On the Use and Performance of Content Distribution Networks

Yin Zhang

Joint work with

Balachander Krishnamurthy and Craig Wills

AT&T Labs Research, WPI

{yzhang,bala}@research.att.com, cew@cs.wpi.edu

ACM SIGCOMM Internet Measurement Workshop

November, 2001

Motivation

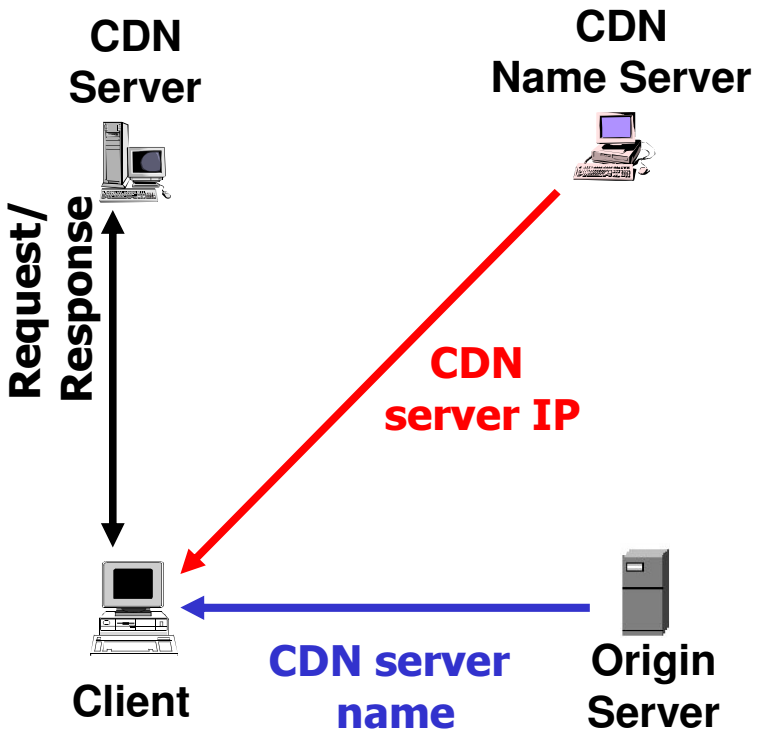
- What is a CDN?
  - A network of servers delivering content on behalf of an origin site
- State of CDNs
  - A number of CDN companies
    - E.g. Akamai, Digital Island, Speedera
  - Used by many popular origin sites
    - E.g., CNN, CNBC, …
- Little has been published on the use and performance of existing CDNs

Research Questions to Answer

- What CDN techniques are being used?
- What is the extent to which CDNs are being used by popular origin sites?
- What is the nature of CDN-served content?
- What methodology can be used to measure the relative performance of CDNs?
- What are specific CDNs performing both relative to origin servers and among themselves?

This talk tries to answer them based on a large-scale, client-centric study conducted in Sept. 2000 and Jan. 2001

What CDN redirection techniques are being used?



- Techniques examined
  - DNS redirection (DR)
    - **Full-site delivery (DR-F)**
    - **Partial-site delivery (DR-P)**
  - URL rewriting (UR)
  - Hybrid scheme (URDR)
    - **URL rewriting + DNS redirection**
- Techniques NOT examined
  - Manual hyperlink selection
  - HTTP redirection
  - Layer 4 switching
  - Layer 7 switching

How widely are CDNs being used?

- Sources of data

Type	Datasets	Date/Duration	Sites
Periodic crawl	HotMM127	2 months: Nov. & Dec. 2000	127
	URL588-MM500		1030
Proxy log	LMC	1 week in Sept. 2000	3
	NLANR	1 week in Jan. 2001	9

- CDN use by popular sites

Nov. 1999	1-2% out of ~600 [KW00]	
Dec. 2000	HotMM127: 31%	(Akamai: 98%)
	URL588-MM500: 17%	(Akamai: 85%)

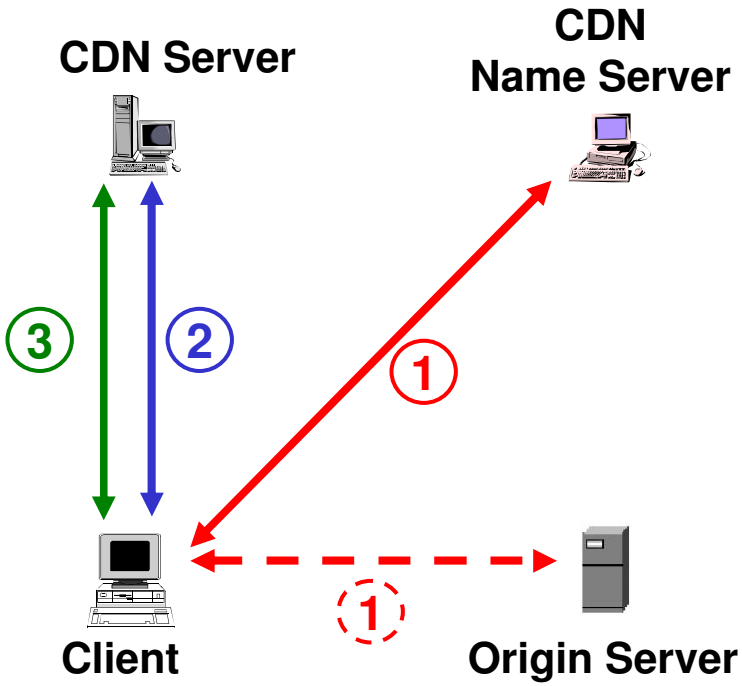
Nature of CDN-served Content

- Daily change characteristics of CDN-served objects

Dataset	HotMM127	URL588-MM500
#Objects	24.9K	75.0K
Prev. seen URL	89%	86%
Prev. seen URL w/ changes	2.2%	3.2%

- Nature of HTTP-requested CDN content
  - Images account for 96-98% CDN-served objects, or 40-60% CDN-served bytes
  - Akamai serves 85-98% CDN-served objects (bytes)
  - Cache hit rates of CDN-served images are generally 20-30% higher than non-CDN served images

Performance Study: Methodology



① Get CDN server IP address

① URL rewriting –
first get CDN server name

② Warm up CDN cache

③ Retrieve pages using "httperf"

Parallel-1.0 – 4 HTTP/1.0

Serial-1.1 -- 2 persistent HTTP/1.1

Pipeline-1.1 – 1 pipelined HTTP/1.1

General Methodology: From N client sites periodically download pages from different CDNs and origin sites.

Content for Performance Study

- Challenge:
  - Different CDNs have different customers.
    How to compare "apples" to "apples"?
- Solution: *Canonical Pages*
  - Create template page based on distributions of the number and size of embedded images at popular sites
    - In our study, we download 54 images and record download time for the first 6, 12, 18, 54 images.
  - For each CDN, construct a canonical page with a list of image URLs currently served by the CDN from a single origin site, that closely match the sizes in the template page.

Measurement Infrastructure

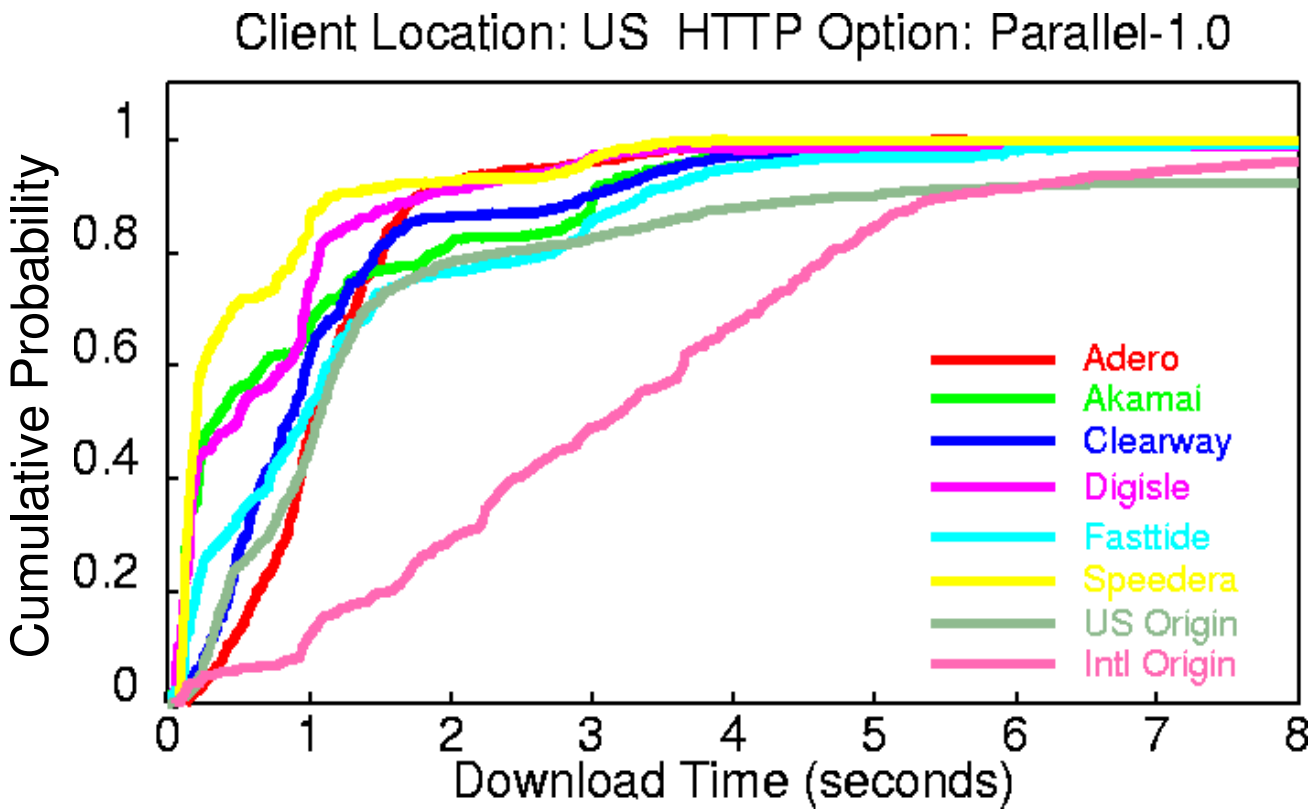
- CDNs

Technique	DR-F	DR-P	UR	URDR
CDNs	**Adero**	**Akamai, Speedera, Digital Island**	**Clearway**	**Fasttide**

**AT&T ICDS NOT tested due to conflict of interest.*

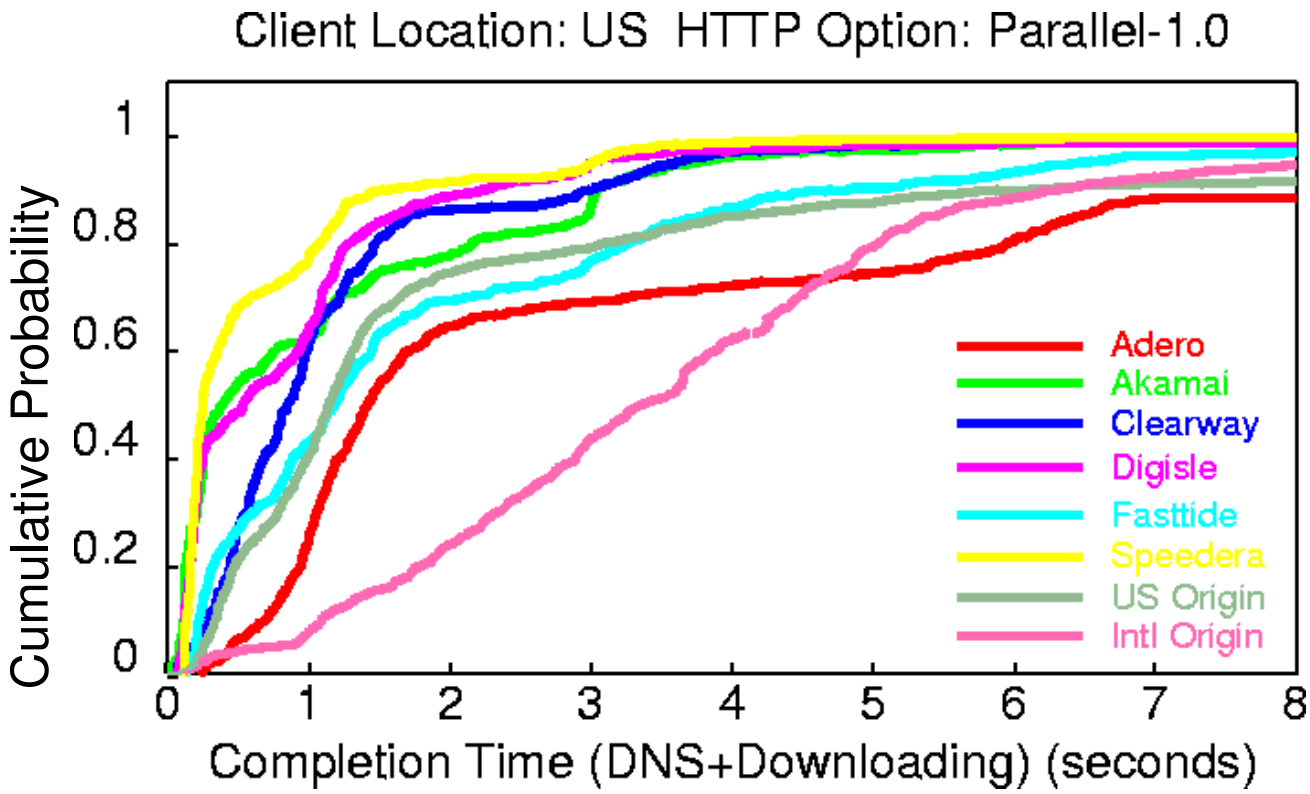
- Origin sites
  - US: Amazon, Bloomberg, CNN, ESPN, MTV, NASA, Playboy, Sony, Yahoo
  - International: 2 Europe, 2 Asia, 1 South America, 1 Australia
- Client sites
  - 24 NIMI client sites in 6 countries
    - NIMI: *National Internet Measurement Infrastructure*
    - Well-connected: mainly academic and laboratory sites

Response Time Results (I)
Excluding DNS Lookup Time



CDNs generally provide much shorter download time.

Response Time Results (II) Including DNS Lookup Time



DNS overhead is a serious performance bottleneck for some CDNs.

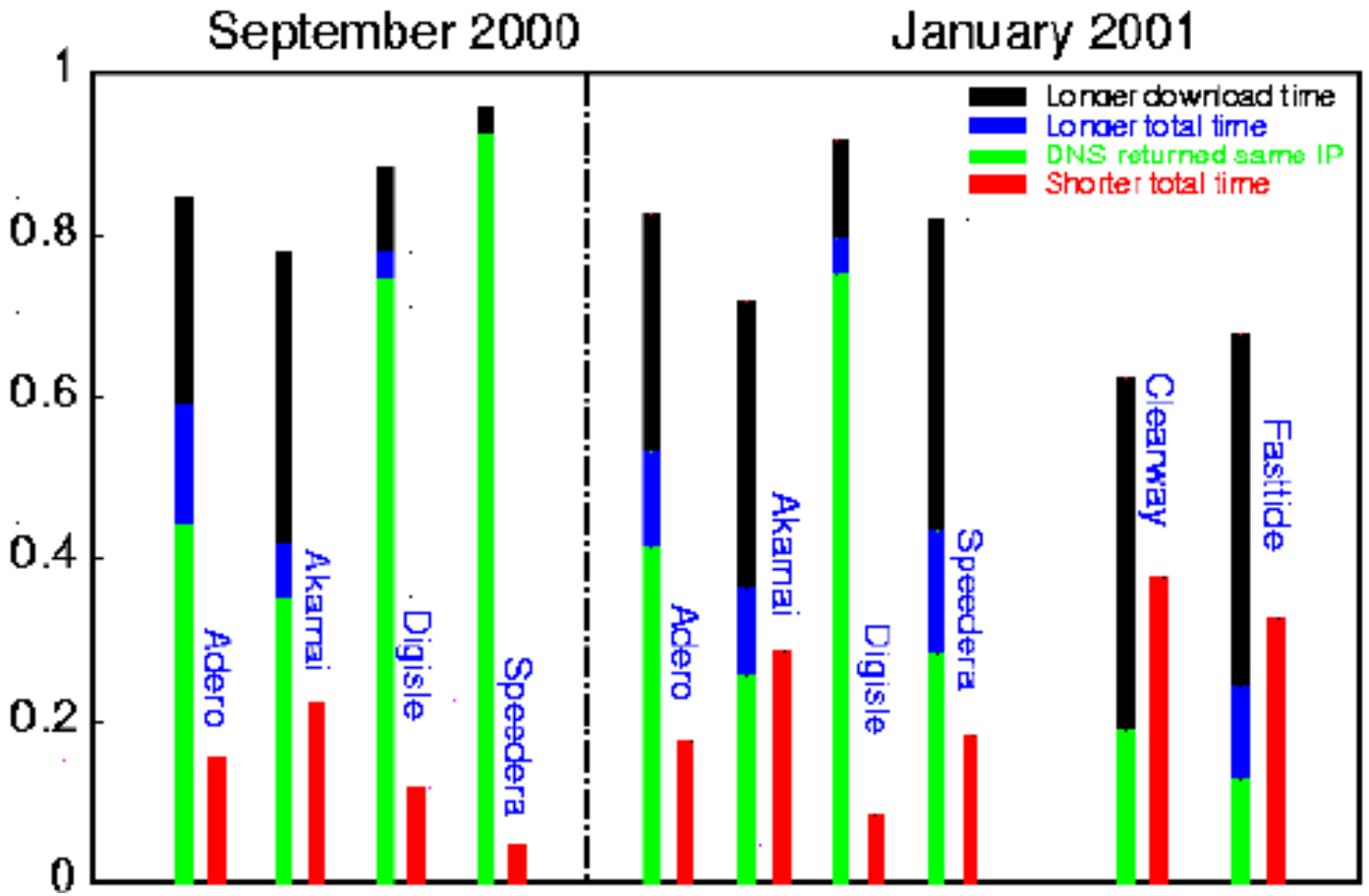
Impact of Protocol Options and the Number of Images

Mean Download Performance Range for Different Numbers of Images and Protocol Options (Jan. 2001)

Protocol Option	Site	Mean Download Time Range (sec.)			
		6 images	12 images	18 images	54 images
Parallel-1.0	CDN	0.26-0.76	0.40-1.23	0.58-1.53	1.49-3.31
	US Origin	1.63	2.45	3.40	8.42
Serial-1.1	CDN	0.27-0.53	0.42-0.81	0.61-1.13	1.46-2.52
	US Origin	1.06	1.46	1.96	4.87
Pipeline-1.1	CDN	0.26-0.50	0.37-0.67	0.47-0.88	1.09-2.04
	US Origin	*Partial Support*			

CDNs perform significantly better than origin sites, although reducing the number of images (e.g. due to caching) and using HTTP/1.1 options reduces the performance difference.

Effectiveness of DNS Load Balancing



Small DNS TTLs generally do not improve download times.

Effectiveness of DNS Load Balancing (cont'd)

Parallel-1.0 Download Performance for CDN Server at New and Fixed IP Addresses (Jan. 01)

CDN (technique)	Mean completion time (sec.)		90% completion time (sec.)	
	New IP	Fixed IP	New IP	Fixed IP
Adero (DR-F)	5.40	1.09	9.60	1.60
Akamai (DR-P)	1.15	1.00	3.05	3.00
Digisle (DR-P)	1.31	1.21	2.30	1.70
Fasttide (URDR)	2.10	1.46	4.72	3.25
Speedera (DR-P)	0.72	0.53	1.53	1.01

Small DNS TTLs generally do not improve download times in either average or worst case situations.

CDN Server Use

Number of Distinct IP Addresses Returned to a Client versus the Mean Download Time (MDT) of Parallel-1.0

CDN (technique)	Sept. 2001				Jan. 2001			
	Mean	**Max**	**Total**	**MDT (sec)**	**Mean**	**Max**	**Total**	**MDT (sec)**
Adero (DR-F)	4.6	9	13	1.66	4.8	8	11	1.16
Akamai (DR-P)	5.8	17	65	2.40	8.5	19	103	1.06
Clearway (UR)		–		–	5.6	6	6	1.26
Digisle (DR-P)	2.7	5	24	1.35	3.4	6	24	1.15
Fasttide (URDR)		–		–	8.7	11	23	1.55
Speedera (DR-P)		–		–	10.3	26	83	0.57

Having more CDN servers does not necessarily imply better download performance.

Ongoing Research: CDN Performance for Streaming Media

- Emerging content – streaming media
  - Streaming media account for less than 1% CDN-served objects, but 14-20% CDN-served bytes
- Methodology
  - Similar to the one for static images
- Streaming content examined
  - ASF (Advanced Streaming Format) streamed over HTTP
- Canonical streaming media object
  - Encoding rates: 38/100/300 Kbps
  - Duration: 10 sec. (specified via HTTP headers)

CDN Performance For Streaming Media: Preliminary Results

CDN Performance on Streaming Media: Mean DNS, First Byte, and Last Byte (relative to Target Delay of 10 sec) Delays

CDN	DNS (sec)	First Byte (sec)	Last Byte (sec)		
			38Kbps	**100Kbps**	**300Kbps**
Akamai	0.42	0.83	1.08	1.01	1.18
Digisle	0.22	3.35	3.55	1.09	1.35
Intel	0.00	0.33	0.30	0.49	0.51
Navisite	0.11	0.28	0.45	0.44	0.54
Yahoo	0.13	0.32	0.52	0.50	0.68

Summary

- There is a clear increase in the number and percentage of popular origin sites using CDNs
  - may have decreased subsequently …
- CDNs performed significantly better than origin sites, although caching and HTTP/1.1 options both reduce the performance difference
- Small DNS TTLs generally do not improve client download times in either average or worst case situations
- Our methodology can be extended to test CDN performance for delivering streaming media
  - More streaming media results available in the TM version: http://www.research.att.com/~bala/papers/abcd-tm.ps.gz

Acknowledgments

- Vern Paxson
  - For being involved in earlier stages of the study and help with NIMI
- Reviewers