Securing Wide-area Storage in WheelFS

Xavid Pretzer
Supervised by Jeremy Stribling and Professer M. Frans Kaashoek

Overview

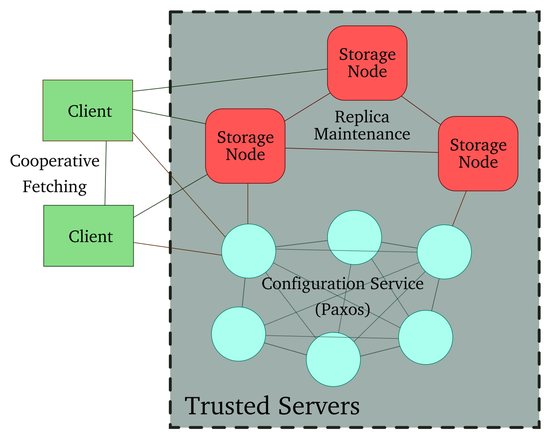
Motivation

- Many advantages to wide-area applications
  - Easy to scale up
  - Put data close to users
  - Handle site-specific faults like natural disasters
- Wide-area storage requires tradeoffs
  - High fault-tolerance prevents strong consistency
  - High availability leads to higher write latency
- The "right answer" depends on the application
- Many applications implement custom storage to fine-tune these tradeoffs

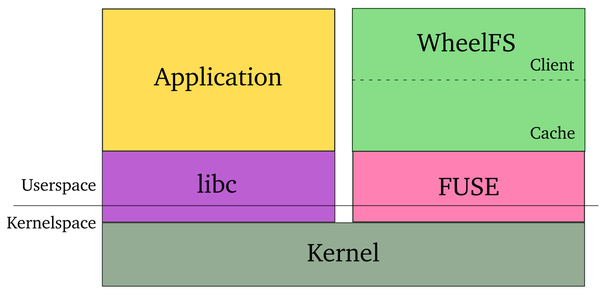
WheelFS

- General, flexible wide-area storage
- Familiar POSIX interface
  - Easy to program
  - Can adapt existing (non-distributed) code
- Configured using 'Semantic Cues'
  - Special pathname components
  - Small number of useful parameters
  - Configure tradeoffs per-file or directory
- Multiple mutually-untrusting applications can securely use WheelFS on the public Internet

System Layout



Client Structure



Semantic Cues

Durability:

- **.RepLevel**=N: keep N backup replicas
- **.SyncLevel**=N: writes succeed when N replicas have the change

Placement:

- **.Site**=X: store at the indicated site
- **.KeepTogether**: store new files with their directory
- **.RepSites**=N: replicas must be in at least N sites

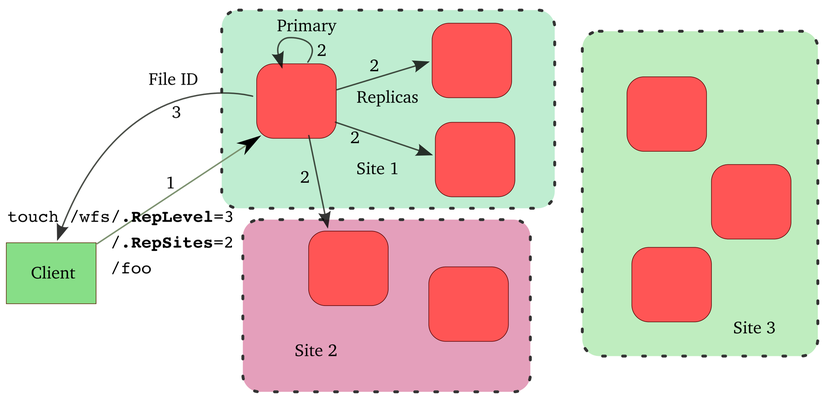
Consistency:

- **.EventualConsistency**: use possibly stale versions if necessary
- **.MaxTime**=T: fail over or return an error after T ms

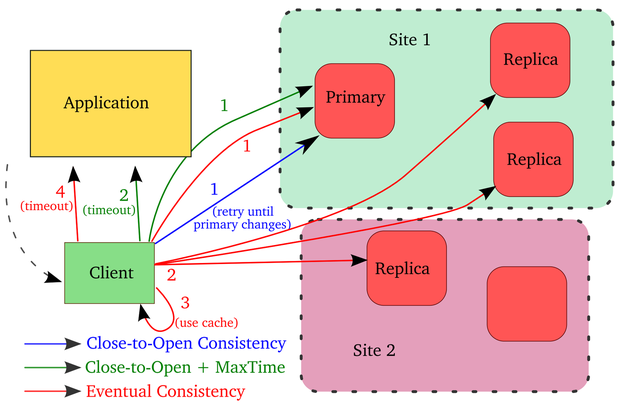
Large reads:

- **.HotSpot**: fetch blocks from other clients' caches
- **.WholeFile**: prefetch later blocks when first block is read

File Creation



Failover for Different Consistency Configurations



Making Apache Caching Proxy into a Cooperative Web Cace

```
CacheRoot /wfs
          /.EventualConsistency
          /.MaxTime=1000
          /.HotSpot
          /cache
```

Security

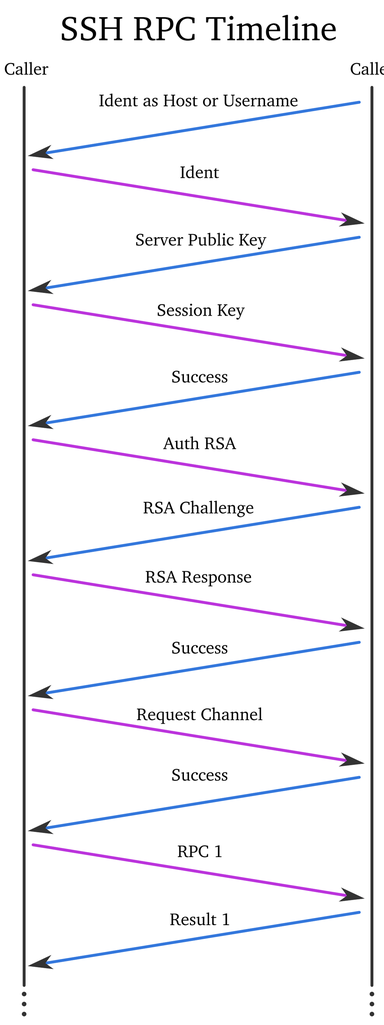
Security Model

- Servers (Storage and Configuration Nodes) trusted
- Clients may be untrusted
  - Privledge separation
  - Reduce impact of app compromise/stolen laptop
- Network untrusted

Security Measures

- SSH connections for encrypted communication
  - Widely available and well tested
  - SSH Agent forwarding (not implemented)
- RSA public key authentication
  - Already used at PlanetLab
  - Server public keys distributed out-of-band
    - Should be maintained by configuration service
  - User public keys stored in the file system
- SHA-256 checksums validate client-to-client data
- Servers and clients both enforce ACLs
  - Clients need to enforce ACLs for .Hotspot
    - Can't stop a hacked client from sharing data

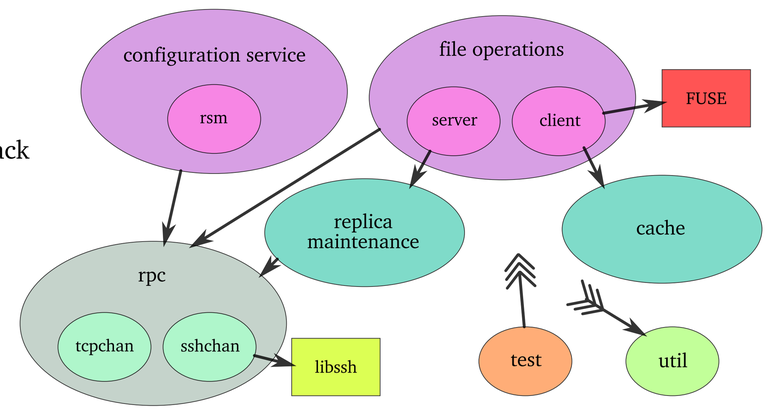
SSH RPC Timeline



Code

- 19,000 lines of C++
- 3800 more for RPC library
  - Uses Vivaldi network coordinates to track network distance between nodes
- Uses pthreads and STL
- Client uses FUSE's low level interface
- SSH channels use openssl and libssh

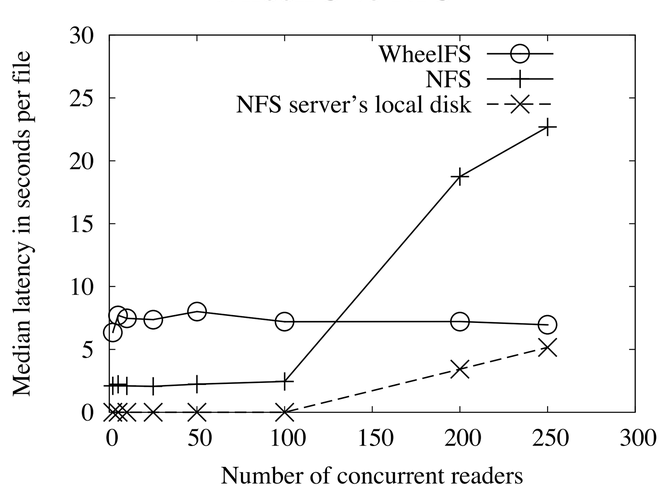
Code Organization



Results

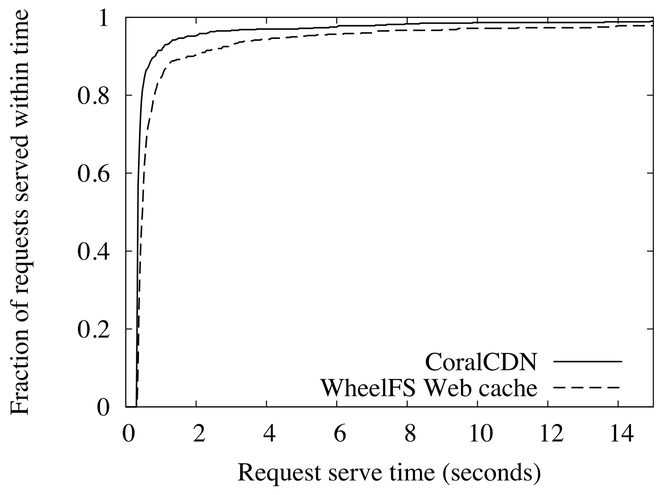
- Useful distributed applications can be easily created from local applications and small code and configuration changes
- Applications on WheelFS have comparable performace and scalability to apps with custom storage layers
- Eventual consistency allows continued high performance even with failures

WheelFS vs NFS

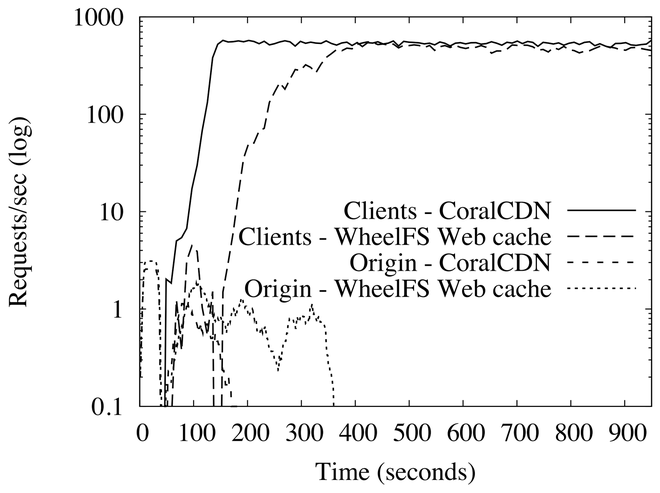


Time to each read 1 MB

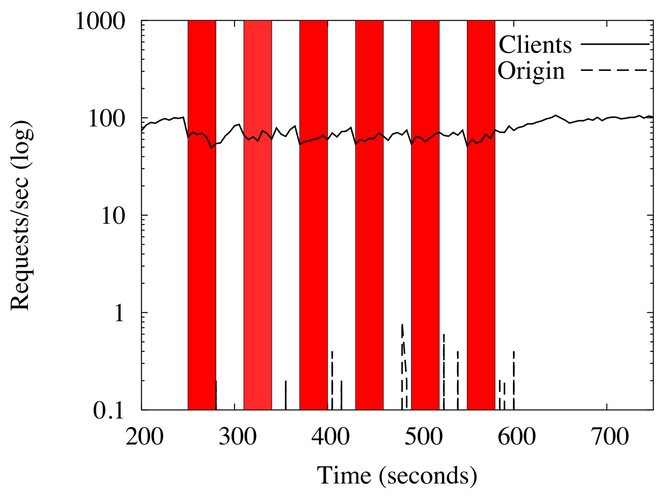
WheelFS Web Cache vs CoralCDN



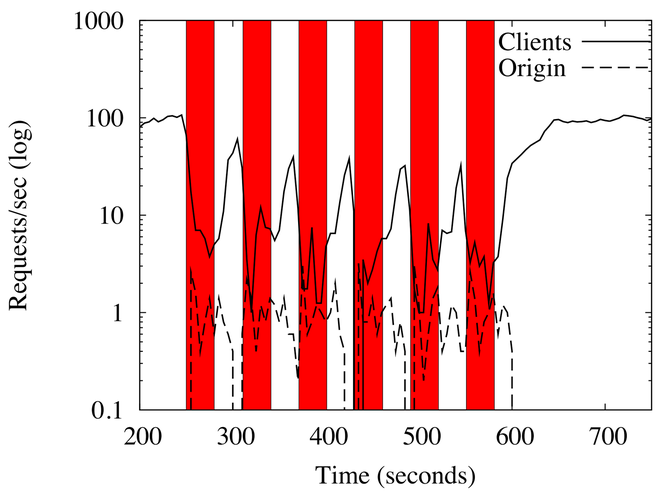
CDF for request latency



Service rate and origin web server load

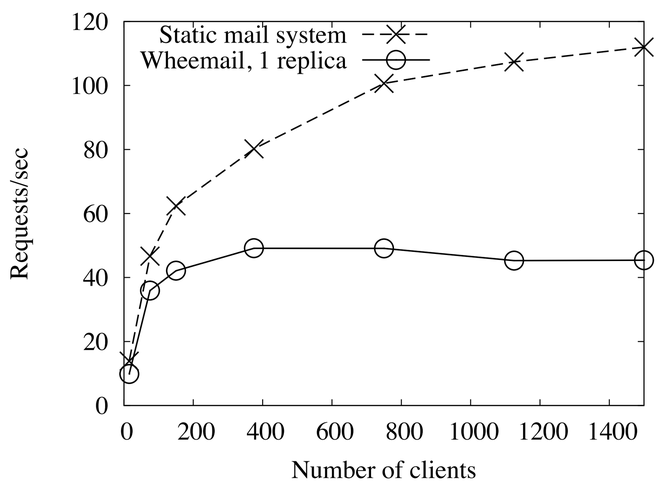


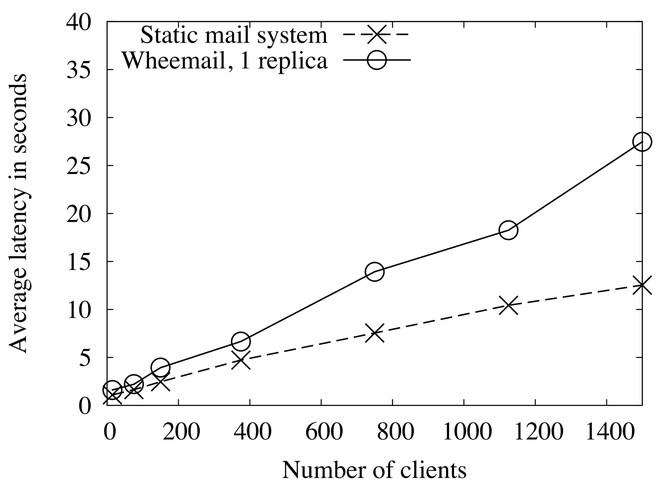
Failures (red) under Eventual Consistency



Failures (red) without Eventual Consistency

WheelFS Replicated Mail vs Static Mail Servers





References

Dabek, Frank, Russ Cox, M. Frans Kaashoek and Robert Morris. "Vivaldi: A Decentralized Network Coordinate System." *Proceedings of the 2004 SIGCOMM*.

Lamport, Leslie. "The Part-time Parliament". *ACM Transactions on Computer Systems*, Vol 16, Issue 2 (May, 1998).

Rivest, Ron, Adi Shamir and Leonard Adleman. "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems". *Communications of the ACM*, Vol 21, Issue 2 (1978).

"Secure Hash Standard (SHS)". Federal Information Processing Standards Publication. Number 180-3 (October, 2008).

Stribling, Jeremy, Yair Sovran, Irene Zhang, Xavid Pretzer, Jinyang Li, M. Frans Kaashoek and Robert Morris. "Flexible, Wide-Area Storage for Distributed Systems with WheelFS". *Proceedings of the 6th USENIX Symposium on Networked Systems Design and Implementation*. (April, 2009).

Ylönen, Tatu. "SSH: Secure Login Connections over the Internet". *Proceedings of the Sixth USENIX UNIX Security Symposium*. (1996)