Web Caching: a Memory-based Architecture

Internal Conference of Computer Architecture
ICCA 2003
Dep. de Informática - Universidade do Minho

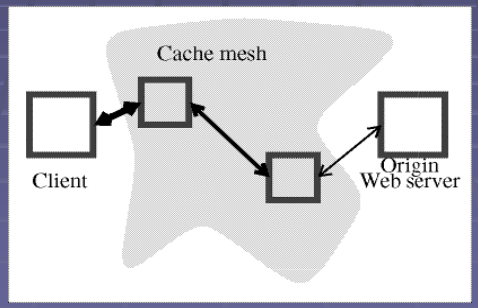
David Sora
david@ipb.pt

Summary

- Overview of Web Caching
  - Functions of a Web Cache
  - Web Caching vs. File System Caching
- Web performance
  - DFS (Distributed File System)
  - Hierarchical Caching
  - Distributed Caching
  - Hierarchical vs. Distributed
  - Performance issues
- AFPA (Adaptive Fast Path Architecture)
- Intel Itanium Architecture
- Conclusions

Overview of Web Caching

- Functions of a Web Cache



David Sora
david@ipb.pt

Overview of Web Caching

- Web Caching vs. File System Caching
  - File size
  - Temporal locality
    - Least Recently Used replacement algorithms (LRU)
    - Least Frequently Used replacement algorithms (LFU)

David Sora
david@ipb.pt

Web Performance

- Distributed File System (DFS)?
  - Copies of files may be cached in one or more servers.
  - Works fine if:
    - Number of Web servers is not too large;
    - Data does not change very frequently;

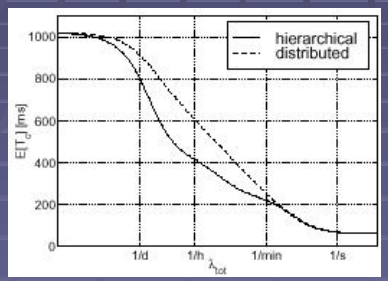
Web Performance

- Caching Architectures
  - Hierarchical Caching
    - Origin server
    - National cache
    - Regional cache
    - Institutional cache
    - Client cache

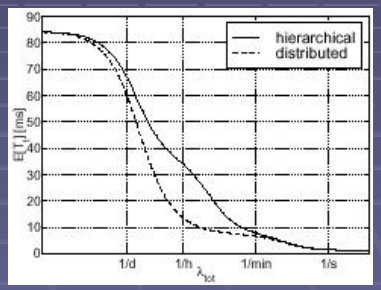
    Cache Levels
  - Distributed Caching
    - Institutional cache

Web Performance

- Hierarchical vs. Distributed



Expected connection time



Expected transmission time

David Sora
david@ipb.pt

Web Performance

- Performance issues
  - Data Copies and Reads
  - Event Notification
    - Single process event driven (SPED)
    - Multiple process / thread (MP)
  - Communication Code Path
    - Reducing the code path (between the Web Server and TCP/IP stack) by system calls and redundant socket layer code elimination

David Sora
david@ipb.pt

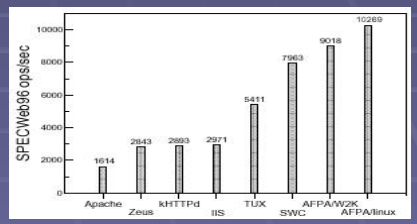
Adaptive Fast Path Architecture (AFPA)

- Kernel-mode platform for high-performance network servers
  - Support for a variety of application protocols
  - Direct integration with the TCP/IP protocol stack
  - Kernel-managed, zero copy cache

Web Server Characteristics

	architecture	cache	0 copy	direct TCP
Apache	MP/user	filesystem	no	no
Zeus	SPED/user	filesystem	no	no
IIS	SPED/user	filesytem	yes	no
kHTTPd	SPED/kernel	filesytem	no	no
TUX	SPED/kernel	memory	yes	no
SWC	SPED/kernel	fs or mem	yes	yes
AFPA	Softint/kernel	fs or mem	Yes	yes

SPECweb96 workload



David Sora
david@ipb.pt

Intel Itanium Architecture (IA 64-bit)

- Predication
- Speculation
- Increased Number of Registers
- Parallelism

David Sora
david@ipb.pt

Conclusions

- Why Web Caching?
  - "(…) a Web cache could eliminate at least 30% of the Web traffic (…)"
- Improved solutions:
  - Adaptive Fast Path Architecture (IBM)
    - Doubles capacity for serving static content
  - Intel Itanium Architecture (Intel)
    - Enable larger and more powerful cache server software

David Sora
david@ipb.pt

Conclusions

- **Future?**

Deliver performance improvements for dynamic content.