Towards Regulatory Compliance in Data Management

CSE690 Lecture

Radu Sion
Stony Brook NSAC Lab
sion@cs.stonybrook.edu
Enron (1930-2001)
Electricity, gas, paper, communications.

2001
Stock drops from $90+ to pennies over allegations of corporate fraud.

December, 2001
Bankruptcy filling.

October 23 – November 9, 2001

November 8, 2001
SEC said: “stop the shredding!”

January 2001
Forensic experts start trying to recover missing documents with limited success.

2002
Congress issues the Sarbanes Oxley Act in direct response to the Enron scandal.
Overview

Finance

Healthcare

Government
e.g., HIPAA

Title I
Continuing health insurance coverage.

Title II
- Privacy Rule (all PHI)
- Transactions and Code Sets Rule
- Security Rule (electronic PHI)
  - Safeguards
    - administrative (policies and procedures)
    - physical
    - technical safeguards
- Unique Identifiers Rule
- Enforcement Rule
SEC. 1173 (d) (“Security Standards for Health Information”) mandates: “safeguards [. . . ] to ensure the integrity and confidentiality [. . . ] of the information” and “to protect against any reasonably anticipated [. . . ] threats or hazards to the [. . . ] integrity of the information” (e.g., once stored).

Impact Layers

Hardware
Tamper-resistance, magnetic Residues, emanations

OS
I/O device drivers and kernel

Storage
Block level: WORM assurances, secure migration (1)
FS level: secure indexing, secure deletion, secure provenance, history independent data structures, secure migration (2)

Databases
History Independence – novel indexing
ACID still holds?

Networks
Physical level: wireless spectrum sharing behavior
Packet level: anti-spam, flow labeling
WORM: Overview

Prevent this!

Owner

Federal Investigators

Data Records

oops: regret $b_2$

We need $b_2$!

Enterprise Data Management

$\text{Federal Investigators}$

$\text{WORM: Overview}$

Stony Brook Network Security and Applied Cryptography Lab
We do not prevent history. Just history *rewriting*.
A bit artificial in scope – why do we trust the owner to correctly log the entries and then mistrust her later? If I were a malicious owner, I would simply not log suspicious emails 😊

Do we trust owner for the next 5 minutes too?
What is $\Delta t = t_3 - t_1$ (“time warp”)? If we know this, we can deploy all kinds of optimizations.

**Trustworthy Indexing.**
When is this an issue? Searches usually conducted through indexes.

**Secure Deletion.**
Is a problem only if trustworthy indexing is required.

**Secure Migration.**
Relatively straightforward. Build trust chain, deal with obsolete encryption, lack of keys.

**Litigation support.**
Need to make sure retention can be prolonged in the case of an ongoing litigation.
**Tape-based**
Assumption: specific reader is used. Checksums (keyed) are written onto tape. Keys are managed inside readers.

**Optical Disks**
Problem: physical storage space, cost, replication attacks, high latency. No secure deletion.

**Hard Disks**
Main problem: “soft”-ware.
DLTSage WORM

Assurances of the tape systems are provided under the assumption that compliant tape-readers are deployed. “DLTSage WORM provides features to assure compliance, placing an electronic key on each cartridge to ensure WORM integrity. This unique identifier cannot be altered, providing a tamperproof archive cartridge that meets stringent compliance requirements to ensure integrity protection and full accessibility with reliable duplication.”
Sony Disk for Data

Holds only 23 GB per disk side. Because it is faster than tape and cheaper than hard disks, optical WORM storage technology is often deployed as a secondary, high-latency storage medium to be used in the framework of a hard disk-based solution. Care needs to be taken in establishing points of trust and data integrity when information leaves the secured hard disk store for the optical media.
EMC Centera

Content addressed storage (CAS) software solution with regulatory compliance capabilities. Data records have “two components: the content and its associated content descriptor file (CDF) that is directly linked to the stored object […] A digital fingerprint derived from the content itself is the content’s locator. […]

The CDF is used for access to and management of the record. Within this CDF, the application will assign a retention period for each individual business record. Centera will permit deletion of a pointer to a record upon expiration of the retention period. Once the last pointer to a record has been so deleted, the object will be eliminated”, and, in the Plus version, also “shredded” (from the media).
Hitachi Message Archive for Compliance

The Data Retention Utility is a software-based “virtual” WORM mechanism for mainstream Hitachi storage systems. It allows customers to “lock down archived data, making it non-erasable and non-rewritable for prescribed periods, facilitating compliance with governmental or industry regulations”. 
IBM LockVault compliance software

Software layer that operates on top of IBM System Storage N series to provide “disk-based regulatory compliance solutions for unstructured data”.

IBM System Storage Archive Manager

The IBM Tivoli Storage Manager is part of the IBM Total Storage Software and provides certain software data retention protection. It “makes the deletion of data before its scheduled expiration extremely difficult. Short of physical destruction to storage media or server, or deliberate corruption of data or deletion of the Archive Manager database, Archive Manager will not allow data [...] to be deleted before its scheduled expiration date.”
Snaplock Compliance/Enterprise Software

A software suite designed to work on top of NetApp NearStore and FAS storage systems. It provides soft-WORM assurances, “preventing critical files from being altered or deleted until a specified retention date”. As opposed to other vendors, NetApp SnapLock supports open industry standard protocols such as NSF and CIFS.
StorageTek Compliance Archiving Software

Software that runs on top of the Sun StorageTek 5320 NAS Appliance to “provide compliance-enabling features for authenticity, integrity, ready access, and security”.
"soft"-WORM

Stony Brook Network Security and Applied Cryptography Lab

Owner

Data Records

oops: regret $b_2$

tamper with checksums

State

We need $b_2$!

Federal Investigators

Lack of tamper-proof hardware makes this possible!
Tamper-proof Hardware.
Achieving WORM in the absence of tamper-proof hardware is not possible.

Q: What kind of tamper-proof hardware?
Microcontroller that stores keys, passwords and digital certificates.
Can the Trusted Platform Module control what software runs?
No. […] it can only act as a 'slave' to higher level services and applications by storing and reporting pre-runtime configuration information. […] At no time can the TCG building blocks 'control' the system or report the status of applications that are running.
Would a TCG/TPM help?

The passive nature of a TPM would require an additional point of blank trust in upper layer code. The ability to virtualize makes this hard to achieve.

**Discussion:** How would Mallory fake a world view to the TPM. Remember we are talking about millions of US dollars worth of incentives here.

And by the way …

… TPMS have been successfully hacked by attackers with almost no resources (see refs).
Active Tamper-proof Hardware.
Achieving WORM in the absence of active tamper-proof hardware is not possible.
**SCPU: Performance**

- RSA1024 Sign: **848/sec**
- RSA1024 Verify: **1157/sec**
- 3DES: **1-8MB/sec**
- DES: **1-8MB/sec**
- SHA1: **1-21MB/sec**

IBM 4764-001: 266MHz PowerPC. 64KB battery-backed SRAM storage. Crypto hardware engines: AES256, DES, TDES, DSS, SHA-1, MD5, RSA. FIPS 140-2 Level 4 certified.
**Strawman Merkle**

**Idea:** no need to be binary

**trust this (store or authenticate)**

**compare**

**Idea: sign stuff (when ?)**

![Diagram of a Merkle tree with hash functions and data points](image)

- **H(.)**
- **H(.)**
- **H(.)**
- **H(.)**
- **H(.)**
- **H(.)**
- **H(.)**
- **H(.)**

Data points:
- **x₁**
- **x₂**
- **x₃**
- **x₄**
- **x₅**
- **x₆**
- **x₇**
- **x₈**
Insight (1)

Stony Brook Network Security and Applied Cryptography Lab

Owner

Storage

Main CPU

Trusted CPU

Enterprise Data Management

write $b_i$

$S_{s}(SN_{base})$

$S_{s}(SN_{current})$

$S_{s}(b_i, SN_{i}, ...)$

$SN_1$

$SN_2$

$SN_3$

$SN_4$

$SN_5$

$SN_3 < SN_{base}$?

$SN_3 > SN_{current}$?

read $SN_3$

$SN_3 = SN_{current}$?

$SN_3 < SN_{base}$?

$SN_3 > SN_{current}$?

monotonically increasing consecutive serial numbers
Main CPU’s proof ("deletion witness") that \( b_3 \) was correctly deleted.

\[ \text{read } SN_3 \]

\[ SN_3 > SN_{\text{current}}? \]

\[ S_d(SN_3) \]

\[ S_d(SN_{\text{current}}) \]

Sorted retention expiration times

Monotonically increasing serial numbers
Efficient Digests?

Issue
SCPU data digestion (hashing) is not very fast.

Fact
We already assume the stored data is accurate.

Question
Why not also trust the main CPU to produce correct data digests *at write time*? This should increase throughput.

How
To prevent cheating we double check during idle times (or mandatory if too much time passes).
How do we maintain the VRDT efficiently.
Hierarchical. Arbitrary “deletion” windows.

How does the SCPU/RM enforce deletion efficiently.
Alarms, efficient index structures of retention expiration times.

How can we “witness” things fast: amortization.
In times of high-load: defer expensive witnessing and use short-lived constructs.
During idle/low-load times: re-enforce short-lived constructs.

How fast can we go.
Writes: 3600-3700 updates/second (4-6hrs. bursts), 450-500 updates/sec (sustained).
Reads: limited only by un-trusted system segment.

What about litigation support.
Authorized regulatory parties present credentials and are allowed to set/reset litigation holds.
What about migration?

1. Owner
2. Compliant Migration Manager
3. Regulatory Authority
4. $S_{RA}(MC(S_1,S_2))$
5. mutual authentication and key exchange
6. data migration over secure channel

Stony Brook Network Security and Applied Cryptography Lab

Systems Security
What Next?

Namespaces, Search Indexes
Trust-worthy Indexing

More Complex Migration
Complex query-driven migration

Secure Deletion
History Independent Data Structures, logging etc.

New Query Languages/Paradigms?
Do transactional semantics still hold in the presence of regulatory compliance? Can we extend SQL to deal with e.g., WORM assurances?
Refs: Regulations