IBM Almaden Research Center

CIS: Content Immutable Storage for Trustworthy Electronic Record Keeping

Lan Huang IBM Almaden

Trustworthy Records as Asset

- Electronic records explosion, paperless trend

  Increase by 64% per year to almost 2EB in 2006
- Regulations

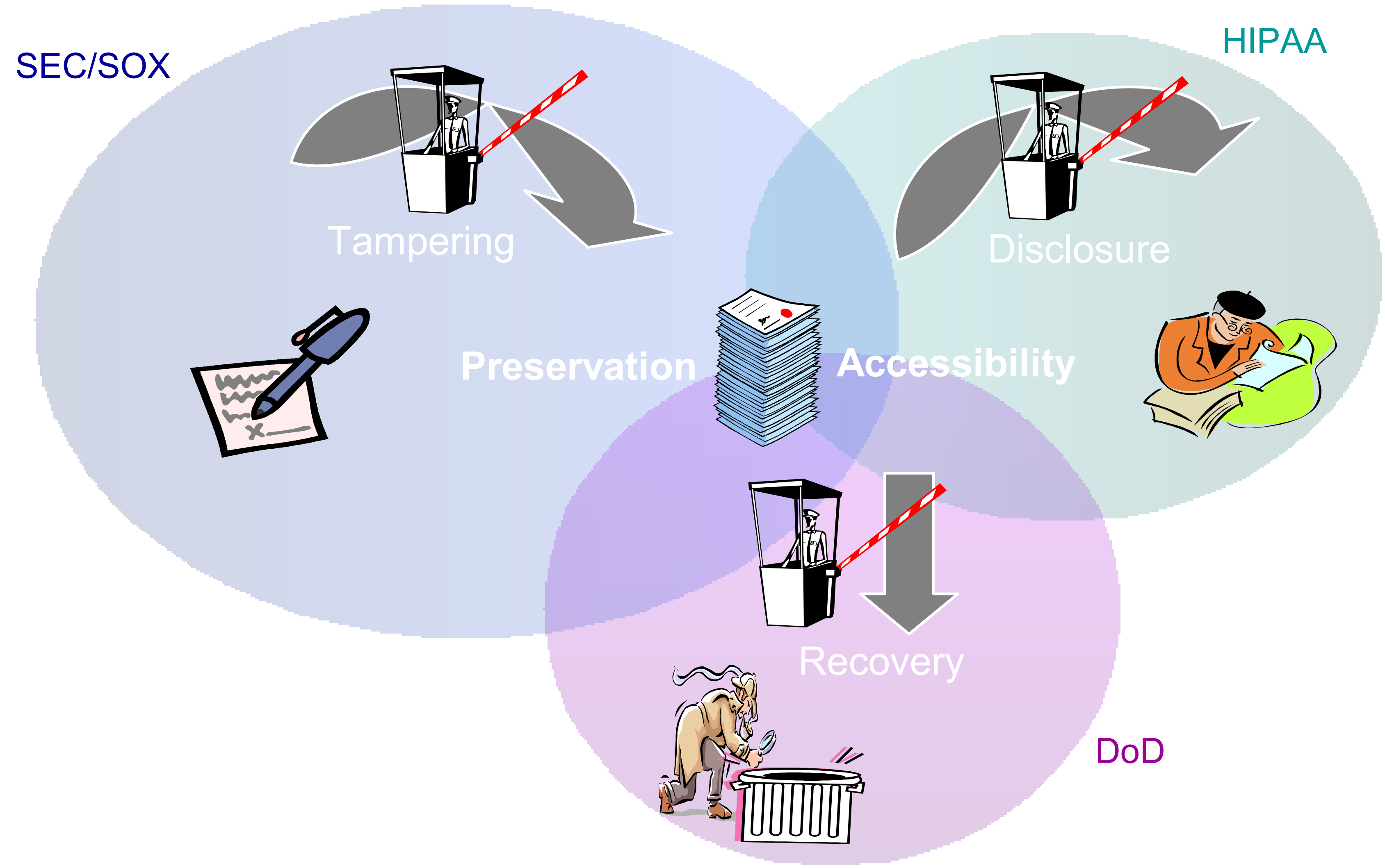
  HIPPA

  Sarbanes-Oxley

  SEC Rule 17a-4

  DoD

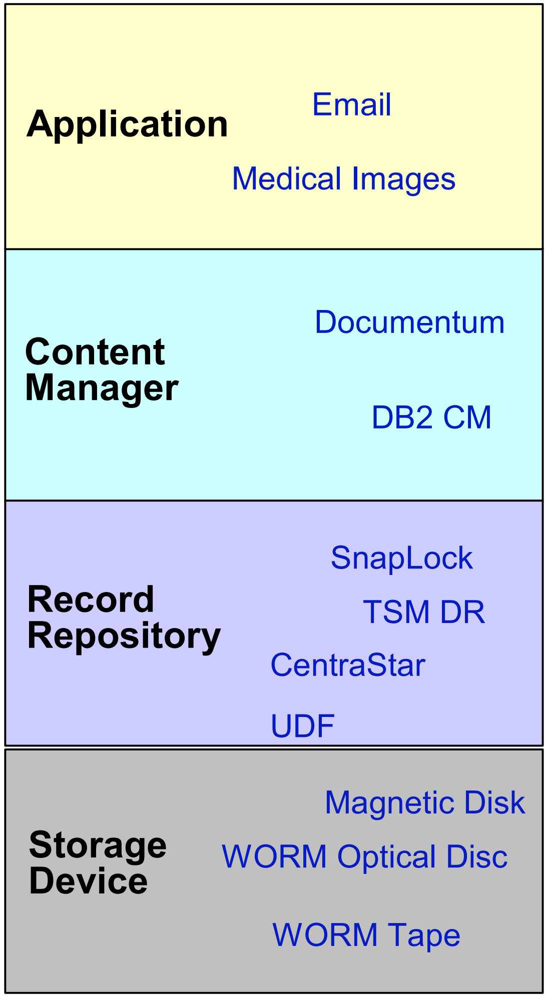
Storage Issues in Record Retention



What Must the Records be Protected Against?

- Accidents, user errors
- Software problems, bugs
- Intentional, malicious attacks since the stakes can be very high for critical records
  - disgruntled employees
  - virus, hacker
  - company insiders
  - conspiracy involving technology experts
- Requires stronger protection than for “security” due to likelihood of inside job

A Typical Record Management Stack



Application

Email

Medical Images

Content Manager

Documentum

DB2 CM

Record Repository

SnapLock

TSM DR

CentraStar

UDF

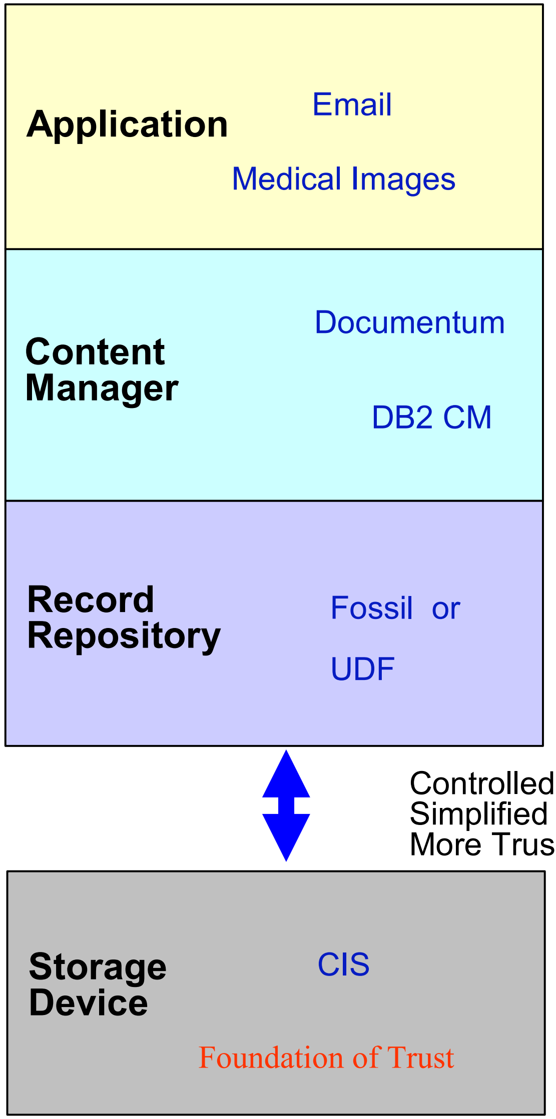
Storage Device

Magnetic Disk

WORM Optical Disc

WORM Tape

A Trustworthy Record Management Architecture on CIS

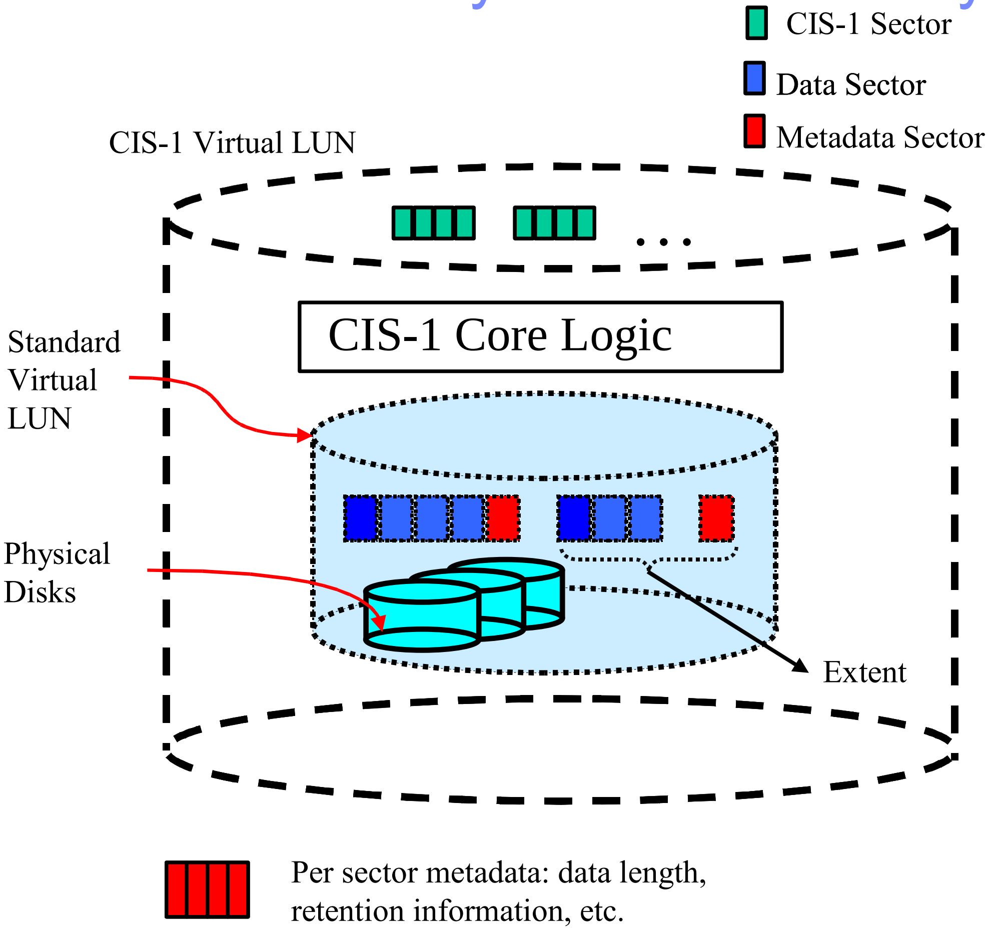


Storage Requirements for Trustworthy Record Keeping

- Secure immutability
  - Protect against even insider attacks
- Efficient index support
  - UDF and other WORM friendly index structures
  - Small append support
- Term-retention and disposition : term-WORM
  - SEC Rule 17a-4
  - DoD
- Low cost and reliable

CIS Prototype Overview: Add-on Modular Layer of Immutability

- Secure immutability of data
  - over-write protection implemented in RAID controller – small trusted computing base
  - disk removal interlock – complete mediation of requests
- Online accessibility of records
  - small-write capability to efficiently support index mechanisms
- Low total cost of ownership



Prototype:
IBM ServeRAID 7t
SATA RAID5 disk array
iSCSI protocol

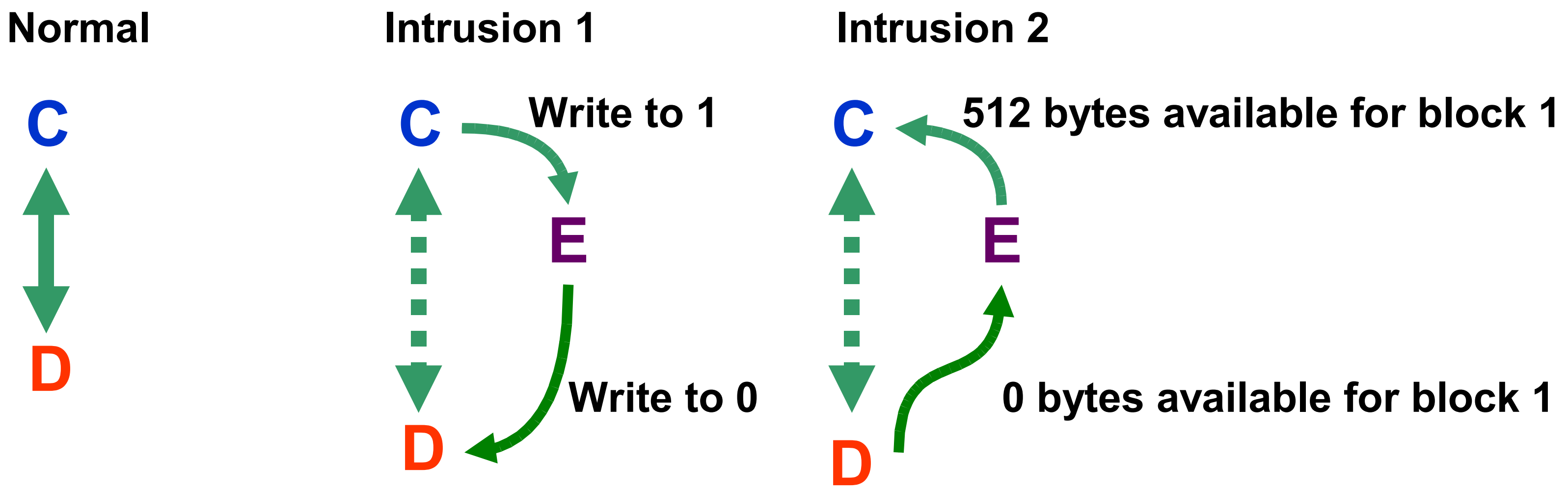
Where to implement CIS-1 core logic?

- Application software
- Network router
- Virtualization software

- Storage controller
  - RAID protection plus programmability
- Hard disk
  - OSD? Cost and infrastructure support barrier

CIS-1 Threat Model

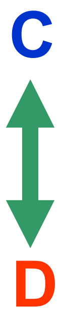
C: Controller D: Disk E : Eve


Normal
C
D
Intrusion 1
C
Write to 1
E
Write to 0
D
Intrusion 2
C
512 bytes available for block 1
E
0 bytes available for block 1
D

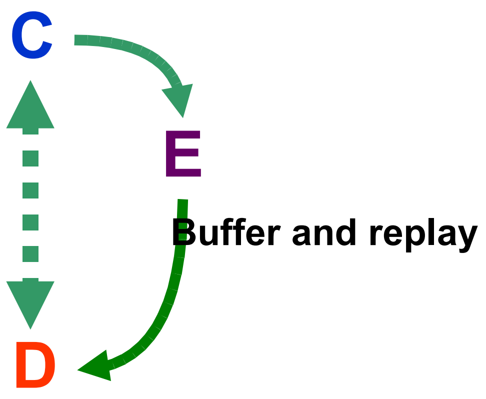
CIS-1 Threat Model

C: Controller **D**: Disk **E** : Eve

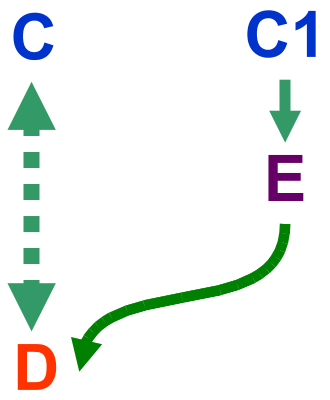
Normal


C
D

Intrusion 3
Write a to block 1
Write b to block 1
Write a to block 1

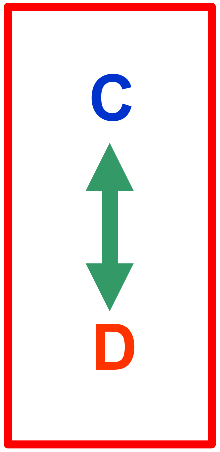

C
E
Buffer and replay
D

Intrusion 4
Buffer and replay
commands
from C1 to D


C
C1
E
D

CIS-1 Overwrite Protection

- Physical binding
  - Physical security
  - Programmed lock
- Virtual binding
  - Mutual authentication



Autovault: Secure Storage by Intelligent Locking

- **Secure enclosure with autolock**

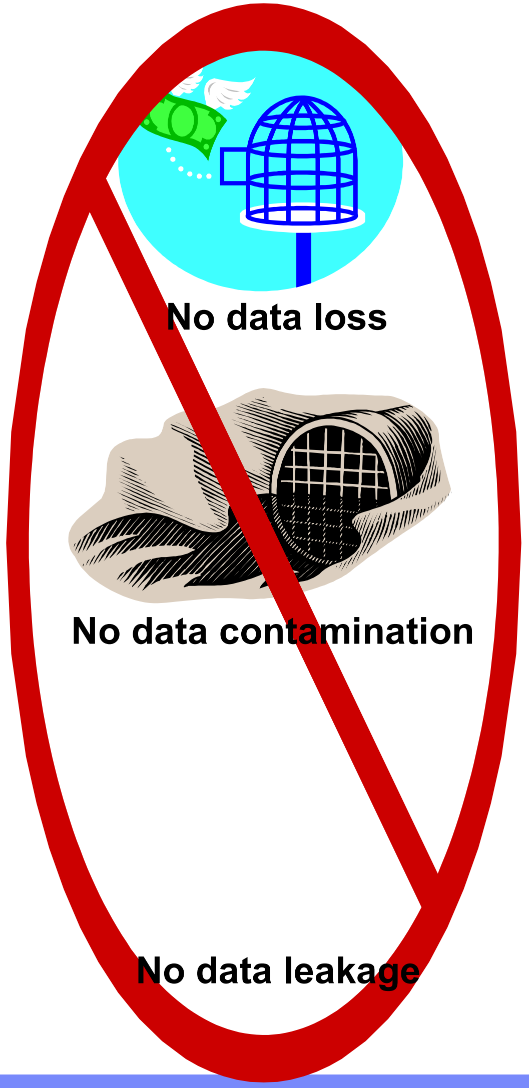
  firmware-controlled locking mechanism for each storage device

- **Advanced data protection features by leveraging locking mechanism**

  no loss of data by disallowing removal of storage devices beyond RAID protection

  no contamination of data from adding or replacing devices

  no leakage of data with removal of devices



No data loss

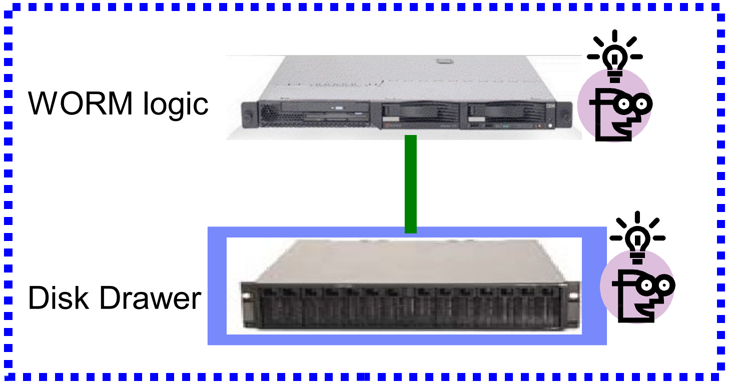
No data contamination

No data leakage

Virtual Binding: Secure Storage by Mutual Authentication

- Public-key cipher based Authentication
- Bytes verification at run time

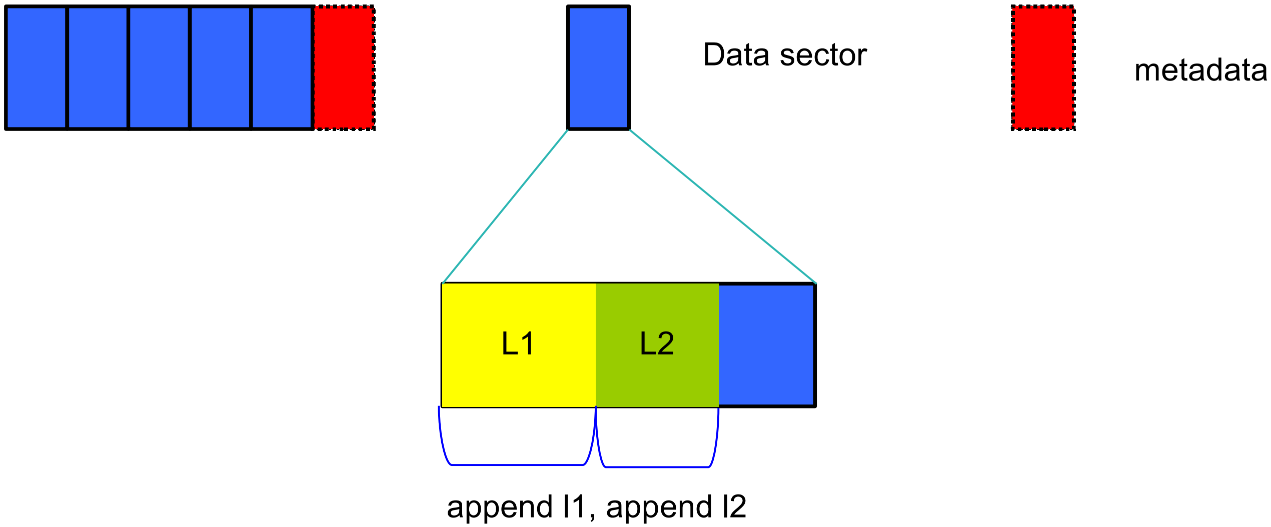
  HMAC based



- *Guaranteeing write-once semantic*
- *No sacrificing ease of storage management*

Block Append Capability

- Byte level granule append
- Space efficiency for index updates



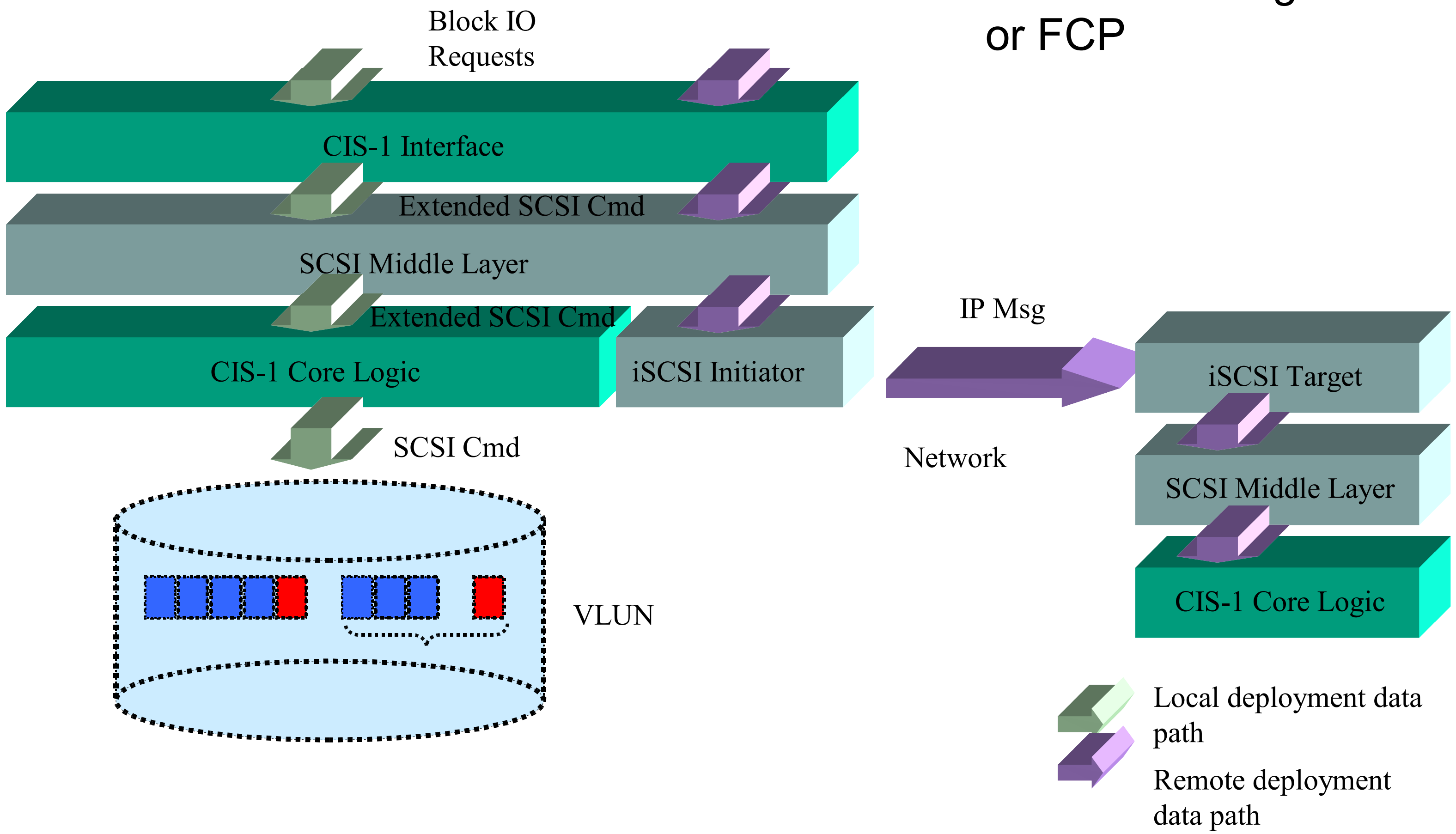
CIS-1 Interface: Standard SCSI Interface with Extended CDB

- Standard Read/Write
- SetRetention: term WORM
- Shred: secure shredding

- Append: efficient meta data writes
- etc

CIS-1 Software Architecture

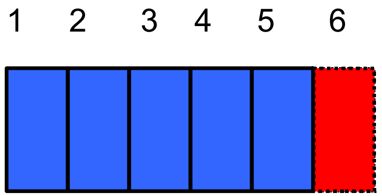
- Loadable block driver module in Linux 2.4.20
- Stackable on any virtual disk interface
- Minimum changes to iSCSI or FCP



CIS-1 Performance Optimization

- Disk layout
- I/O clustering
- Reduce lock contention

Disk Layout and I/O Clustering



- Reduce disk head movement align page size and extent size

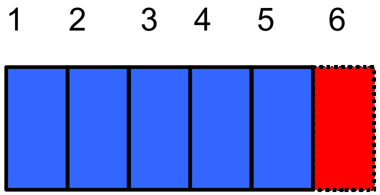
- I/O Clustering

  Operation: write block 4

  Read 6 ; read 4; write 4; write 6 ➔

  Read 4 5 6; write 4 5 6;

Reduce Lock Contention



Operation: write block 4

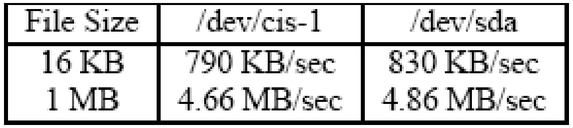
1. ***Lock 6***
2. Read 6 into memory if not cached
3. Write 4 to disk
4. Update 6 in memory
5. Write 6 to disk
6. ***Unlock 6***
7. Acknowledge write success

→

1. Read 6 into memory if not cached
2. ***Lock 6; Update 4 in memory;***
3. ***Update 6 in memory; Unlock 6***
4. Ack success if write cache is non-volatile
5. Write 4 to disk
6. Write 6 to disk
7. Ack success if write cache is volatile

CIS-1 Performance Evaluation

- **Synthetic file creation trace**
- **Synthetic I/O trace**



File Size	/dev/cis-1	/dev/sda
16 KB	790 KB/sec	830 KB/sec
1 MB	4.66 MB/sec	4.86 MB/sec

Table 1. File creation throughput for CIS-1 and rewritable storage. CIS-1 only adds a 5% or less throughput degradation.

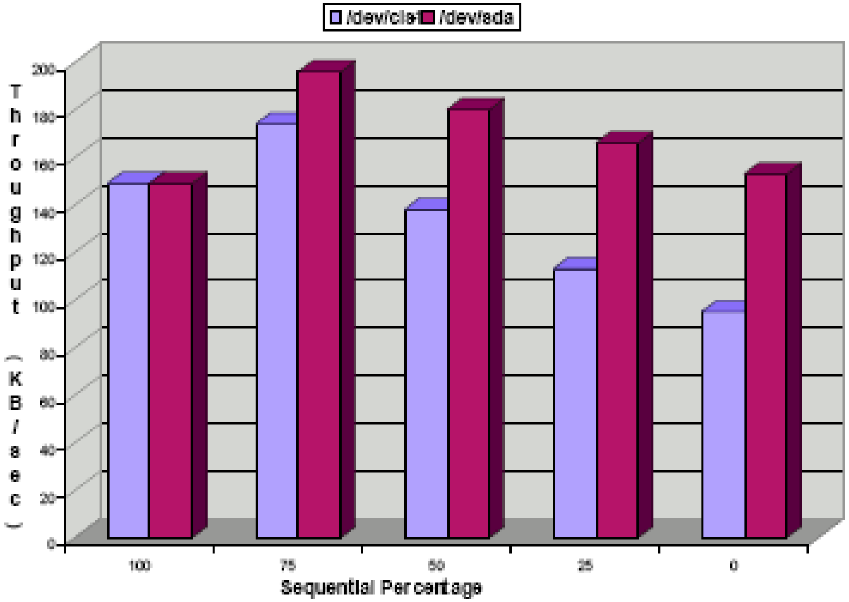


Figure 4. Throughput for CIS-1 and rewritable storage with varying percentage of sequential operations in the workload.

Related Work

- Venti
- Centera
- SnapLock
- OSD object-based storage device
- Self-Securing Storage (S3)

Summary

- Content Immutable Storage for trustworthy record keeping
  - Secure immutability
  - Efficient index support
  - Term-retention and disposition : term-WORM
  - Low cost and reliable
- Working prototype that provides comparable performance to rewritable media for target workload

Backup