

**ΟΙΚΟΝΟΜΙΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ**



**ATHENS UNIVERSITY
OF ECONOMICS
AND BUSINESS**

Information-Centric Networks

Section # 5.1: Content Distribution

Instructor: George Xylomenos

Department: Informatics



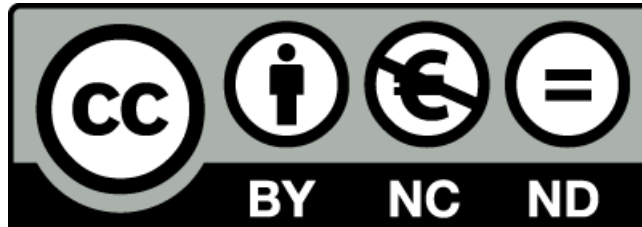
Funding

- These educational materials have been developed as part of the instructors educational tasks.
- The **“Athens University of Economics and Business Open Courses”** project only funded the reformatting of these educational materials.
- The project is being implemented as part of the Operational Program “Instruction and Lifelong Learning” and is co-financed by the European Union (European Social Fund) and national funds.



Licencing

- These educational materials are subject to a Creative Commons License.



Week 5 / Paper 1

- On the use and performance of content distribution networks
 - Balachander Krishnamurthy, Craig Wills, Yin Zhang
 - IMC Workshop, 2001
- Main point
 - CDNs offload work from origin servers
 - How are CDNs used?
 - DNS redirection and URL rewriting
 - How do CDNs perform?
 - DNS lookups may not be worthwhile
 - Note: things have changed a lot since 2001!

Introduction

- Content Distribution Network (CDN)
 - A collection of servers for content delivery
 - May be collocated with origin server or widespread
 - The origin server's is cached/replicated at CDN servers
 - The CDN tries to locate a server “close” to the client
 - Origin servers are aware of CDNs
- CDNs use mostly proprietary algorithms
 - What techniques are employed?
 - How much are CDNs used?
 - What kind of content is offloaded?
 - How can their performance be measured?
 - How do CDNs perform against each other and the origin server?

CDN techniques

- How to direct a client to the right CDN server?
- DNS redirection: select server during DNS resolution
 - The CDN controls the authoritative DNS server
 - Decides based on load and proximity
 - Full-site: origin server is hidden behind the CDN
 - Partial-site: URLs embedded in pages modified for the CDN
 - `www.foo.com/bar.gif->foo.cdn.net/www.foo.com/bar.gif`
- URL rewriting: dynamic rewriting of embedded URLs
 - The returned page has the selected IP addresses replaced
- Mixed mode
 - First URL rewriting with DNS name of CDN server
 - Then DNS redirection depending on load and location

Use of CDNs

- How much are CDNs used?
 - This study is 10 year old, around when CDN use exploded
 - 17-30% of popular sites were found to use CDNs
 - Found either CDN DNS names or CDN DNS servers
 - Most sites used Akamai at that time
 - A custom list of sites was created to reflect other CDNs
- Change characteristics of content
 - Less than 1% of the returned content changed
 - CDNs are used mostly for static content (images)
- Nature of requested content
 - 96-98% of objects were images
 - 40-60% of bytes were images
 - CDN cached images had a 20-30% higher hit rate

Measurement methodology

- Client-side study of image content delivery
 - Response latency and download time
- Content for study: the canonical page
 - A large set of pages was analyzed
 - A page with 18 images of specific sizes was chosen
 - A similar page was found in each CDN
 - Similar pages were found in sites that do not use a CDN
- Study description
 - Record DNS lookup time
 - Use httpperf to download all images so as to prime the CDN
 - Fetch all images (from the CDN) and record the download time
 - Clients from the NIMI measurement infrastructure, no caching
 - Measurements were repeated one year later

Response time results

- Total DNS plus download time results
 - CDNs performed better than origin servers
 - CDNs improved during the second measurement period
 - One CDN had problems with high DNS lookup times
 - URL rewriting CDNs avoid additional DNS lookups
 - DNS redirection CDNs worked best overall
 - But they also had the highest number of servers!
- Download time only
 - Isolates the results of CDN server selection
 - Akamai was the most inconsistent!

DNS load balancing

- DNS load balancing versus CDN server selection
 - CDNs return low TTLs to allow timely selection of servers
 - Is this worth the trouble of the additional lookups?
 - Modified test to remember CDN returned IP address
 - Mixed results for download time
 - Some times the new IP address is not better (may even be worse)
 - Bad results for total response time
 - Even if download time is OK, the extra DNS lookup has a cost
 - Result: the small TTL is not generally a good idea!
 - It may lead to small or no gains in download time
 - Plus it requires an extra DNS lookup
 - Also, it overloads DNS (but CDNs do not really care)
 - But, again, this is a 10 year old study

End of Section # 5.1

Course: Information-Centric Networks, **Section # 5.1: Content Distribution**

Instructor: George Xylomenos, **Department:** Informatics

