Web Servers Overview

Chris Chin, Gregory Seidman, Denise Tso

February 21, 2001

Relating to Other Chapters

- *Dynamic content* comes from web servers
- *Proxy caches* sit between users and web servers
- Web sites are *customized*
- Web *search engines* crawl web servers

HTTP Server Basics

- HTTP is an application-level protocol on top of TCP
- Simplest version: respond to HTTP requests with files on disk
  1. accept connection and read request
  2. translate requested URL into file path
  3. read file from disk and write to network
  4. close connection
- More complicated versions involve
  - generating responses dynamically
  - caching responses in memory

The First Problem: Too Many Requests!

Several possible bottlenecks:

- Network bandwidth
- Disk bandwidth (and latency)
- Memory capacity
- Processing time

Solutions (1/2): Clustering

"A cluster, as I understand it, is a series of machines with the ability to crash in parallel." —A Microsoft Employee

- Several servers appear to the outside world as a single machine
- Advantages include
  - more disk bandwidth
  - more memory capacity
  - more processing power
  - sometimes support failover
- Disadvantages include
  - may not survive failed elements
  - load may not be balanced
  - difficult to maintain a consistent overall state

Solutions (2/2): Load Balancing

- Fair and efficient allocation of workload to clustered elements
- Hard problem
- Depends on standards for inter-node communication

The Second Problem: Security

- Some information is sensitive
- Only designated users should have access
- Listening on the network should be difficult or impossible

Solutions: Authorization and Encryption

- HTTPS and SSL
- Password protection