

*Review of:*

## An Analysis of Internet Content Delivery Systems

- Paper by:
  - Stefan Saroiu, Krishna Gummadi, Richard Dunn, Steven Gribble, Henry Levy (Univ. of Washington)
- Published in:
  - 5<sup>th</sup> Symposium on Operating Systems Design and Implementation (USENIX OSDI)
  - December, 2002 (Boston, MA)
- Review by:
  - Bharath Madhusudan

## Style of Paper

- Introduction
- Overview of Content Delivery Systems
- Measurement Methodology
- Trace Analysis
- CDN/P2P Cacheing
- Conclusions

## Introduction

- Motivation for the work
  - An Increase in the use of specialized content delivery systems
    - P2P Systems, CDNs
    - A need to understand these systems
  - Goal: Analysis of collected traces to learn how they work
    - Web Traffic, Akamai, Kazaa, Gnutella

## Introduction – Overview Of CDNs

Increasingly dynamic nature of web content

-- Web caches are insufficient

**DNS Redirection**

-- TTL Field for Load balancing

**URL Rewriting**

<http://a388.g.akamaitech.net/7/388/21/fc35ed7f236388/cnn.com/images/hub2000/ad.info.gif>



## Introduction – Overview of P2P Systems

- An overlay of peers
  - Clients as well as Servers
  - Participation Voluntary
- Centralized database
  - Napster
- Flooding
  - Gnutella, Freenet
- Hybrid
  - Kazaa

## Trace Collection

- Passive Monitoring
  - Monitoring Port that sends copies of incoming and outgoing packets
  - Kernel packet filter

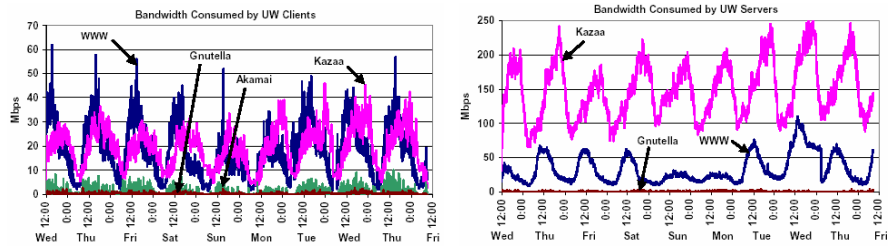
## Trace Interpretation

- **WWW Traffic**
  - Ports 80, 8080, 443
- **Gnutella**
  - Ports 6346, 6347
- **Kazaa**
  - Ports 1214
- **Akamai**
  - WWW Content Served by an Akamai Server

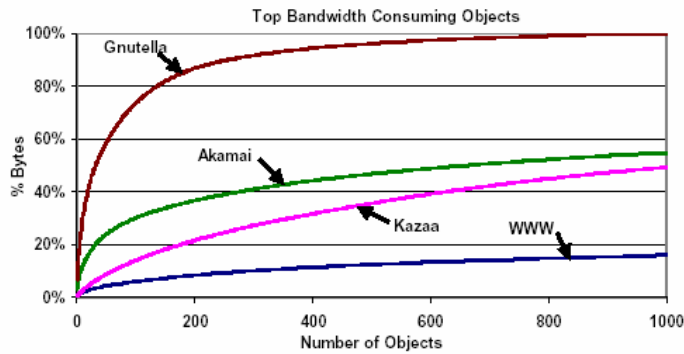
## Trace Summary Stats

	WWW		Akamai		Kazaa		Gnutella	
	inbound	outbound	inbound	outbound	inbound	outbound	inbound	outbound
<b>HTTP transactions</b>	329,072,253	73,001,891	33,486,508	N/A	11,140,861	19,190,902	1,576,048	1,321,999
<b>unique objects</b>	72,818,997	3,412,647	1,558,852	N/A	111,437	166,442	5,274	2,092
<b>clients</b>	39,285	1,231,308	34,801	N/A	4,644	611,005	2,151	25,336
<b>servers</b>	403,087	9,821	350	N/A	281,026	3,888	20,582	412
<b>bytes transferred</b>	1.51 TB	3.02 TB	64.79 GB	N/A	1.78 TB	13.57 TB	28.76 GB	60.38 GB
<b>median object size</b>	1,976 B	4,646 B	2,001 B	N/A	3.75 MB	3.67 MB	4.26 MB	4.08 MB
<b>mean object size</b>	24,687 B	82,385 B	12,936 B	N/A	27.78 MB	19.07 MB	19.16 MB	9.78 MB

## Client and Server b/w Comparison



## Top b/w Consuming Objects



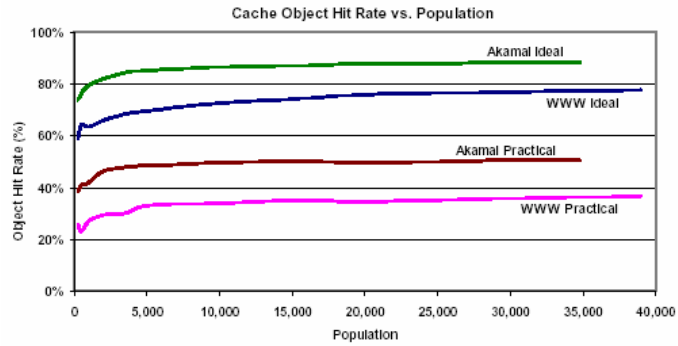
## Some Observations (Clients)

- Reverse Caching may be beneficial
- Small portion of clients account for large portion of the traffic
- Kazaa accounts for the greater fraction of HTTP bytes than WWW
- Kazaa request rate is far lower than WWW request rate
- Simultaneous open Kazaa connections exceed open WWW connections

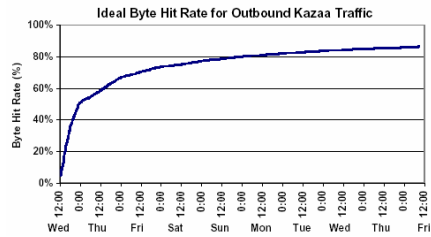
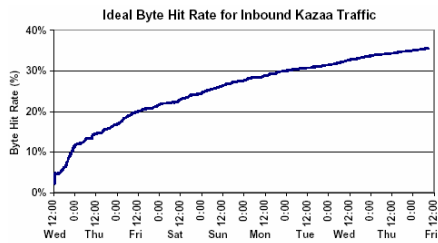
## More Observations (Servers)

- Small number of Internal Servers take up lots of b/w
- 938 out of 400K External Servers giving 50% of the WWW bytes (expected)
- 600 external peers out of 281K supply 26% of the Kazaa bytes to internal peers (!!)
- High failure rate (80%) for P2P
- Questions about scalability

# CDN Caching



# P2P Caching



## Conclusions

- P2P traffic accounts for majority of bytes transferred
- P2P objects much larger than Web objects
- Objects exhibit Zipf like popularity
- Small number of clients and servers hog most of the b/w
- Bandwidth requirement of Kazaa peer 90X single web client