffic in one day)
    - Maintain a normal operation of U.S. gov. websites

  - MySpace
    - Offload 98% of its traffic to Akamai
    - 2.6X ↑ speedup to U.S. users, 6X ↑ to international users

## Overall Benefits and Results

- Application Delivery (3 examples)
  - Enterprise applications
    - Help enterprise overcome performance and reliability challenges
    - Global performance improvements from 100% to 700%

  - Large file transfers
    - 5X ↑ large file (2 GB) transfer (Europe to the US)
    - 4-5X ↑ large files transfer using SFTP
    - 2.3X ↑ file transfers over a VPN (India and the US)
  - EdgeComputing: Sony Ericsson
    - Phone configurator, shopping cart, and dealer locator—to the edge
    - Response time 3X↓ & infrastructure needs 65% ↓
    - Application availability ↑ from 92% to 100%

## Advantages

- Caching
  - Cache content from a centralized server to local cluster
  - Reduce load on the origin server
  - Absorb and mitigate service attack

- Faster page loads and faster transactions

- Strengthen customer loyalty

## Weaknesses & Challenges

- Difficult to configure
  - complicated API & hard-to-understand configs

- Operate too much as a black box
  - customers have little transparency into how it works

- Challenges
  - Some content provider builds their 'DIY' CDN
  - Lots of competitors

## Competitors

- Other CDN providers
  - MaxCDN
  - Level 3
  - Amazon Cloudfront
  - Google Pagespeed
  - CloudFlare
  - Verizon DMS
  - ...

## Akamai in the news

- 216,000 servers in over 120 countries and within more than 1,500 networks around the world
- Eighty-five percent of the world's Internet users are within a single "network hop" of an Akamai CDN server.
- Stock prices are dropping, and the revenue for Q3 is only 5% up year-over-year
- Apple, Microsoft & Facebook are moving much of their traffic to in-house CDNs
- Daily web traffic reaching more than 30 Terabits per second.
- Delivers nearly 3 trillion Internet interactions each day.
- Enables more than $250 billion in annual e-commerce for its online retail customers

Discussion / Q&A