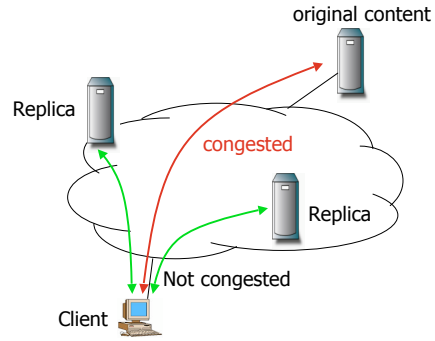
Outline for today

- Oceanstore: An architecture for Global-Scale Persistent Storage – University of California, Berkeley. ASPLOS 2000
- Chord
- Content Distribution Network

Content Distribution Networks (slides courtesy Girish Borkar: Udel)



Persistent store

E.g. files (traditional operating systems), persistent objects (in a object based system)

- Applications operate on objects in persistent store
  - Powerpoint operates on a persistent .ppt file, mutating its contents
  - Palm calendar operates on my calendar which is replicated in myYahoo, Palm Desktop and the Pilot itself
- Storage is cheap but maintenance is not
  ~ 4 $/GB

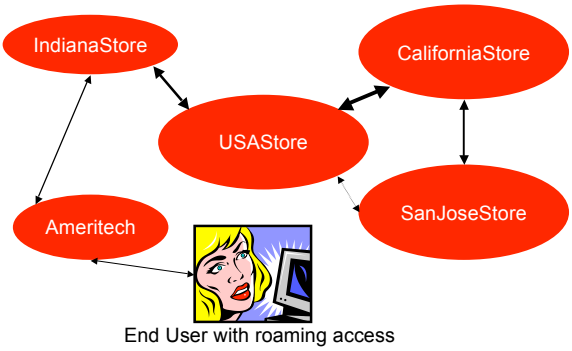
Global Persistent Store

- Persistent store is fundamental for future ubiquitous computing because it allows "devices" to operate transparently, consistently and reliably on data.
- Transparent: Permits behavior to be independent of the device themselves
- Consistently: Allows users to safely access the same information from many different devices simultaneously.
- Reliably: Devices can be rebooted or replaced without losing vital configuration information

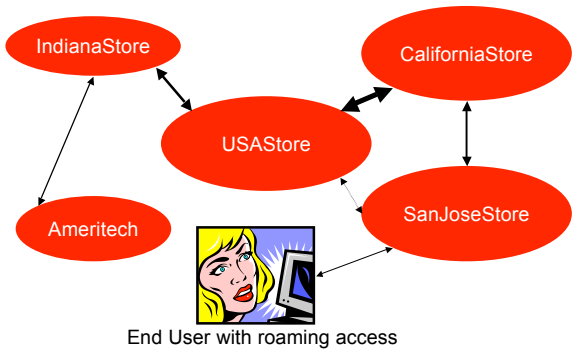
Persistent store on a wide-scale

- 10 billion users, 10,000 files per user = 100 trillion files!!
- Information:
  - should be separated from location. To achieve uniform and highly-available access to information, servers must be geographically distributed, but exploit caching close to clients for performance
  - must be secure
  - must be durable
  - must be consistent

Oceanstore system model: Data Utility



Oceanstore system model: Data Utility



End User with roaming access

Oceanstore Goals

- Untrusted infrastructure (utility model – telephone)
  - Only clients can be trusted
  - Servers can crash, or leak information to third parties
  - Most of the servers are working correctly most of the time
  - Class of trusted servers that can carry out protocols on the clients behalf (financially liable for integrity of data)
- Nomadic Data Access
  - Data can be cached anywhere, anytime (promiscuous caching)
  - Continuous introspective monitoring to locate data close to the user

Oceanstore Persistent Object

- Named by a globally unique id (GUID)
- Such GUIDs are hard to use. If you are expecting 10 trillion files, your GUID will have to be a long (say 128 bit) ID rather than a simple name
  - passwd vs 12agfs237dfdfhj459uxzozfk459ldfnhgga
- self-certifying names
  1. secureHash(/id=surendar,ou=uga,key=<SecureKey>/etc/passwd) -> uniqueId
  2. Map uniqueId->GUID
  - Users would use symbolic links for easy usage
    - /etc/passwd -> uniqueId

SecureHash

- Pros:
  - The self-certifying name specifies my access rights
- Cons:
  - If I lose the key, the data is lost
    - Key management issues
      - Keys can be upgraded
      - Keys can be revoked
  - How do we share data?

Access Control

- All read-shared-users share an encryption key
  - Revocation:
    - Data should be deleted from all replicas
    - Data should be re-encrypted
    - New keys should be distributed
    - Clients can still access old data till it is deleted in all replicas
- All writes are signed
  - Validity checked by Access Control Lists (ACLs)
  - If A says trust B, B says trust C, C says trust D, what can you infer about A ? D

Oceanstore Persistent Object

- Objects are replicated on multiple servers. Replicated objects are not tied to particular servers i.e. floating replicas
- Replicas located by a probabilistic algorithm first before using a deterministic algorithm
- Data can be active or archival.
  - Archival data is read-only and spread over multiple servers – deep archival storage

Updates

- Objects are modified through updates (data is never overwritten) i.e. versioning system
- Application level conflict resolution
- Updates consist of a predicate and value pair. If a predicate evaluates to true, the corresponding value is applied.
  1. <room 453 free?>, <reserve room>
  2. <room 527 free?>, <reserve room>
  3. <else> <go to Jittery Joes>
- This is similar to Bayou which we will explore later in the semester

Introspection

- Oceanstore uses introspection to monitor system behavior
- Use this information for cluster recognition
- Use this information for replica management