Open Issues in Content Distribution

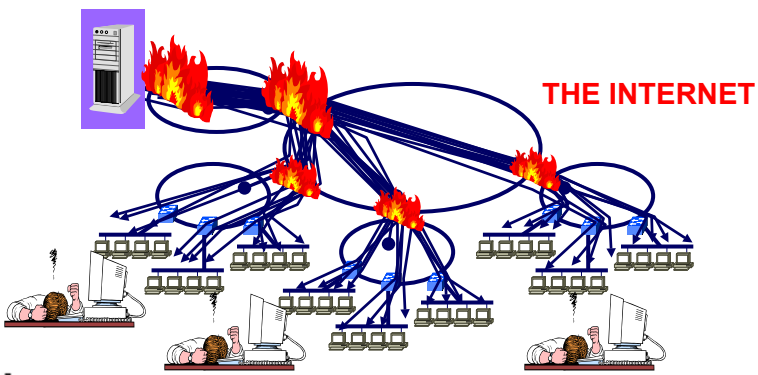
Pablo Rodriguez

Systems and Networking Department
Microsoft Research, Cambridge. UK
pablo@microsoft.com

Microsoft Research

The Problem

- The Internet has been growing very fast, with a growing number of users accessing a growing amount of content
  - Servers and network links are overloaded
  - Users get frustrated



Microsoft Research

Grand Convergence of Computing,

Content Distribution History…

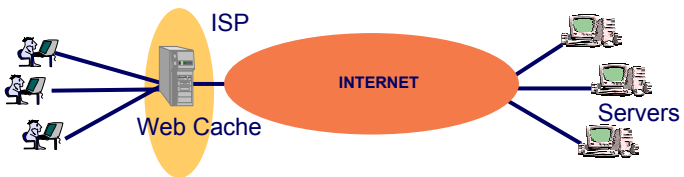
"With 25 years of Internet experience, we've learned exactly one way to deal with the exponential growth: Caching".

(1997, Van Jacobson)

… and he was right, but with a different business model

Content Distribution History…

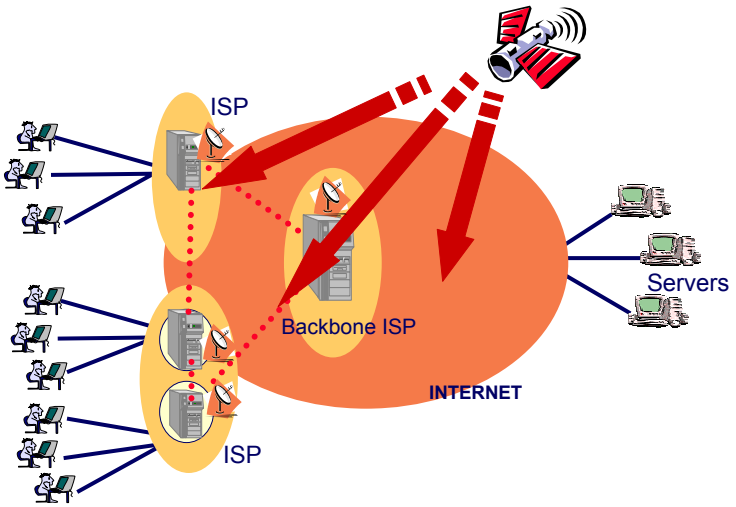
- Web Proxy Caching:



- Caches *save ISPs bandwidth* , reduce clients *latency*, and *avoid flash crowds* and bandwidth usage in the origin servers' access link
- Web caching gives good performance because very often
  - » a single client repeatedly accesses the same document
  - » a nearby client also accesses the same document
- **Cache Hit ratio increases logarithmically with number of users**

Grand Convergence of Computing,

Content Distribution History…



What went wrong with Web Caches?

- Web protocols evolved extensively to accommodate caching, e.g. HTTP 1.1
- However, Web caching was developed with a strong ISP perspective, leaving content providers out of the picture
  - It is the ISP who places a cache and controls it
  - ISPs only interest to use Web caches is to reduce bandwidth
- In the USA: Bandwidth was very cheap.
  - No interest for ISPs in Caching
- In Europe, there were many more Web caches
  - However, ISPs can arbitrarily tune Web caches to deliver stale content
- European Union tried to ban Web caching. Some US content providers started suing ISPs using Web caching…

Grand Convergence of Computing,

Content Provider's Point of View

- Content providers care about
  - User experience latency
  - Content freshness
  - Avoid flash crowds
  - Minimize bandwidth usage in their access link
  - Accurate access statistics
- In an ideal world, all ISPs would use cooperative caches with enough capacity, delivering fresh content, and reporting accurate access statistics
- However, the real world is that many ISPs did not implement caching and the ones that did, abused of it
  - Content providers defeated caches (Pragma: No-cache) and started thinking about building infrastructures to deliver their content…

Content Distribution History…

- Some large content providers decided to use their network of *Mirror Servers*
  - » But content providers prefer to outsource the distribution to a third party…
  - » Plus, it is more cost effective (no need to dimension all systems for the pick)
- *Content Distribution Networks* (CDNs) build an overlay networks of caches to provide fast, cost-effective, and reliable content delivery, while working tightly with content providers

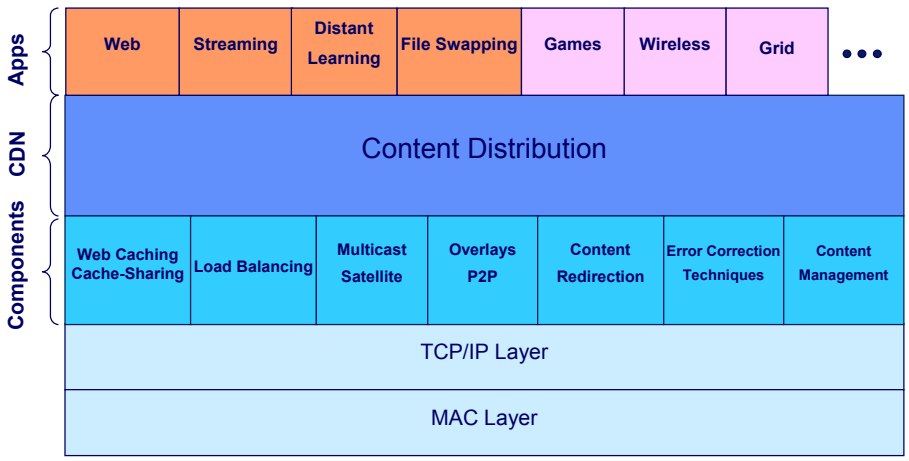
Grand Convergence of Computing,

Why are CDNs important?

- Content Distribution Networks:
  - Provide control over content
  - Bypass bottlenecks to reduce latency and provide more reliable performance
  - Offload servers from flash crowds
  - Provide economy of scale and reduce infrastructure and management cost (sharing)
  - Allow for more sophisticated Web content authoring
  - Eliminate needs to dimension all servers for pick (multiplexing)
  - Shield servers from denial of attacks
  - Provide application-level agreements

- CDNs are used to:
  - Relieve end-user latency for the most important Web sites (e.g. CNN, Yahoo)
  - Minimize impact of flash crowd events (e.g. Olympics, US Open)
  - Provide significant bandwidth savings (e.g. 30-40%)
  - Distribute enterprise content (e.g. remote learning)

Content Distribution Middleware

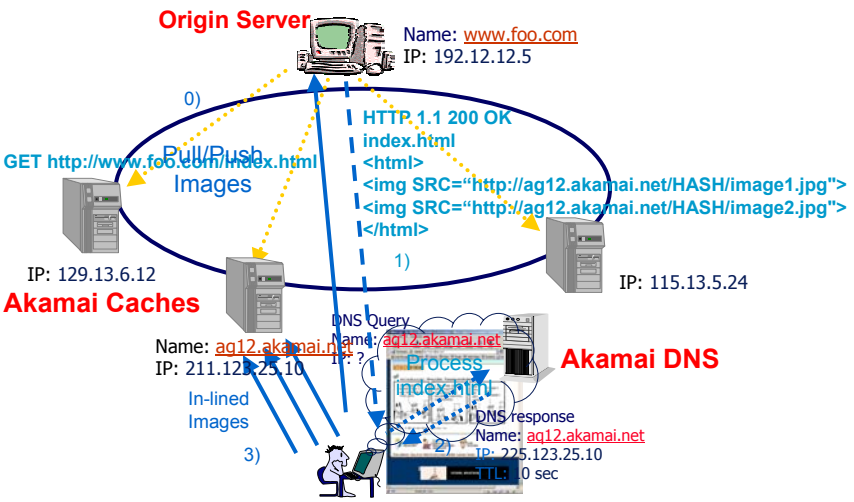


Grand Convergence of Computing,

CDN Case Study: Akamai

- Akamai (AH kuh my) is Hawaiian for intelligent, clever and informally "cool". Founded Apr 99, Boston MA by MIT students
- [Nasdaq: AKAM], had an explosive opening-day gain of 458.4 % on October 29th, 2000
- Akamai can be considered to be the first CDN in the Internet (others also at the time: Sandpiper, Digital Island)
- More than 1250 content providers use their network. 14000 servers in 40 countries
  - Still fewer countries than UN…
- Delivers text/images as well as streaming of stored and life media. $2000 per Mbps/month. $300 for region-specific service

Akamai: How it works?

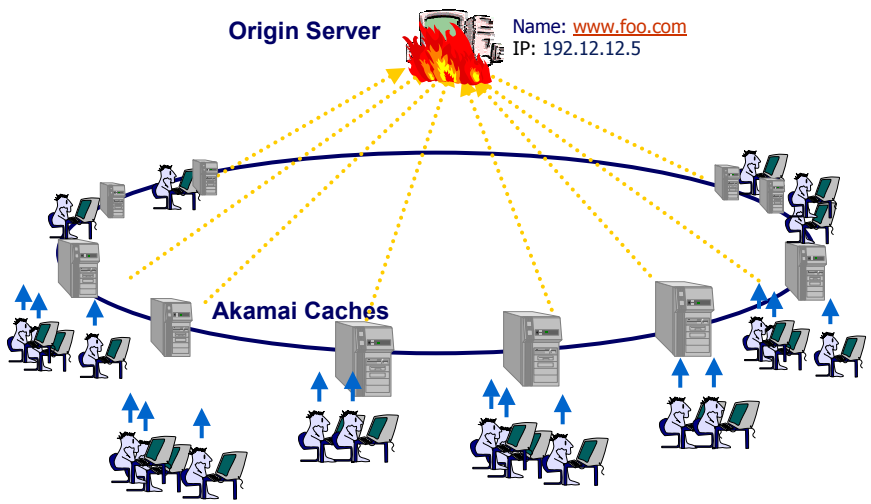


Grand Convergence of Computing,

More Akamai information

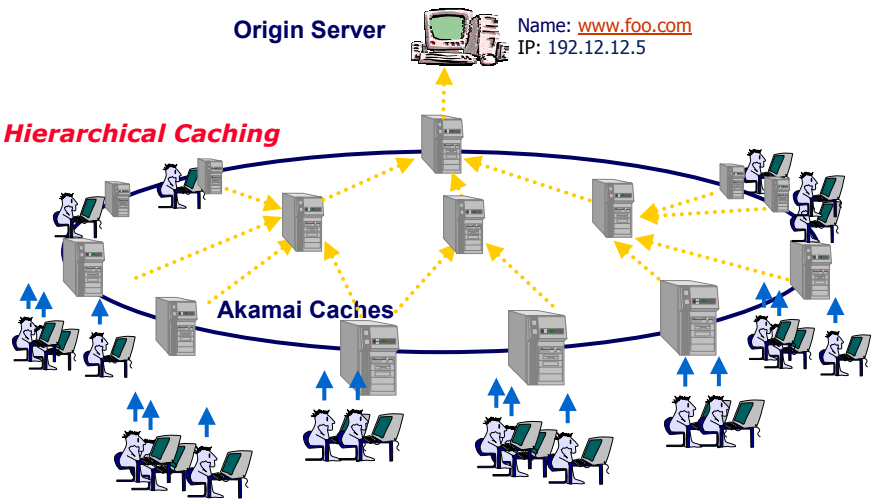
- URL akamaization is becoming obsolete and only supported for legacy reasons
  - Currently most content providers prefer to use DNS CNAME techniques to get all their content served from the Akamai servers
    - » Still content providers need to run their origin servers
- Akamai Evolution:
  - Files/Streaming
  - Secure pages and whole pages
  - Dynamic Page assembly at the edge (ESI)
  - Distributed applications
    - » First step is to replicate read-only databases

Akamai: How to avoid flash crowds?



Grand Convergence of Computing,

Akamai: How to avoid flash crowds?



CDN Challenges….

- Distributing Web content has proven to be an easier problem than expected
  - Bandwidth and servers are becoming cheaper and cheaper
  - P2P is a becoming a serious alternative to using CDNs
- Multimedia content is still waiting for the last-mile
  - Plus no clear business model
- After the bubble burst, there are fewer content providers
  - Content concentration. Fewer chances for statistical multiplexing
- Lots of content tends to be of local interest only (no need to have global presence)
  - Few nodes with well connected links can provide most of the benefit (e.g. data centers in major countries)
- Managing a large number of servers around the world is difficult and expensive

Grand Convergence of Computing,

CND's future….

- The "edge" is shrinking: Most ISPs can reach content providers within only few hops
  - » Bottleneck is moving towards the servers
- Not much sense for world-wide CDNs with huge number of servers
  - Left over from the bubble's era
  - Cost inefficient
- Most likely CDNs will use much fewer servers placed only in well-connected data hosting centers
  - » Provides better multiplexing since sharing happens at fewer points
  - » These servers will enable dynamic resource allocation for efficient infrastructure sharing
- CDNs will still be deployed in enterprises and countries with bandwidth problems
- P2P will take over part of the CDN's business, specially in the enterprise market
- Dynamic content delivery will not require moving the processing power to the edge
  - Processing at the edge does not reduce the processing time

Research: Been there, done that…

- Cache replacement policies
- Inter-Cache sharing protocols
- Redirection Mechanisms
- Content Management…

Grand Convergence of Computing,

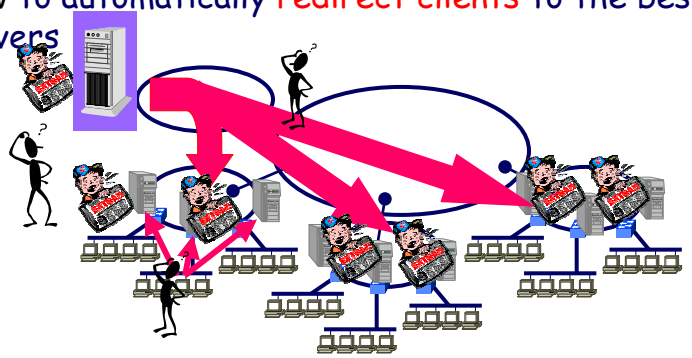
Several Possible Research Areas

- Distributing Dynamic Content
- CDNs for Interactive Applications
- Wireless Content Distribution
- Minimizing maintenance/deployment costs

Microsoft Research

Content Distribution Issues

- How to scale the distribution of content to the network edge
- How to ensure that the content is up-to-date
- How to automatically redirect clients to the best edge servers

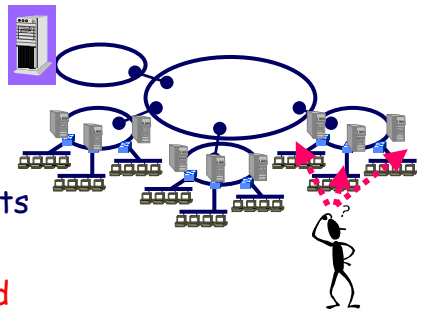


Microsoft Research

Grand Convergence of Computing,

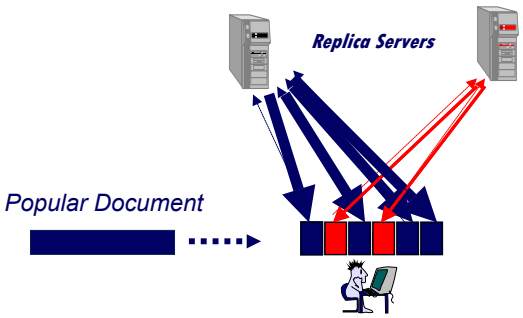
Selecting the Best Server

- The copy selected should ideally have the lowest access time
- Techniques for server selection use
  - Client origination address
  - Geographical proximity
  - Number of hops
  - RTT
  - Server load
  - Application level measurements
- Problems
  - High complexity and overhead
    - » periodic measurements, statistics
  - The selected copy may not always be the best one



Parallel Downloads

- Instead consider parallel-downloads!
  - Speeds-up download times
  - Relaxes server selection
  - Performs load balancing and increases fault-tolerance



Grand Convergence of Computing,

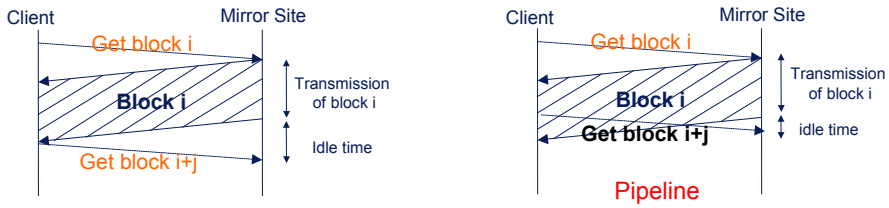
Requirements

- Documents should be large enough
- Replicated content must be bit-by-bit identical
- Clients and servers are able to utilize range requests as specified in HTTP 1.1
- Clients and servers implement persistent connections to avoid TCP slow-start

How to choose the block size?

- The number of blocks should be larger than the number of mirror sites accessed in parallel
- Each block should be small enough to rapidly adapt to changing conditions and ensure that the last block requested from each server terminates at about the same time

- Each block should be sufficient large to reduce the influence of the idle times and reduce the number of negotiations (transmission time>>RTT)



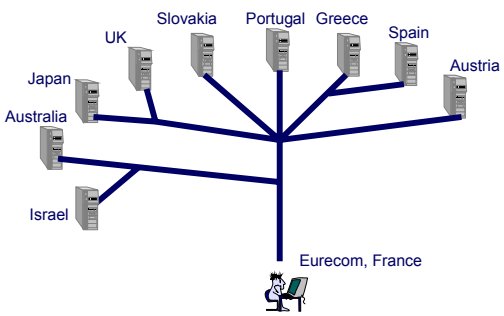
Grand Convergence of Computing,

Cost of Parallel-Access

- Processing overhead at the client for scheduling block assignments
  - There are already several products that do a parallel-access to a single server
  - It is easy to scale the client
- Additional network traffic generated by the block request messages
  - Can be reduced by adaptively increasing the block size
- Higher number of TCP connections opened to the mirror sites
  - For large files, TCP set-up is relatively insignificant
  - TCP connections last for a smaller period of time
  - Drop TCP connections for slow servers
  - Use idle-servers in different time-zones

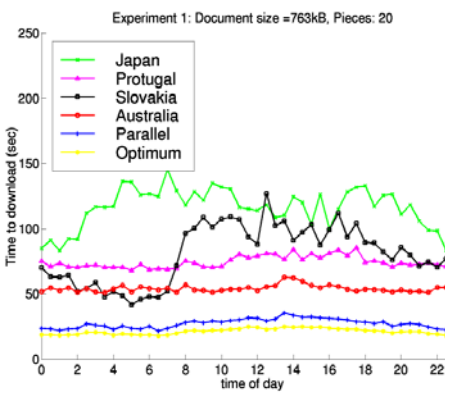
Parallel-Access: Experiment

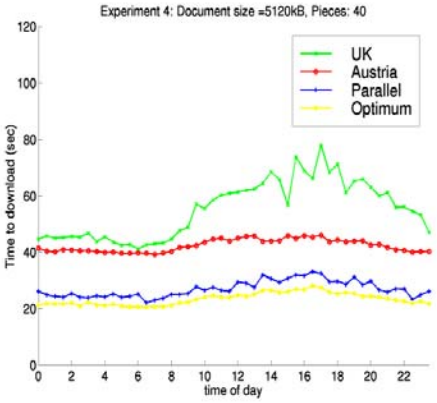
- Client written in Java.
- We used 9 NLANR mirror sites as the mirror servers
- We computed the average delay of the parallel-access system every 15 minutes during a period of 10 days



Grand Convergence of Computing,

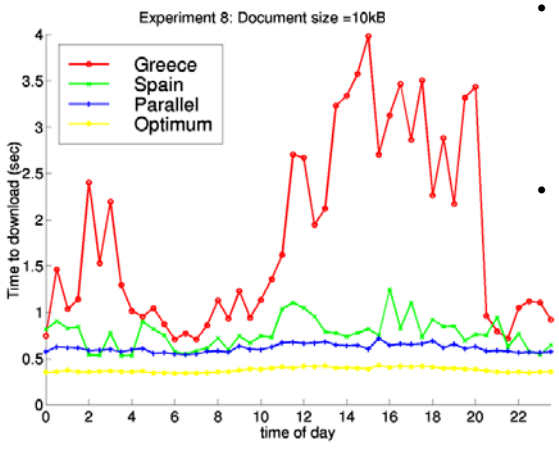
Parallel-access: Performance





- Smoothes out bandwidth variations, increases resilience against congestion and failures, and reduces downloading times

Small Documents

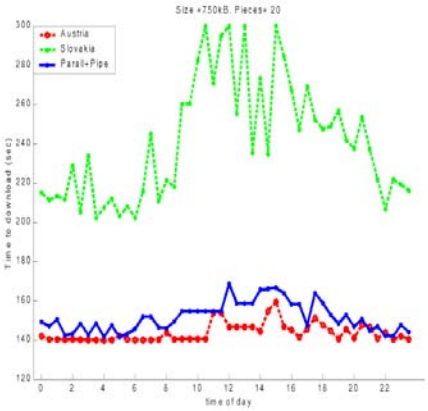


- For small documents, the connection time accounts for most of the download time
- Pipelining is difficult due to the small transmission times

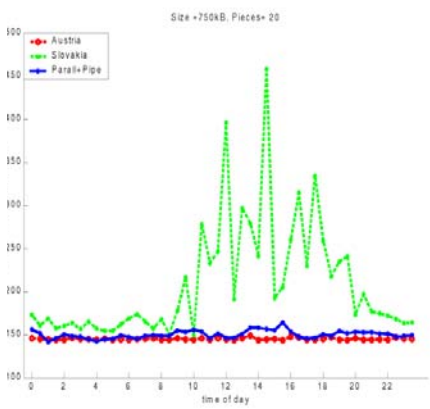
Grand Convergence of Computing,

Parallel-Access on a Modem line

No Pipelining



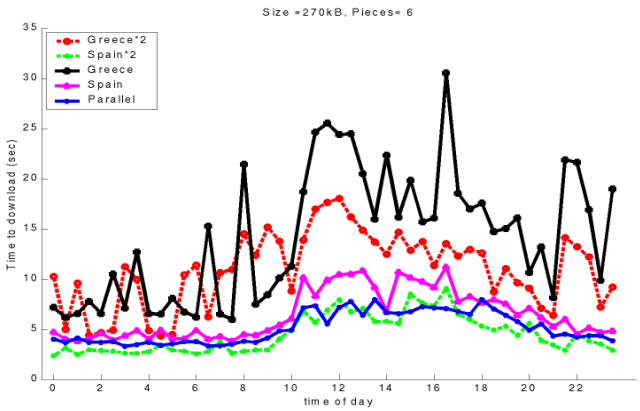
Pipelining



- Parallel-access over a modem line gives as good performance as the fastest server with no server selection, however, idle times are more relevant
- Using pipelining, the parallel-rate is very close to the best possible rate

Microsoft Research

Parallel-Access: Multiple sites vs One site



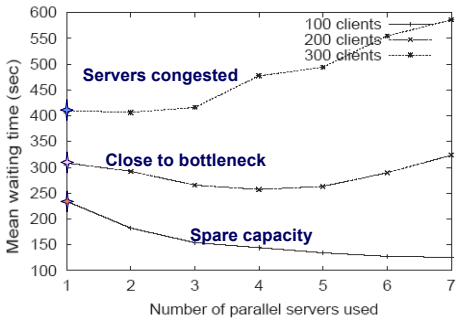
- If the wrong server is chosen, the performance is worse than a parallel-access.
- Using a single server there is no load balancing and there is still the problem of server selection

Microsoft Research

Grand Convergence of Computing,

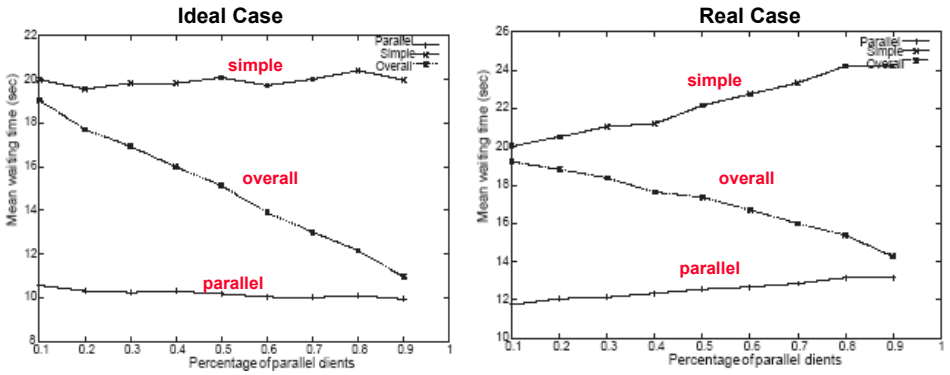
Large Scale Deployment

- What if everybody does parallel-downloads?
  - If *all* clients share the same bottleneck => Speedup is reduced



  - Still,
    - » Client's experience is more uniform
      - across clients and within a single client
    - » Load is automatically distributed among servers
    - » No need for a server selection algorithm

Fairness



100 clients, 10 servers, 100 routers, NS2
(average loaded system)

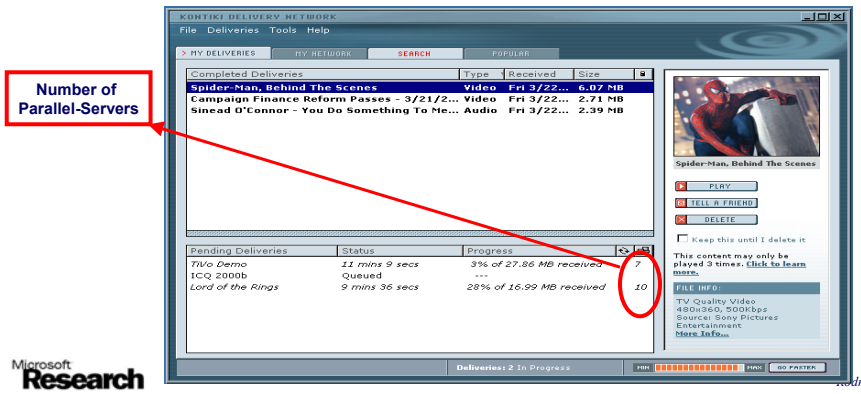
- As more users perform parallel-downloads, the overall download time decreases but the performance of simple-downloads is slightly reduced => Overhead of having many connections and block requests

Microsoft Research

Grand Convergence of Computing,

Peer-to-Peer Parallel-Donwloads

- Parallel-access works very well in peer-to-peer applications
  - Peer-servers performance fluctuates
  - Content tends to be quite large
  - Clients become new sources: Efficient way to scale a parallel-access



Microsoft Research

Future Extensions

- Parallel-downloads under P2P systems
- Applications to streaming audio and video

Microsoft Research

Grand Convergence of Computing,

Further Reading

- *"On the scale and performance of cooperative Web proxy caching".* Alec Wolman, Geoffrey M. Voelker, Nitin Sharma, Neal Cardwell, Anna Karlin, and Henry M. Levy. In *Proceedings of SOSP '99*
- Lee Breslau, Pei Cao, Li Fan, Graham Phillips, and Scott Shenker. *"Web caching and Zipf-like distributions: Evidence and implications".* In Proceedings of the INFOCOM '99 conference, March 1999
- *Pablo Rodriguez, Andreas Kirpal, and Ernst W. Biersack. Parallel-access for mirror sites in the Internet. In Proceedings* of IEEE INFOCOM'2000, March 2000
- *"Distribution of Stored Information in the Web". A tutorial on Web caching. Keith Ross (http://cis.poly.edu/~ross/)*
- *Content Distribution Networks Overview :* http://www.telin.nl/Middleware/cdn/ENindex.htm
- *"Web Caching and Replication".* Rabinovich and Spatscheck. Addison Wesley.

Conclusions

- Content Distribution is an important part of today's Internet
- Content Distribution needs to take a step forward to accommodate new applications and bottlenecks
- Interactivity, distributed applications, reduced maintenance/deployment costs
- Wireless networks bring a new set of problems that can be overcome with intelligent CDN techniques

Grand Convergence of Computing,