On efficient delivery of web content

Mangesh Kasbekar
Akamai Technologies

On efficient delivery of web content

Larger context: Improving the energy efficiency of an entire organization

- The importance of understanding the unique profile of the organization
- What is Akamai's unique story?

Three short stories

- The Akamai Content Delivery Network
- Trends in web content and traffic
- Opportunities for energy efficiency

- Questions

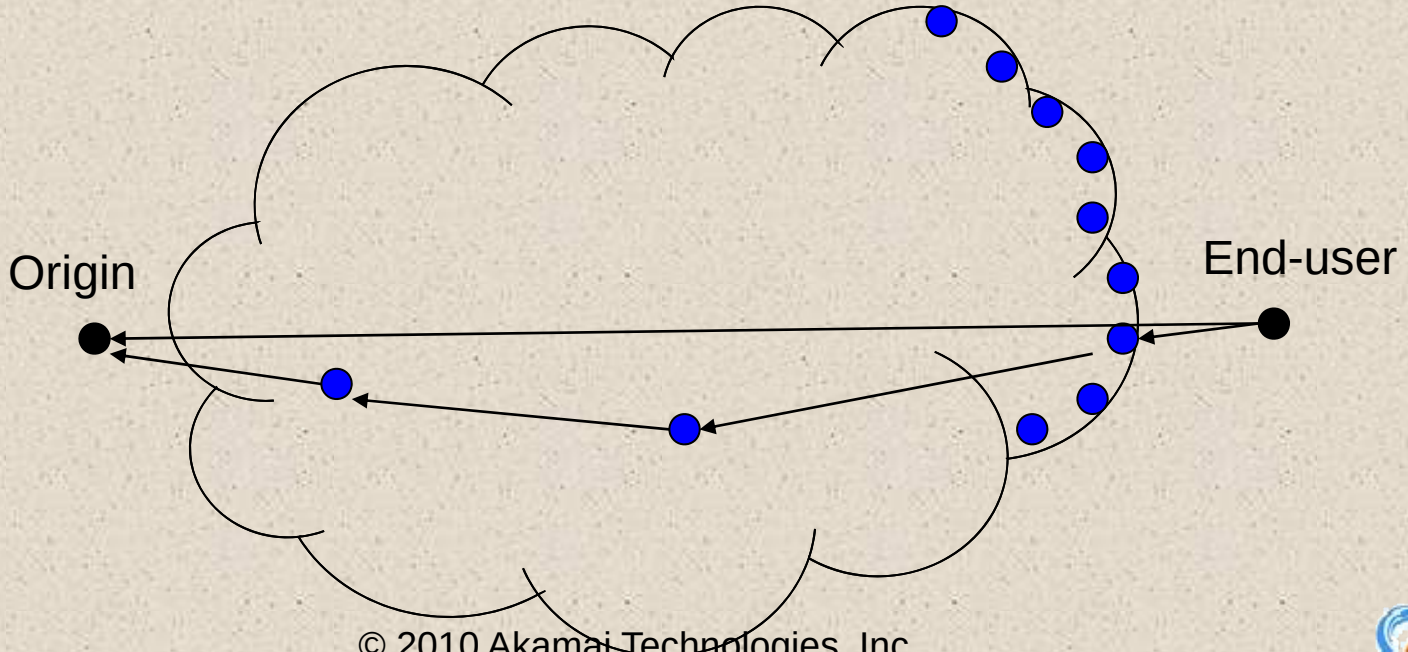
Akamai

First short story

- The Akamai Content Delivery Network
- Trends in web content and traffic
- Opportunities for energy efficiency

The Akamai CDN

- What is a CDN
- Why use a CDN
- What does a CDN offer



The Akamai CDN

- Akamai deployment
  - 66,000+ servers
  - 120+K CPUs, 400+K disks
  - Located in 1000+ datacenters in rented space
- Akamai Traffic
  - Petabytes delivered each day
  - Recent traffic record of 3.5 Tbps
  - Growing extremely fast

Energy efficiency of this massively distributed platform

- Areas that we can't control
  - Datacenter PUE
  - Energy source
  - Hardware's energy efficiency
  - Growth in the web traffic
- Areas that we do control
  - Server software efficiency
  - Operational practices around traffic management
- Currently, datacenters are inefficient
  - PUE of 1.7-3.0
  - Any savings in the reduction of machine count are amplified by the same factor

The Akamai CDN

- How best to control what we can control
  - Thorough understanding of the workload

Second short story

- The Akamai Content Delivery Network
- Trends in web content and traffic
- Opportunities for energy efficiency

Trends

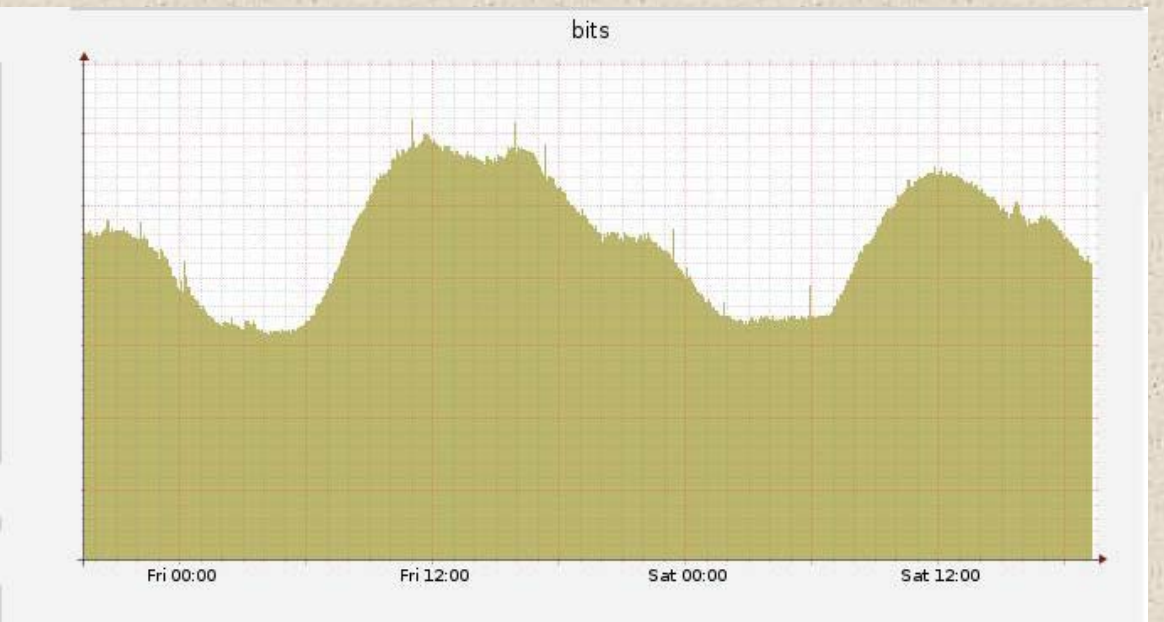
General web traffic

against

- Software downloads
- Transactional application acceleration
- Large-footprint long-tail content
- Media downloads and full movie delivery

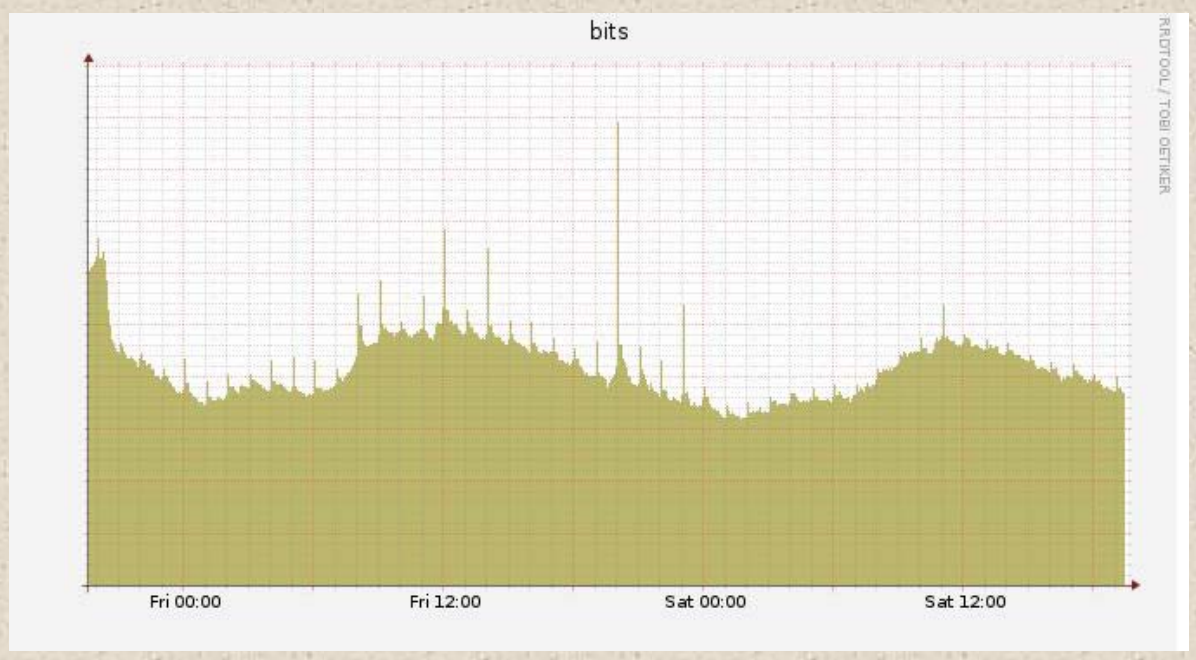
General web traffic

- Selling point: High performance for cacheable content
- Content type: html, images, stylesheets, javascripts
- Peak to valley ratio = 1.8
- Peak traffic at 1000-1600 EST



Software downloads

- Selling point: software downloads and frequent updates
- Requirements: cheap delivery and origin offload
- Content type: very large files, high cache hit rates
- Peak to valley ratio = 1.4
- Peak traffic hr: early in the morning, but fairly flat



Akamai

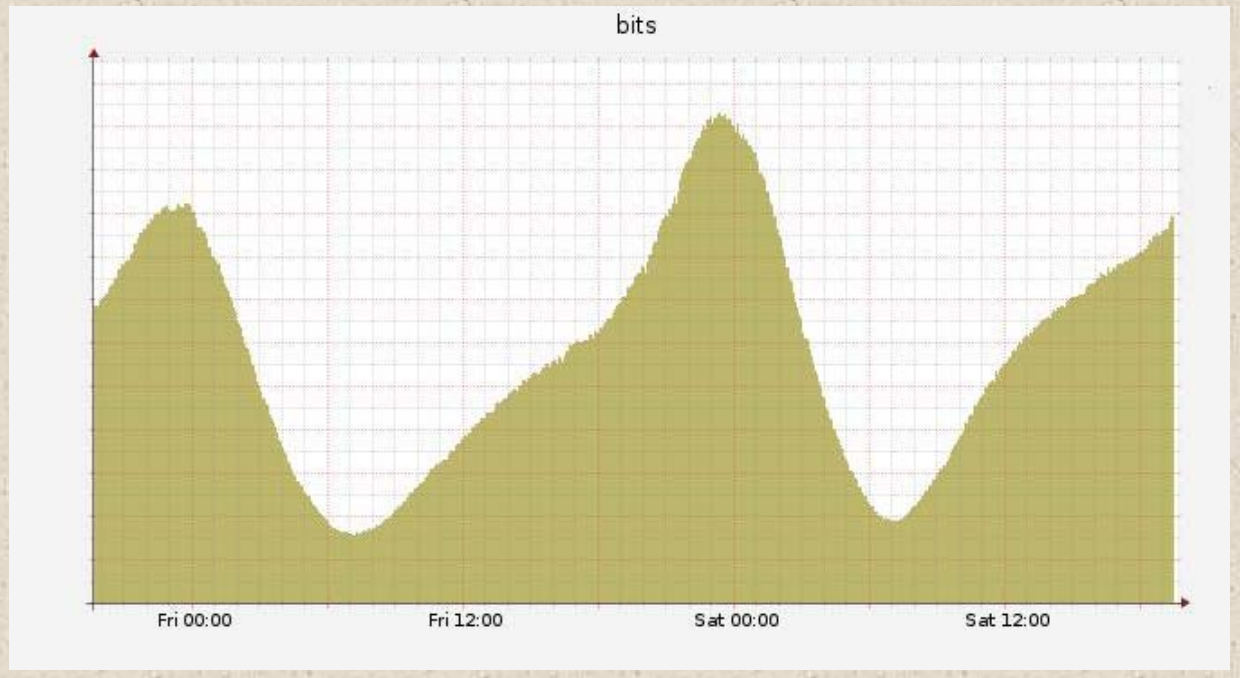
Large-footprint long-tail traffic

- Selling point: delivery of user-generated content
- Requirement: Extremely high origin offload
- Content type: thumbnails - small videos, bad cacheability
- Peak to valley ratio = 2
- Peak traffic at 1200-1800 EST

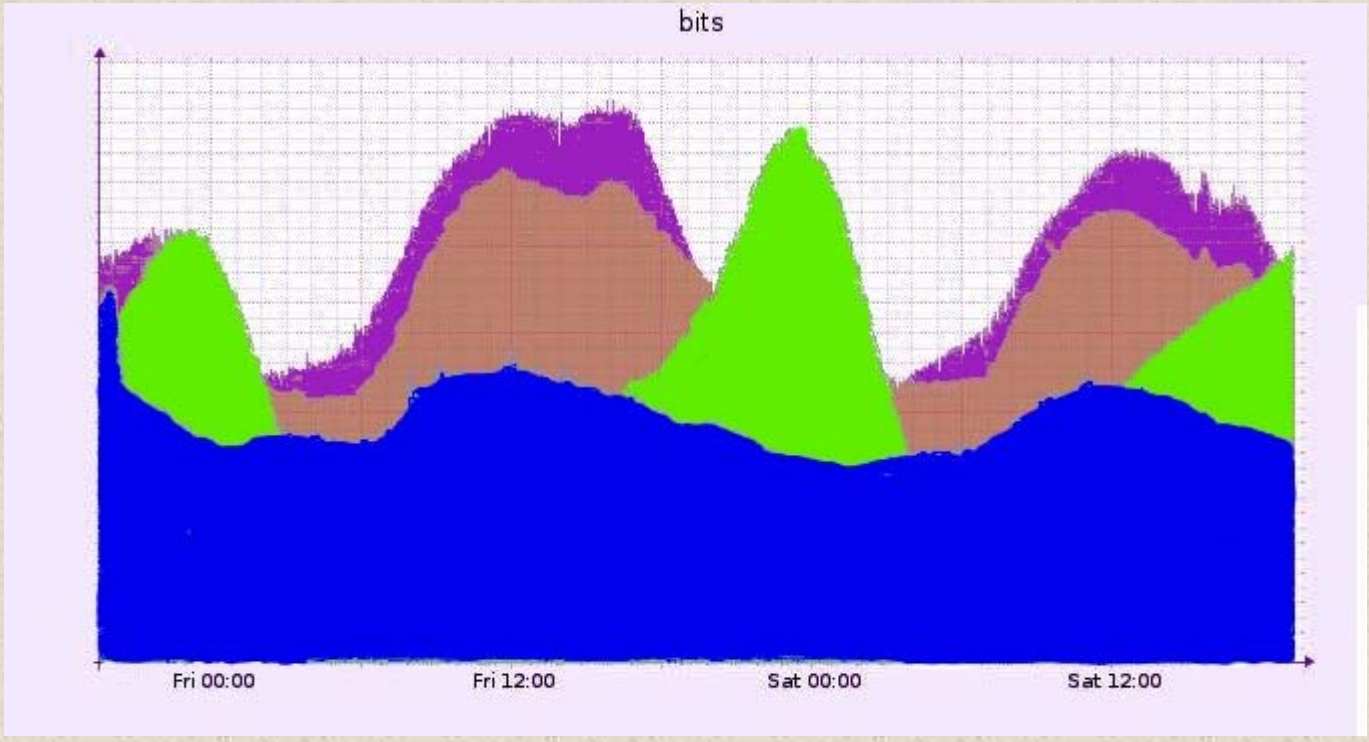
Akamai

Media downloads and movie delivery

- Selling point: movie and media delivery
- Requirement: real-time performance
- Content-type: very large media files
- Peak to valley ratio = 7
- Peak traffic at 2200-0100EST



Combined traffic pattern



Not to scale

Metrics

- Different resources in a server
  - CPU, Disk I/O, Memory, Network Bandwidth, Various internal limits
- Define U = utilization of the most constrained resource for a given level of traffic
- Define a metric for resource-intensiveness of a class of traffic, and normalize:
  - R = U * constant / traffic

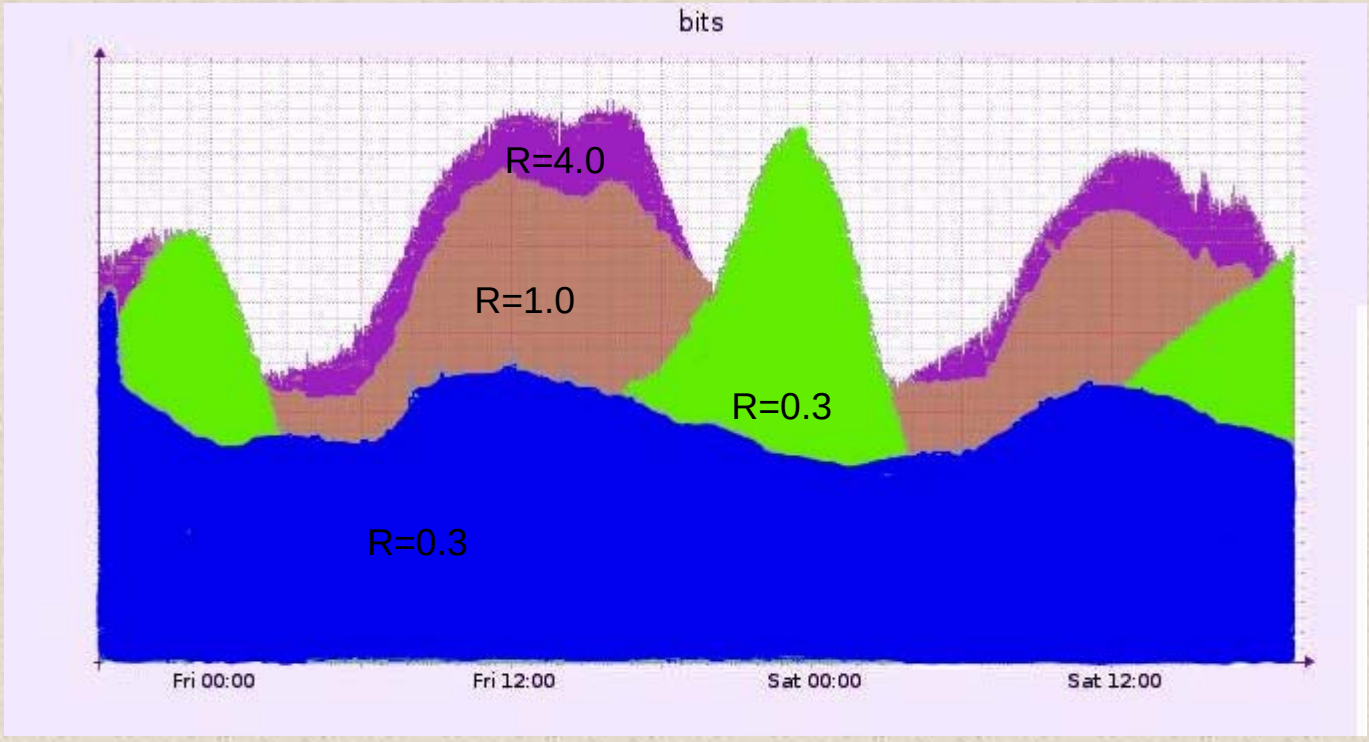
R for different classes of traffic

Category	R	CPU	DISK
General	1.0	100%	60-70%
Software downloads	0.3-0.4	100%	<50%
Large-footprint long-tail	1.5-4.0	70-95%	100%
Media downloads	0.3-0.4	100%	<50%

Low R = High efficiency

Akamai

Combined traffic pattern



Not to scale

Akamai

Trends

- Summary
  - Four large classes of traffic
  - Different requirements from the customer
  - Different traffic patterns
  - Different levels of resource intensiveness

Third short story

- The Akamai Content Delivery Network
- Trends in web content and traffic
- Opportunities for energy efficiency

Opportunities

- Effect of traffic consolidation
- Hosting in energy efficient data centers, new generations of hardware, OS, and software
- Using fewer servers to serve the same traffic
  - Micro: Server efficiency projects
  - Macro: Traffic mixing

- Traffic consolidation

  Several thousand websites hosting their own web-server infrastructures v/s all serving their traffic from a common CDN platform

  – Better server utilization for the CDNs due to flatter peaks and lower peak to valley ratio
  – Shared headroom
  – Efficient platform

Improving things we don't directly control:

- Hosting in energy efficient data centers as they become available
- New generations of hardware, OS and software
  - New generation of processors
  - SSDs
  - Rejecting inefficient 3rd party software

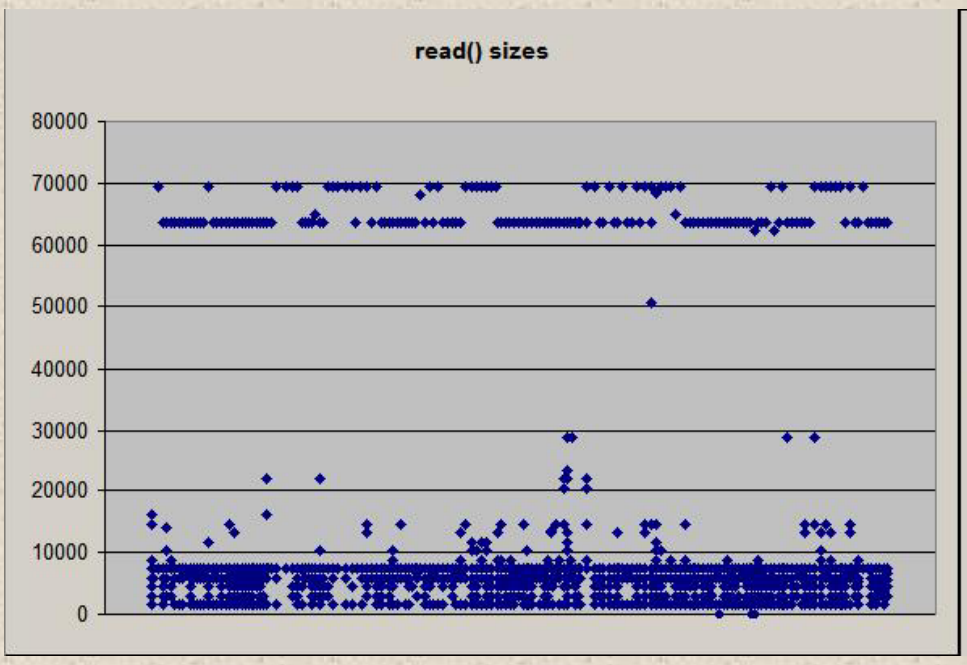
Micro: server efficiency projects

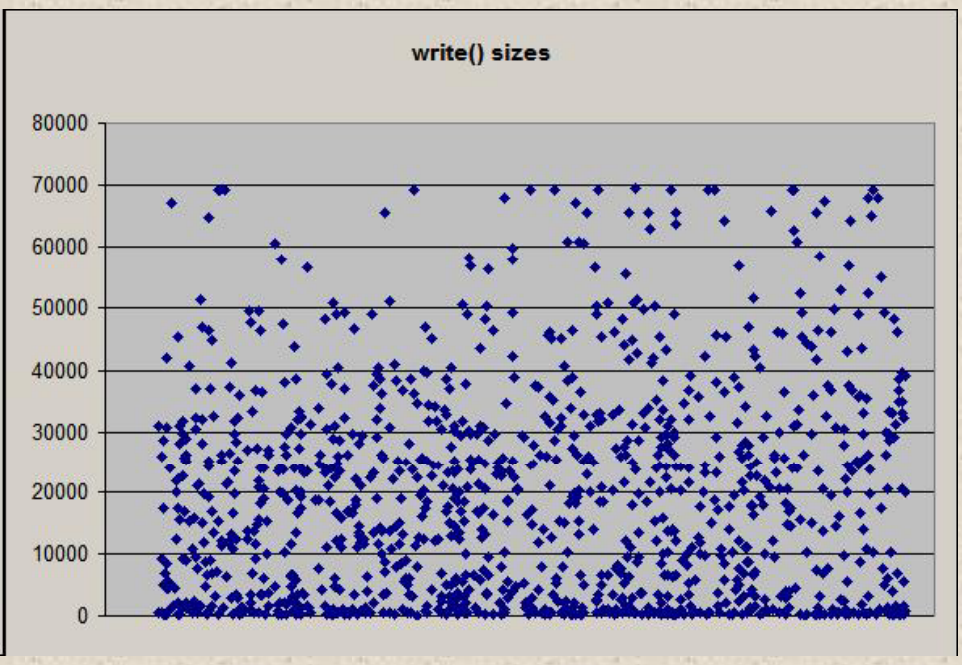
Reducing the usage of the most constrained resource in each traffic class.

2 examples:

- Reducing system calls related to network I/O for large files
- Using probabilistic methods for making caching decisions for large-footprint long-tail content

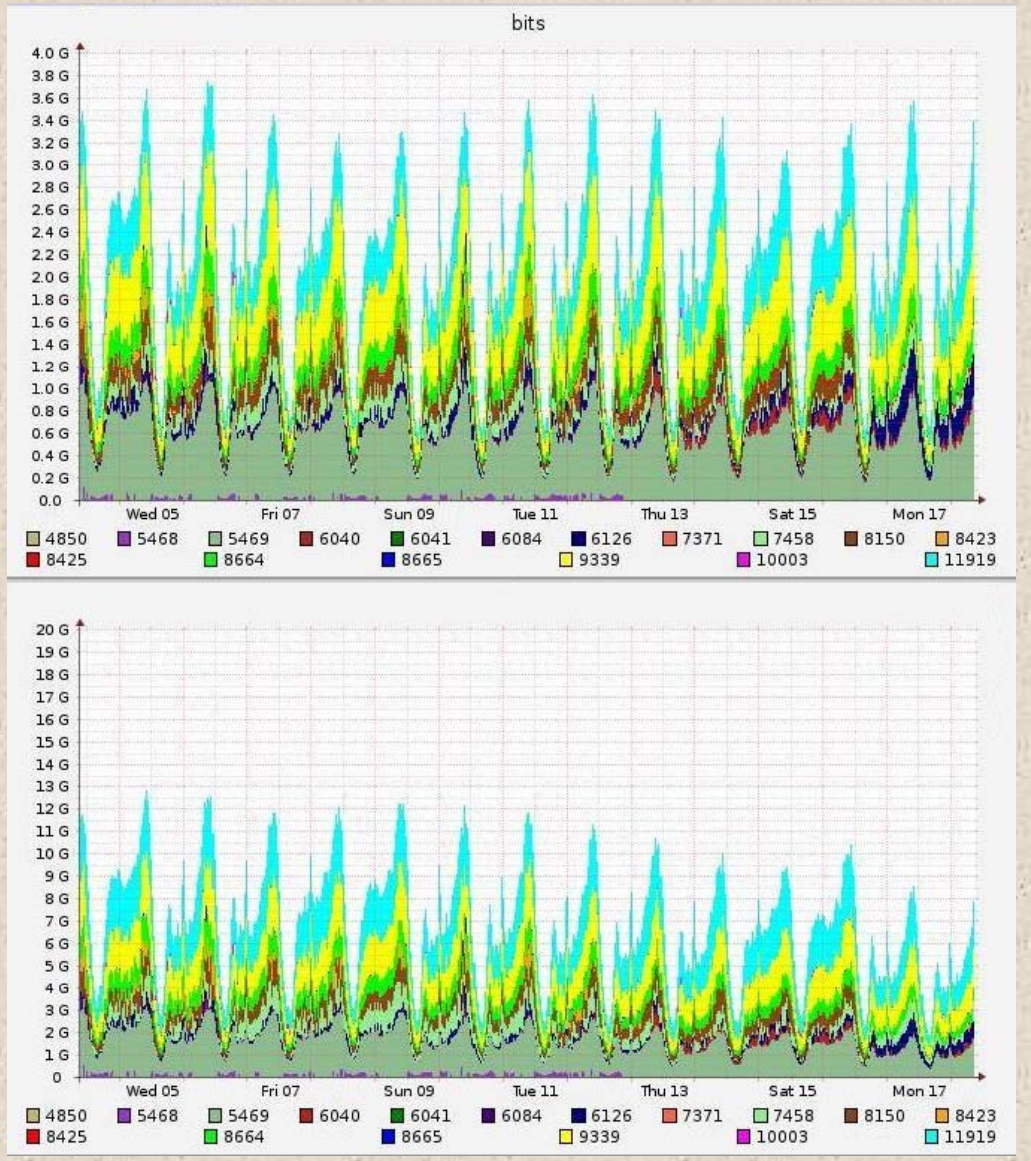
- Reducing network I/O calls
  - 64KB forward-read sizes
  - 64KB client-write sizes





Akamai

- Probabilistic caching for long-tail content



Macro: Traffic mixing

CDNs get all classes of web traffic, smart traffic mixing is very beneficial to minimize unusable resources

- Opportunities
  - Different peak hours
  - Different resource requirements
  - Different performance needs
- Challenges
  - Cache churn
  - Disk fragmentation
  - Precise traffic control by the global load-balancer

Conclusion

Akamai's unique story in improving its energy efficiency

- Main trends in web traffic
- Resource intensiveness characterization of traffic classes
- Reducing the number of machines needed to serve the same amount of traffic
- Smart traffic mixing