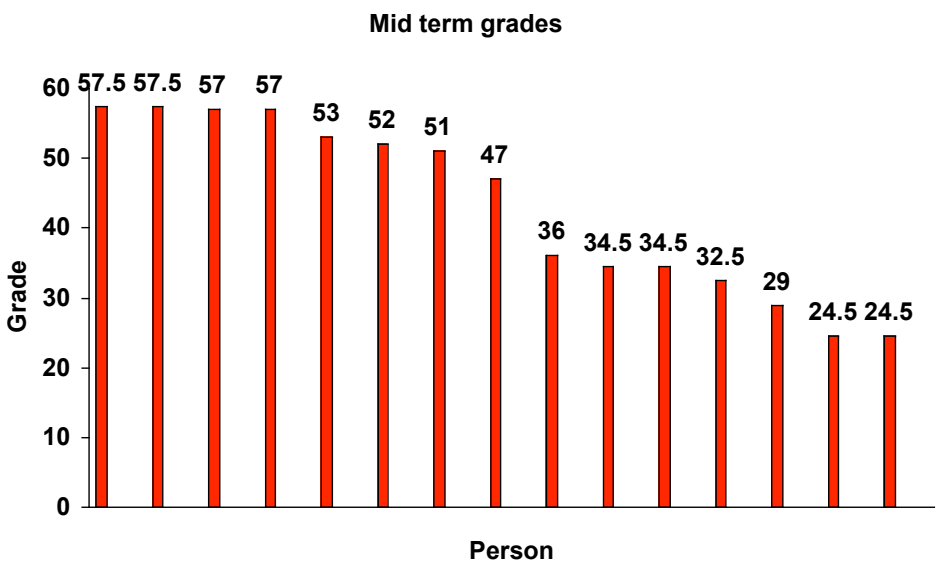


Outline

- Extensibility, Safety and Performance in the SPIN Operating System
  - Brian Bershad, Stefan Savage, Przemyslaw Pardyak, Emin Gun Sirer, Marc E. Fiuczynski, David Becker, Craig Chambers, Susan Eggers
- MSDOS: Extensibility and Performance
- Mach: Extensibility and Safety
- UNIX: Safety and Performance
- Goal: SPIN should have all three

Extensibility

- Applications can dynamically extend system to provide specialized services
- Put extension code in the kernel
  - Communication cost is cheap
- SPIN implements minimal services: Processor execution state, MMU, IO/DMA, Dynamic linker
- Compare with
  - Micro-kernels: Cost to cross protection boundaries
  - Library based: Offers minimal protection boundaries

Safety

- Kernel is protected from actions of extension
- Use language protection features
  - Static safety
- Modula3
  - Memory safe
  - Interfaces for hiding resources
  - Cheap capabilities
  - Restrict access to interfaces at dynamic link-time

Performance

- Extensibility and safety have low cost
- Extensions provide specialized service
  - Customized for the specific task with no extraneous code
- Extensions close to kernel service
  - Invocations cheap
  - Low latency response to interrupts

Outline

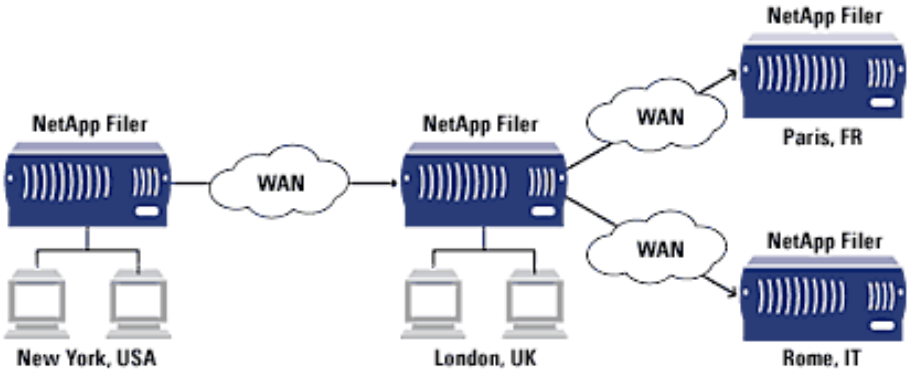
- SnapMirror: File System based asynchronous Mirroring for disaster recovery
- Trends:
  - Persistent and reliable data is crucial for businesses
  - Disks are getting cheaper and bigger, backup technologies are not keeping up
  - RAID to guard against disk failures
    - Hybrid levels (level 50) can provide redundancy and performance
  - Disaster recovery
    - Create off-site online backups to guard against disasters

Challenges

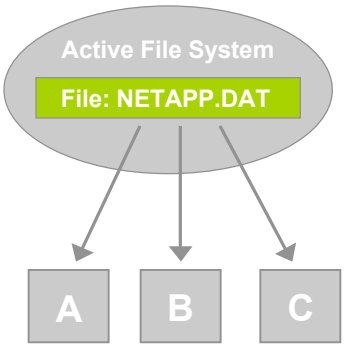
- Backup restore from tapes are cheap but slow
  - Tapes can achieve around 60 GB/hour for restore
  - Terabyte data stores can take a long time to restore
- Online remote backup
  - Expensive (network bandwidth requirements)
  - Performance slow because transaction cannot complete till WAN update finishes
- Asynchronous backups
  - Backup at regular intervals
  - If backup goes to multiple devices, then the event ordering can create inconsistent backups
- We want cheaper, faster restore capable mechanism

Remote mirrors for disaster recovery

- Courtesy: NetApp

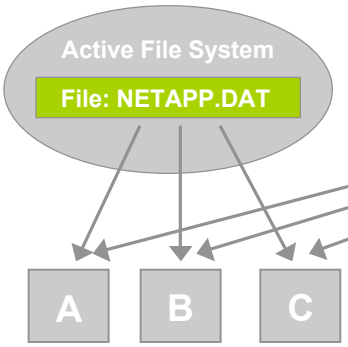


Snapshot Internals (1)



- Data actually resided in block C on disk

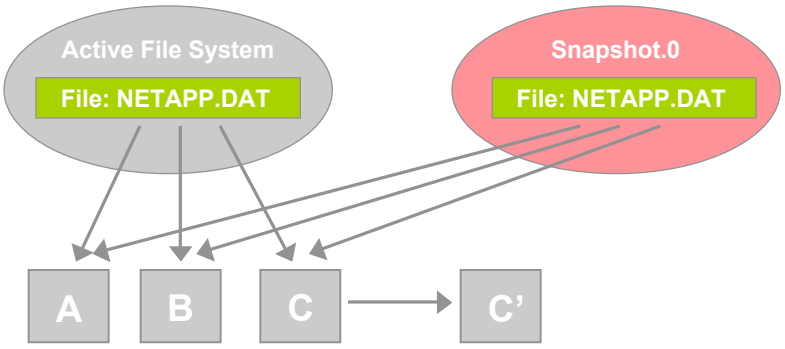
Snapshot Internals (2)



- Makes copy of root inode
- Updates Block Map File

- Data actually resided in block C on disk

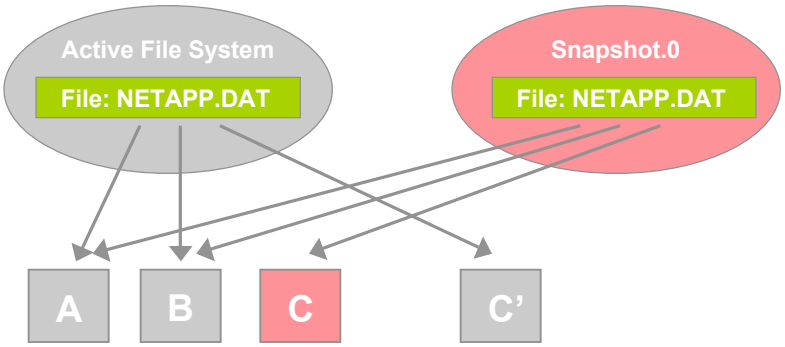
Snapshot Internals (3)



- WAFL writes modified data block to new location on disk (C')

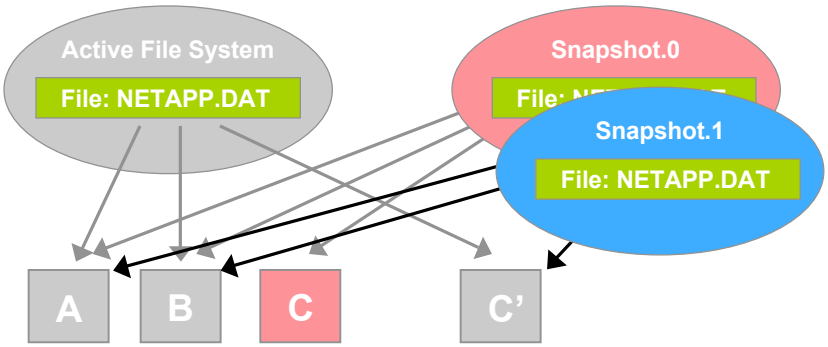
- Client modifies data at end of file
- Data actually resided in block C on disk

Snapshot Internals (4)



- Active file system version of NETAPP.DAT is now composed of disk blocks A, B & C'.
  - Snapshot.0 file system version of NETAPP.DAT is still composed of blocks A, B & C

Snapshot Internals (5)



- Snapshot.1 file system version of NETAPP.DAT is composed of blocks A, B & C'

SnapMirror

- Can use this mechanism to mirror data across WAN
- Can reduce data storage requirements by not backing up deleted/updated data
- Identifying dirty blocks are easier than logical, file system aware mechanisms