NSAC

Network Security and Applied Cryptography Laboratory

http://crypto.cs.stonybrook.edu

Towards Regulatory Compliance in Data Management

CSE690 Lecture

Radu Sion
Stony Brook NSAC Lab
sion@cs.stonybrook.edu

National Science Foundation
WHERE DISCOVERIES BEGIN

STONY BROOK
COMPUTER SCIENCE

Let’s talk about … ENRON ☺

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
Accounting firm Arthur Andersen destroys tens of thousands of digital/paper documents, *despite knowledge of a forthcoming U.S. SEC subpoena*.

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

9>8 ☺

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

National Association of Insurance Commissioners, **Graham-Leach-Bliley Act**, 1999; The U.S. Securities and Exchange Commission, **Rule 17a-3&4, 17 CFR Part 240**: Electronic Storage of Broker-Dealer Records, 2003; U.S. Public Law No. 107-204, 116 Stat. 745, The Public Company Accounting Reform and Investor Protection Act, 2002 (**Sarbanes-Oxley**)

Healthcare

U.S. Dept. of Health & Human Services, The Health Insurance Portability and Accountability Act (**HIPAA**), 1996; The U.S. Department of Health and Human Services Food and Drug Administration, **21 CFR Part 11**: Electronic Records and Signature Regulations 1997

Government

U.S. Public Law 107-347. The E-Government Act, 2002 (Federal Information Security Management Act **FISMA**); The U.S. Department of Defense, **Directive 5015.2**: DOD Records Management Program, 2002; The U.S. Department of Education. 20 U.S.C. 1232g; 34 CFR Part 99:The Family Educational Rights and Privacy Act (**FERPA**), 1974

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

e.g., HIPAA (cont'd)

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).

http://www.cms.hhs.gov/HIPAAGenInfo/Downloads/HIPAALaw.pdf

Impact Layers

Hardware

Tamper-resistance, magnetic Residues, emanations

OS

I/O device drivers and kernel

Storage

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

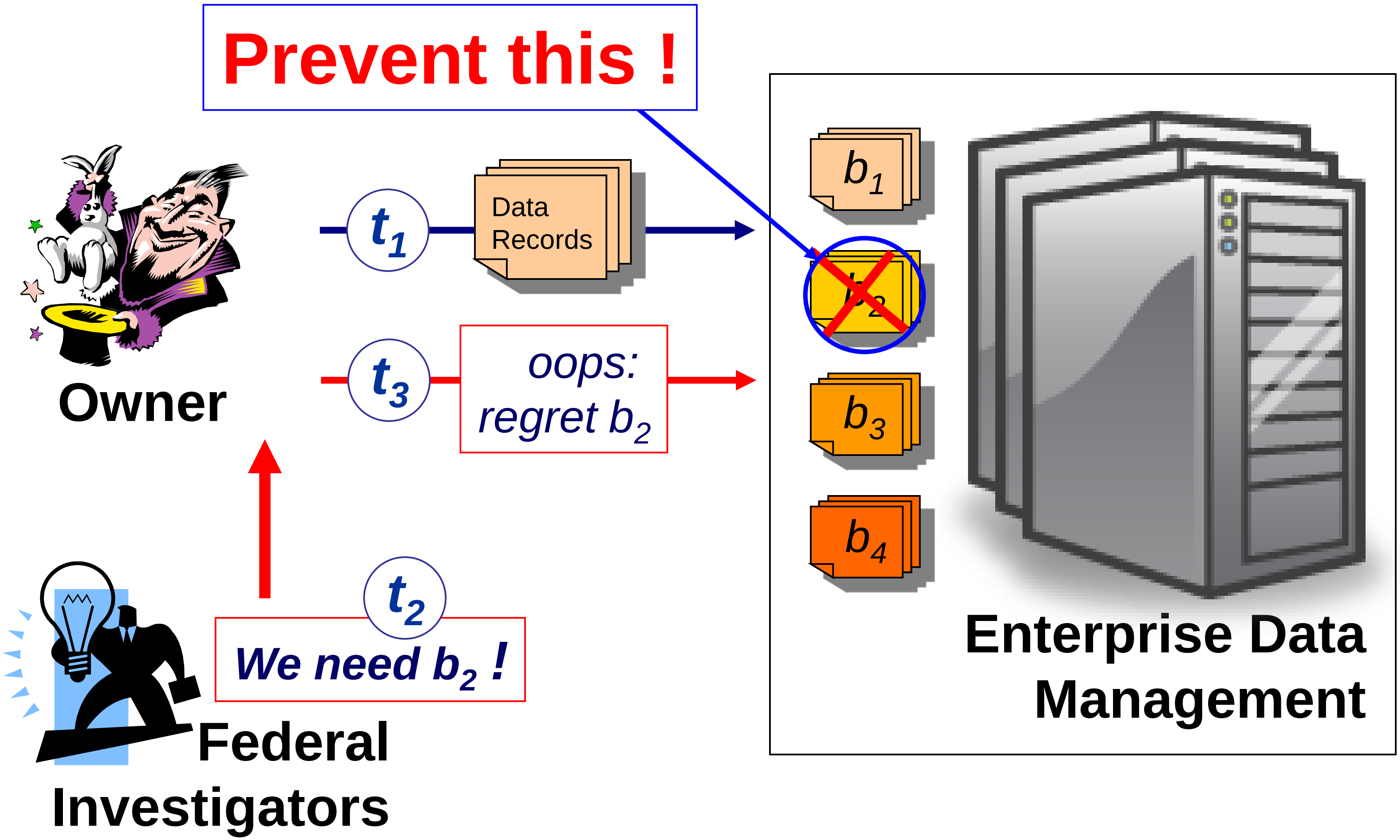
Databases

History Independence – novel indexing
ACID still holds ?

Networks

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

WORM: Overview



WORM: Some Properties

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.

WORM: Existing Systems

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.

Tape: Quantum

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.”

CD: Sony

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) <u>software</u> 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.”

Disk: Network Appliance

Snaplock Compliance/Enterprise Software

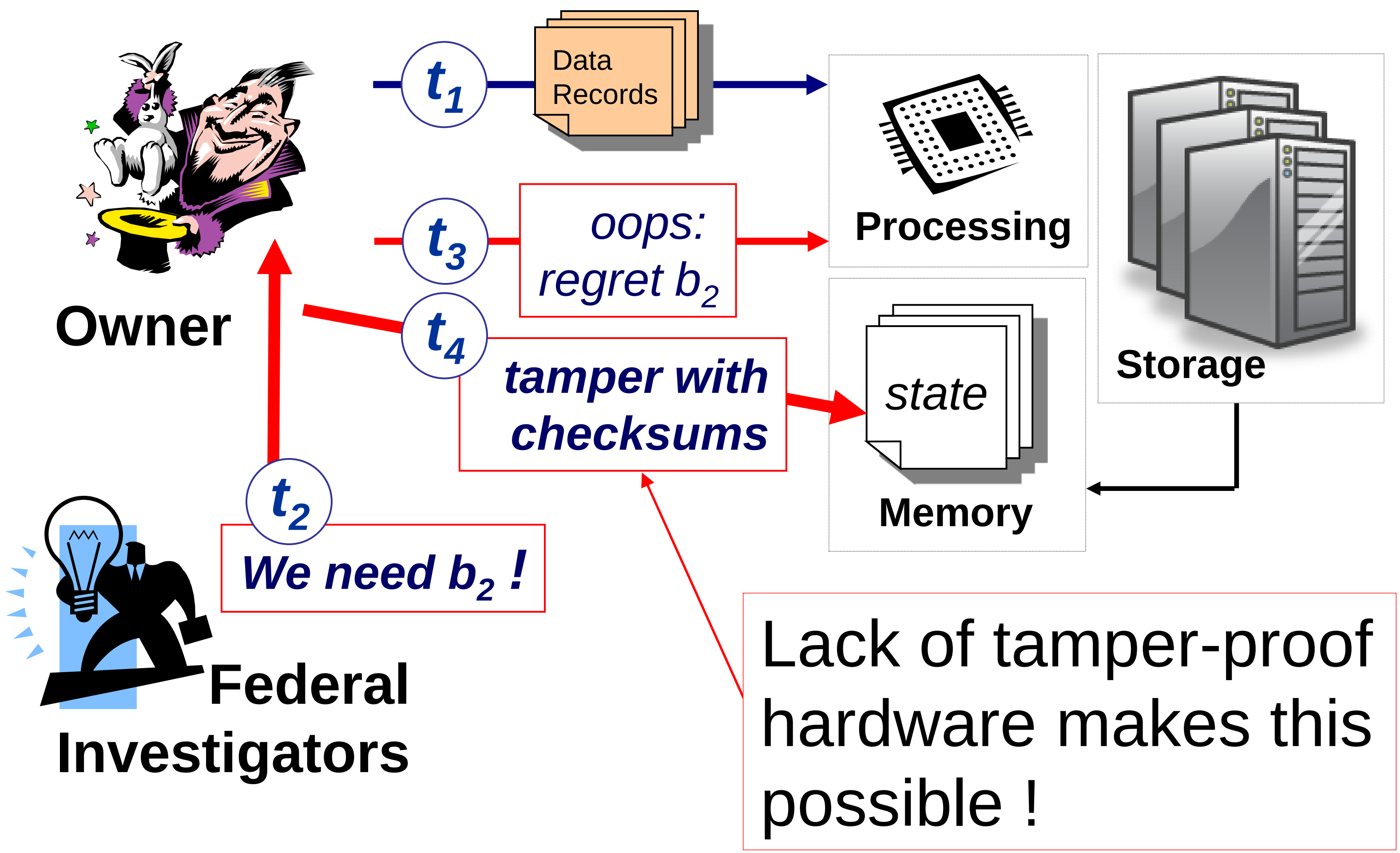
A software suite designed to work on top of NetApp NearStore and FAS storage systems. It provides soft-WORM assurances, “preventing critical files from being altered or deleted until a specified retention date”. As opposed to other vendors, NetApp SnapLock supports open industry standard protocols such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

Disk: Sun

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



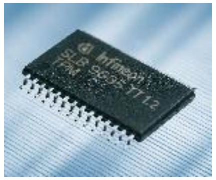
Owner
t_1
Data Records
Processing
Storage
t_3
oops: regret b_2
t_4
tamper with checksums
state
Memory
t_2
We need b_2 !
Federal Investigators
Lack of tamper-proof hardware makes this possible !

WORM: Claim One

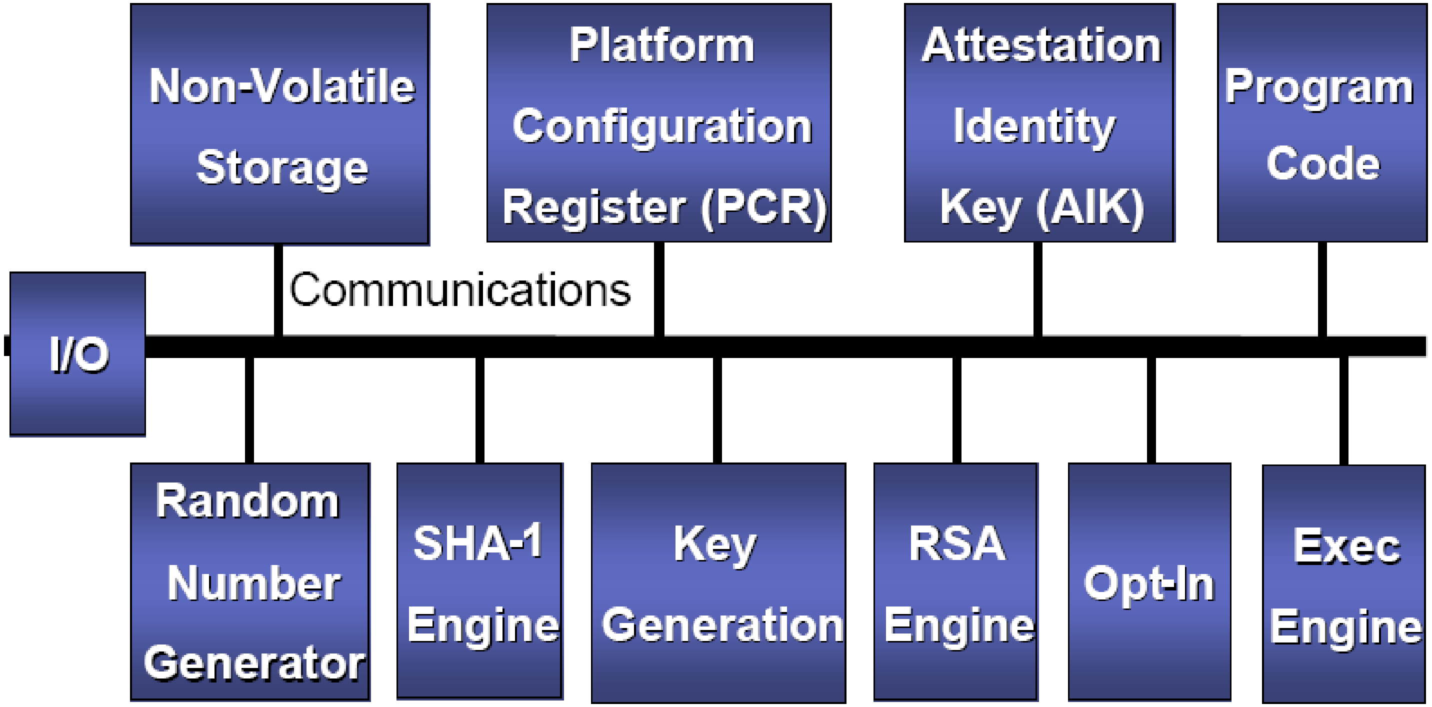
Tamper-proof Hardware.
Achieving WORM in the absence of tamper-proof hardware is not possible.

Q: *What kind of tamper-proof hardware ?*

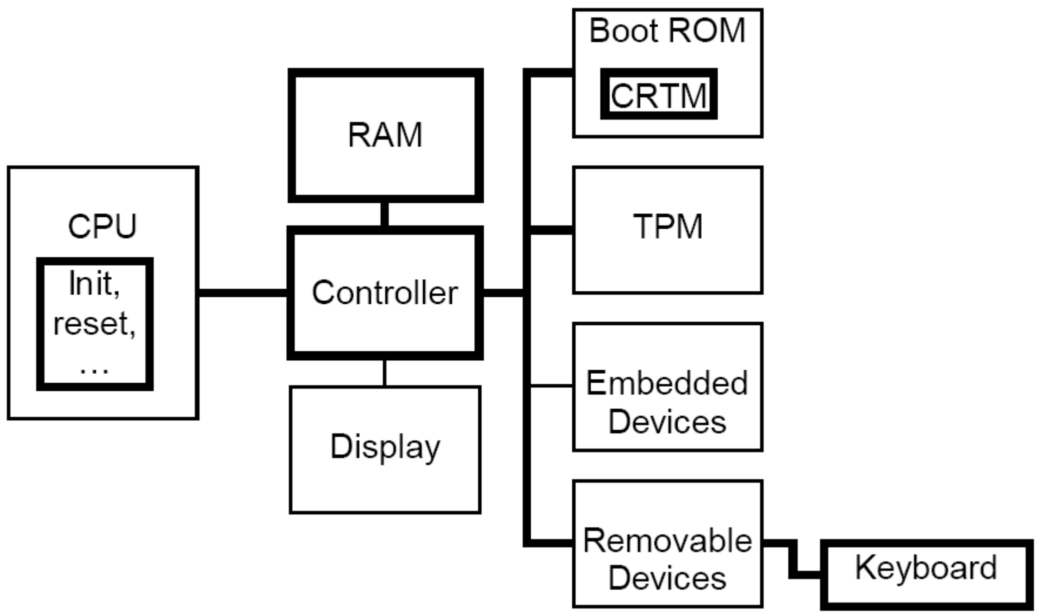
TPM



Microcontroller that stores keys, passwords and digital certificates.



TPM: Trust Chain



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.

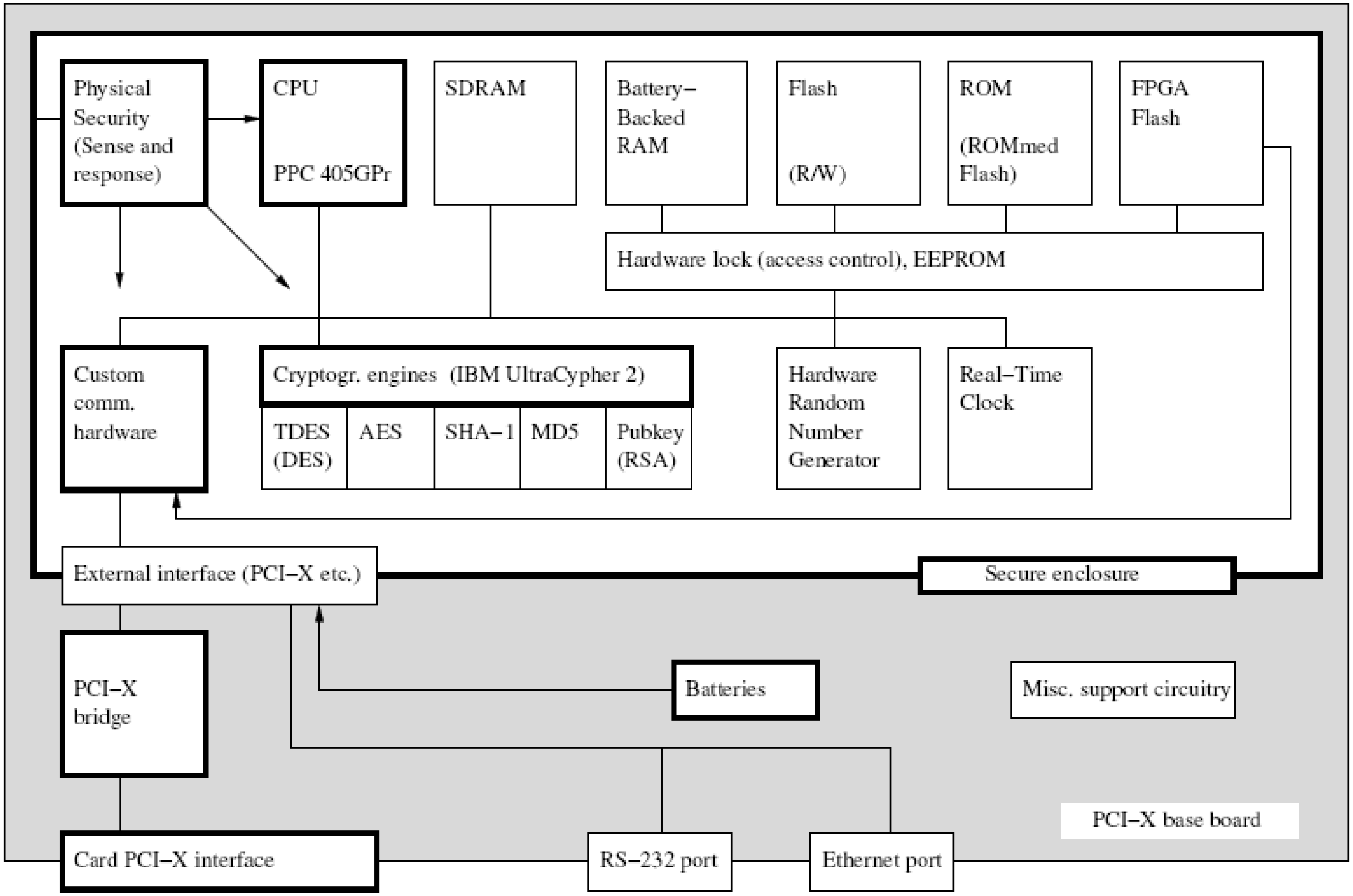
Discussion: How would Mallory fake a world view to the TPM. Remember we are talking about *millions of US dollars* worth of incentives here.

And by the way …

… **TPMs have been successfully hacked** by attackers with almost no resources (see refs).

SCPUs (IBM 4764)

Stony Brook Network Security and Applied Cryptography Lab



WORM: Claim Two

Active Tamper-proof Hardware.

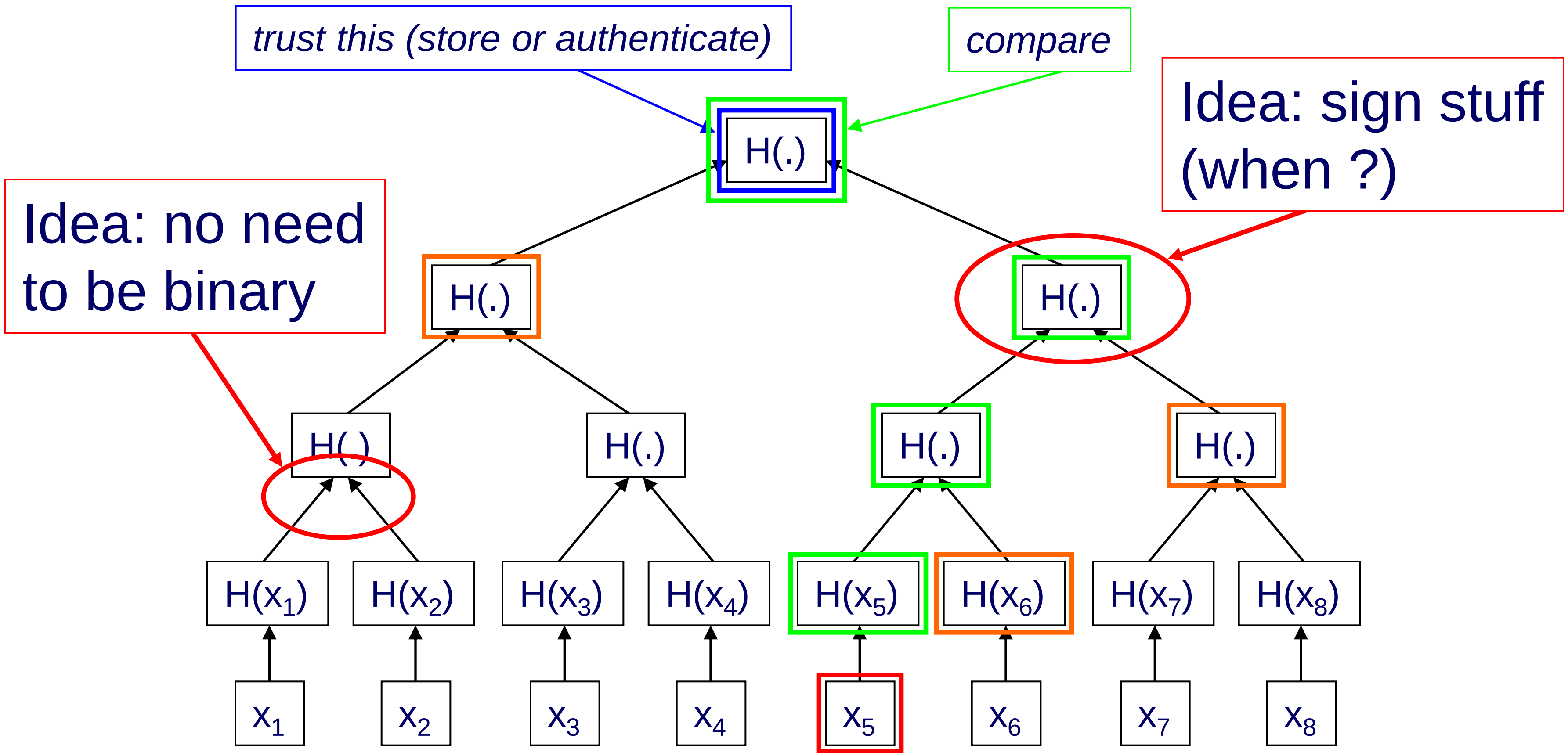
Achieving WORM in the absence of active tamper-proof hardware is not possible.

SCPU: Performance

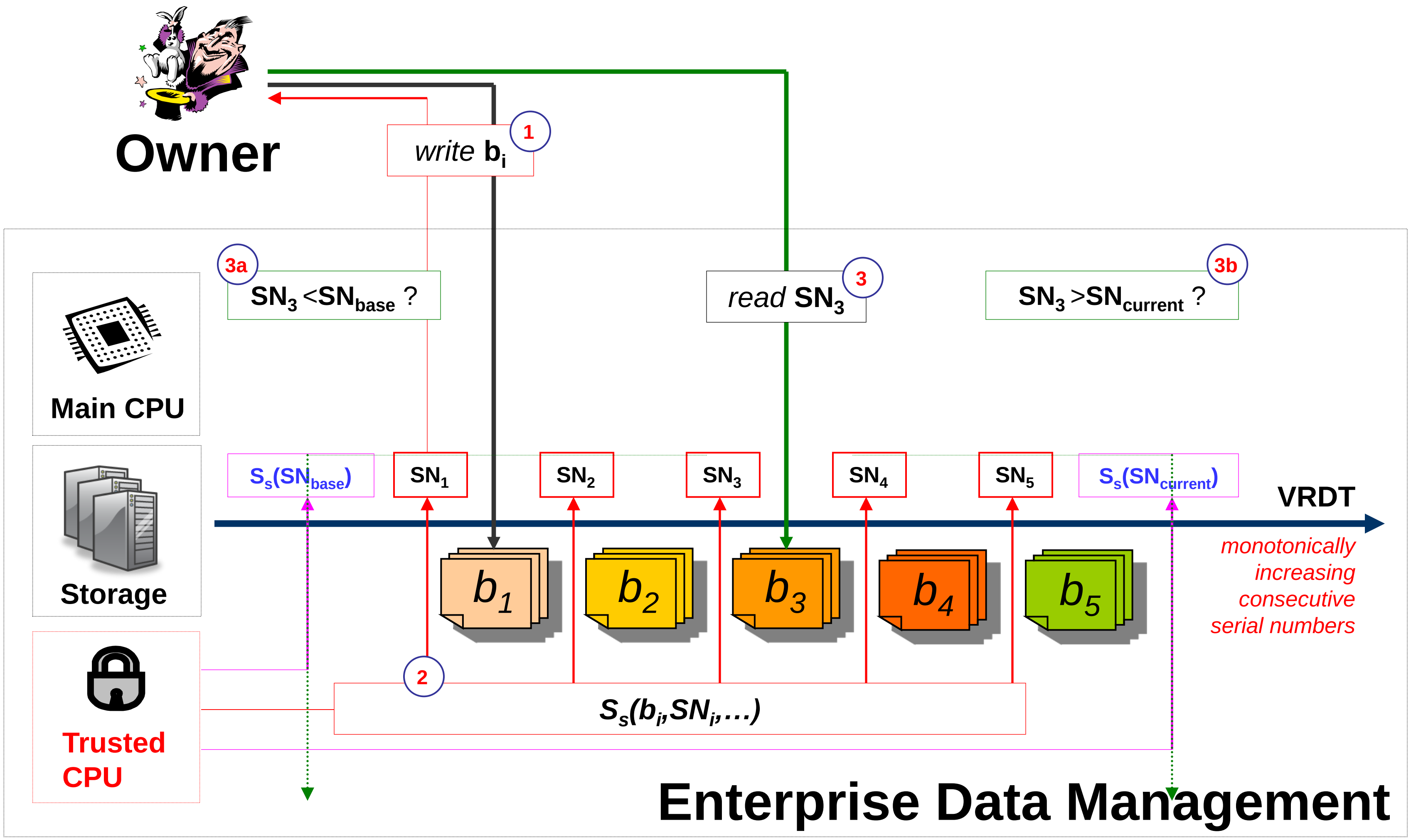
RSA1024 Sign: **848/sec**
RSA1024 Verify: **1157/sec**
3DES: **1-8MB/sec**
DES: **1-8MB/sec**
SHA1: **1-21MB/sec**

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.

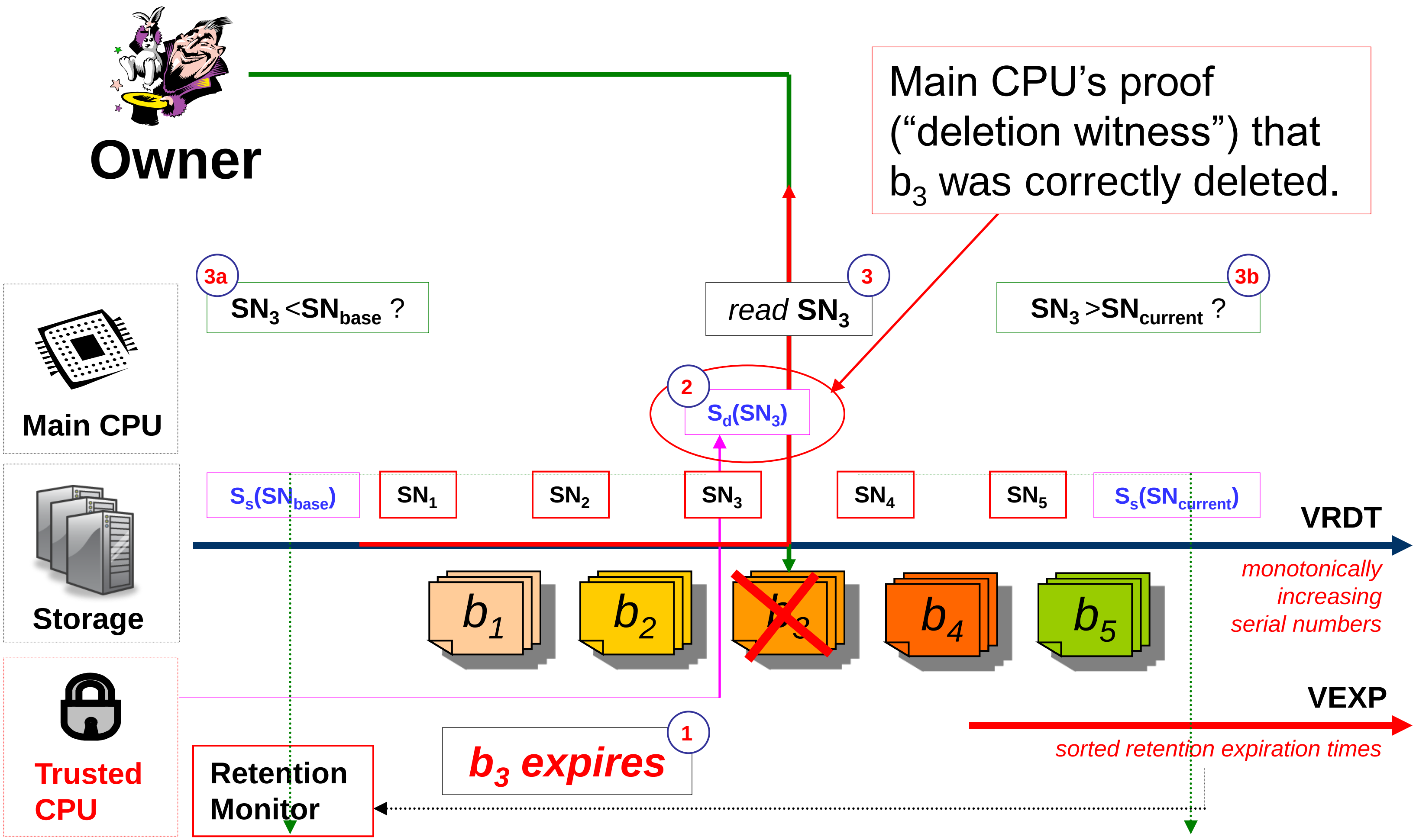
Strawman Merkle



Insight (1)


Owner
write b_i
1
3a
$SN_3 < SN_{base}$?
read SN_3
3
3b
$SN_3 > SN_{current}$?
Main CPU
Storage
Trusted CPU
$S_s(SN_{base})$
SN_1
SN_2
SN_3
SN_4
SN_5
$S_s(SN_{current})$
VRDT
b_1
b_2
b_3
b_4
b_5
monotonically increasing consecutive serial numbers
2
$S_s(b_i, SN_i, \ldots)$
Enterprise Data Management

Insight (2)


Owner
Main CPU's proof ("deletion witness") that b3 was correctly deleted.
3a
SN3 <SNbase ?
read SN3
3
SN3 >SNcurrent ?
3b
Main CPU
2
Sd(SN3)
Storage
Ss(SNbase)
SN1
SN2
SN3
SN4
SN5
Ss(SNcurrent)
VRDT
b1
b2
b3
b4
b5
monotonically increasing serial numbers
VEXP
Trusted CPU
Retention Monitor
b3 expires
1
sorted retention expiration times

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 <u>main</u> 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).

Can We Eat The Cake Too ?

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.

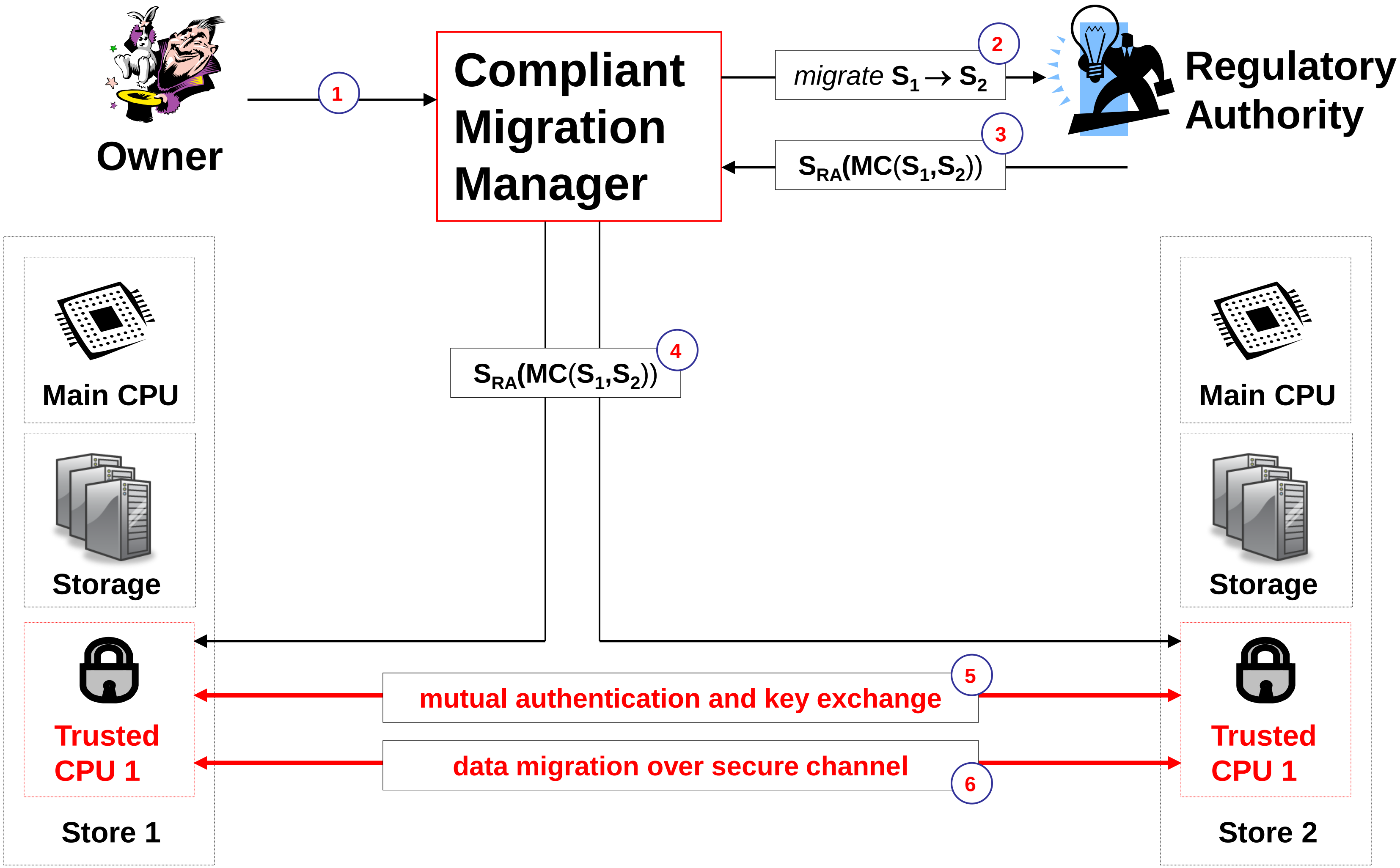
How can we “witness” things fast: amortization.
In times of high-load: defer expensive witnessing and use short-lived constructs.
During idle/low-load times: re-enforce short-lived constructs.

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?

Stony Brook Network Security and Applied Cryptography Lab



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: Trusted Hardware

1. Attacks on TPMs. Online at http://www.cs.dartmouth.edu/~pkilab/sparks/.
2. TWIRL: The Weizmann Institute Relation Locator. Online at http://www.wisdom.weizmann.ac.il/~tromer/twirl/.
3. IBM PCI-X Cryptographic Coprocessor. Online at http://www-03.ibm.com/security/cryptocards/pcixcc/overperformance.shtml, 2003.
4. IBM 4758 PCI Cryptographic Coprocessor. Online at http://www-03.ibm.com/security/cryptocards/pcicc/overview.shtml, 2006.
5. IBM Common Cryptographic Architecture (CCA) API. Online at http://www-03.ibm.com/security/cryptocards//pcixcc/overcca.shtml, 2006.
6. Trusted Computing Platforms Storage: Compliance, Security, and Policy. Online at https://www.trustedcomputinggroup.org/news/presentations/SNIA Security Summit 2006.pdf, January 2006.
7. IBM Cryptographic Hardware. Online at http://www-03.ibm.com/security/products/,2007.
8. Trusted Computing Group. Online at https://www.trustedcomputinggroup.org/, 2007.
9. Trusted Platform Module (TPM) Specifications. Online at https://www.trustedcomputinggroup.org/specs/TPM,2007.
10. Bernhard Kauer. OSLO: Improving the Security of Trusted Computing. In USENIX Security Symposium, 2007.
11. Smart Card Alliance. HIPAA compliance and smart cards: Solutions to privacy and security requirements. Online at http://www.datakey.com/resources/HIPAA_Compliance_and_Smart_Cards_FINAL.pdf, Sep. 2003.

Refs: Regulations

1. NIST Federal Information Processing Standards. Online at http://csrc.nist.gov/publications/fips/, 2007.
2. Protiviti Consulting. Frequently Asked Questions About J-SOX. Online at http://www.protiviti.jp/downloads/JSOXOverviewfinal\ E.pdf, 2006.
3. National Association of Insurance Commissioners. Graham-Leach-Bliley Act, 1999. www.naic.org/GLBA.
4. Ministry of Finance. Bill 198 of 2002. An Act to implement budget measures and other initiatives of the Government. Legislative Assembly of Ontario, 2002.
5. British Parliament. Data protection act of 1998. Online at http://www.staffs.ac.uk/legal/privacy/dp10rules/, 1998.
6. European Parliament. European directives. Online at http://ec.europa.eu/justice\ home/fsj/privacy/law/index\ en.htm, 2006.
7. Australian Securities and Exchange Commission. Clerp 9 corporate reporting and disclosure laws. Online at http://www.asic.gov.au, 2004.
8. The Enterprise Storage Group. Compliance: The effect on information management and the storage industry. Online at http://www.enterprisestoragegroup.com/, 2003.
9. The U.S. Department of Defense. Directive 5015.2: DOD Records Management Program. Online at http://www.dtic.mil/whs/directives/corres/pdf/50152std 061902/p50152s.pdf, 2002.
10. The U.S. Department of Education. 20 U.S.C. 1232g; 34 CFR Part 99:Family Educational Rights and Privacy Act (FERPA). Online at http://www.ed.gov/policy/gen/guid/fpco/ferpa, 1974.
11. The U.S. Department of Health and Human Services Food and Drug Administration. 21 CFR Part 11: Electronic Records and Signature Regulations. Online at http://www.fda.gov/ora/compliance ref/part11/FRs/background/pt11finr.pdf, 1997.
12. The U.S. Securities and Exchange Commission. Rule 17a-3&4, 17 CFR Part 240: Electronic Storage of Broker-Dealer Records. Online at http://edocket.access.gpo.gov/cfr2002/aprqtr/17cfr240.17a-4.htm, 2003.
13. U.S. Dept. of Health & Human Services. The Health Insurance Portability and Accountability Act (HIPAA), 1996. www.cms.gov/hipaa.
14. U.S. Public Law 107-347. The E-Government Act, 2002.
15. U.S. Public Law No. 107-204, 116 Stat. 745. The Public Company Accounting Reform and Investor Protection Act, 2002.
16. N. Lawson, J. Orr, and D. Klar. The HIPAA privacy rule: An overview of compliance initiatives and requirements. Defense Counsel Journal, 70:127–149, 2003.
17. Occupational Safety and Health Administration. Access to employee exposure and medical records. - 1910.1020 regulations (standards - 29 cfr). Online at http://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=STANDARDS&p_id=10027.
18. U. S. Congress. Federal rules of civil procedure. Online at http://www.law.cornell.edu/rules/frcp/, 2006.

Refs: Crypto

1. TWIRL: The Weizmann Institute Relation Locator. Online at http://www.wisdom.weizmann.ac.il/~tromer/twirl/.
2. Mihir Bellare and Daniele Micciancio. A New Paradigm for Collision-Free Hashing: Incrementality at Reduced Cost. In Walter Fumy, editor, Advances in Cryptology — Proceedings of EuroCrypt 1997
3. Dwaine E. Clarke, Srinivas Devadas, Marten van Dijk, Blaise Gassend, and G. Edward Suh. Incremental multiset hash functions and their application to memory integrity checking. In Chi-Sung Laih, editor, ASIACRYPT 2003
4. A.K. Lenstra et al. Analysis of Bernstein's Factorization Circuit. Advances in Cryptology, pages 1–26, 2002.
5. J. Franke et al. SHARK: A Realizable Special Hardware Sieving Device for Factoring 1024-Bit Integers. In Cryptographic Hardware and Embedded Systems, 2005.
6. W. Geiselmann et al. Scalable Hardware for Sparse Systems of Linear Equations, with Applications to Integer Factorization. In Cryptographic Hardware and Embedded Systems CHES, 2005.
7. W. Geiselmann et al. A Simpler Sieving Device: Combining ECM and TWIRL. In Intl. Conf. on Information Security and Cryptology, 2006.
8. N. Ferguson and B. Schneier. Practical Cryptography. Wiley & Sons, 2003.
9. W. Geiselmann and R. Steinwandt. Hardware for Solving Sparse Systems of Linear Equations over GF(2). In Cryptographic Hardware and Embedded Systems CHES, 2003.
10. W. Geiselmann and R. Steinwandt. Special Purpose Hardware in Cryptanalysis: The Case of 1024-bit RSA. IEEE Security and Privacy, pages 63–66, January 2007.
11. O. Goldreich. Foundations of Cryptography, Cambridge University Press, 2001.
12. Jetico, Inc. BestCrypt software home page. www.jetico.com, 2002.
13. A.K. Lenstra and A. Shamir. Analysis and Optimization of the TWINKLE Factoring Device. In EuroCrypt, 2000.
14. Alfred J. Menezes, Paul C. van Oorschot, and Scott A. Vanstone. Handbook of Applied Cryptography. CRC Press, 2001.
15. R. Merkle. Protocols for public key cryptosystems. In IEEE Symposium on Research in Security and Privacy, 1980.
16. C. Pomerance. A Tale of Two Sieves. Notices of the ACM, pages 1473–1485, December 1996.
17. B. Schneier. Applied Cryptography: Protocols, Algorithms and Source Code in C. Wiley & Sons, 1996.
18. A. Shamir. Factoring Large Numbers with the TWINKLE Device. In Cryptographic Hardware and Embedded Systems, 1999.
19. Robert D. Silverman. A cost-based security analysis of symmetric and asymmetric key lengths. Online at http://www.rsasecurity.com/rsalabs/bulletins/index.html.

Refs: Industry WORM

1. EMC. Centera Compliance Edition Plus. Online at http://www.emc.com/centera/ and http://www.mosaictech.com/pdfdocs/emc/centera.pdf, 2007.
2. Hitachi Data Systems. The Message Archive for Compliance Solution, Data Retention Software Utility. Online at http://www.hds.com/solutions/data life cycle archiving/achievingregcompliance.html, 2007.
3. HP. WORM Data Protection Solutions. Online at http://h18006.www1.hp.com/products/storageworks/wormdps/index.html, 2007.
4. IBM Corp. IBM System Storage N series with LockVault compliance software. Disk-based regulatory compliance solutions for unstructured data. Online at http://www-03.ibm.com/systems/storage/network/software/lockvault/, 2007.
5. IBM Corp. IBM Total Storage Family: Tivoli System Storage Archive Manager. Online at http://www-306.ibm.com/software/tivoli/products/storage-mgr-data-reten/, 2007.
6. IBM Corp. IBM TotalStorage Enterprise. Online at http://www-03.ibm.com/servers/storage/, 2007.
7. IBM Corporation and Daniel James Winarski and Kamal Emile Dimitri. United States Patent 6879454: Write-Once Read-Many Hard Disk Drive, 2005.
8. Network Appliance Inc. SnapLock Compliance and SnapLock Enterprise Software, Online at http://www.netapp.com/products/software/snaplock.html, 2007.
9. Quantum Inc. DLTSage Write Once Read Many Solution. Online at http://www.quantum.com/Products/TapeDrives/DLT/SDLT600/DLTIce/Index.aspx and http://www.quantum.com/pdf/DS00232.pdf, 2007.
10. StorageTek Inc. VolSafe secure tape-based write once read many (WORM) storage solution. Online at http://www.storagetek.com/, 2007.
11. Sun Microsystems. Sun StorageTek Compliance Archiving Software. Online at http://www.sun.com/storagetek/management software/data protection/compliance archiving/,2007.
12. Sun Microsystems. Sun StorageTek Compliance Archiving system and the Vignette Enterprise Content Management Suite (White Paper). Online at http://www.sun.com/storagetek/white-papers/Healthcare Sun NAS Vignette EHR 080806 Final.pdf, 2007.
13. Zantaz Inc. The ZANTAZ Digital Safe Product Family. Online at http://www.zantaz.com/,2007.

Refs: Research on WORM

1. Malcolm C. Easton. Key-Sequence Data Sets on Indelible Storage. IBM Journal of Research and Development, May 1986.
2. W. Hsu and S. Ong. Fossilization: A Process for Establishing Truly Trustworthy Records. IBM Research Report, (10331), 2004.
3. Lan Huang, Windsor W. Hsu, and Fengzhou Zheng. CIS: Content Immutable Storage for Trustworthy Record Keeping. In Proceedings of the Conference on Mass Storage Systems and Technologies (MSST), 2006.
4. Soumyadeb Mitra, Windsor W. Hsu, and Marianne Winslett. Trustworthy Keyword Search for Regulatory-Compliant Records Retention. In Proceedings of UIUC, 2006.
5. Soumyadeb Mitra and Marianne Winslett. Secure Deletion from Inverted Indexes on Compliance Storage. In Proceedings of the StorageSS Workshop, 2006.
6. Peter Rathmann. Dynamic Data Structures on Optical Disks. In 1st International Conference on Data Engineering, 1984.
7. Douglas J. Santry, Michael J. Feeley, Norman C. Hutchinson, and Alistair C. Veitch. Elephant: The file system that never forgets. In Workshop on Hot Topics in Operating Systems, pages 2–7, 1999.
8. Qingbo Zhu and Windsor W. Hsu. Fossilized index: the linchpin of trustworthy non-alterable electronic records. In SIGMOD ’05: Proceedings of the 2005 ACM SIGMOD international conference on Management of data, pages 395–406, New York, NY, USA, 2005. ACM Press.
9. W. Hsu and S. Ong. WORM Storage is not Enough. IBM Systems Journal, April 2007.
10. C. Johnson and T. Grandison. Compliance with data protection laws using hippocratic database active enforcement and auditing. IBM Systems Journal, 46(2), April 2007.
11. M. Mesnier, G. Ganger, and E. Riedel. Object-based storage: pushing more functionality into storage. IEEE Potentials, 24(2):31 – 34, April-May 2005.
12. S. Mitra and M. Winslett. Secure deletion from inverted indexes on compliance storage. In StorageSS ’06: Proceedings of the Second ACM Workshop on Storage Security and Survivability, pages 67–72, New York, NY, USA, 2006. ACM Press.
13. D. Reiner, G. Press, M. Lenaghan, D. Barta, and R. Urmston. Information lifecycle management: The EMC perspective. ICDE 2004.

Refs: Provenance (1)

1. U. Braun, S. Garfinkel, D. Holland, K.-K. Muniswamy-Reddy, and M. Seltzer. Issues in automatic provenance collection. In Proceedings of the International Provenance and Annotation Workshop, pages 171–183, 2006.
2. M. Mesnier, G. Ganger, and E. Riedel. Object-based storage: pushing more functionality into storage. IEEE Potentials, 24(2):31 – 34, April-May 2005.
3. Y. Simmhan, B. Plale, and D. Gannon. A survey of data provenance in e-science. SIGMOD Rec., 34(3):31–36, September 2005.
4. R. S. Barga and L. A. Digiampietri. Automatic generation of workflow provenance. In Proceedings of the International Provenance and Annotation Workshop (IPAW), pages 1–9, 2006.
5. U. Braun, S. L. Garfinkel, D. A. Holland, K.-K. Muniswamy-Reddy, and M. I. Seltzer. Issues in automatic provenance collection. In Proceedings of the International Provenance and Annotation Workshop (IPAW), pages 171–183, 2006.
6. P. Buneman, A. Chapman, and J. Cheney. Provenance management in curated databases. In SIGMOD '06: Proceedings of the 2006 ACM SIGMOD international conference on Management of data, pages 539–550, New York, NY, USA, 2006. ACM Press.
7. P. Buneman, A. Chapman, J. Cheney, and S. Vansummeren. A provenance model for manually curated data. In Proceedings of the International Provenance and Annotation Workshop (IPAW), pages 162–170, 2006.
8. P. Buneman, S. Khanna, and W. C. Tan. Data provenance: Some basic issues. In FST TCS 2000: Proceedings of the 20th Conference on Foundations of Software Technology and Theoretical Computer Science, pages 87–93, London, UK, 2000. Springer-Verlag.
9. P. Buneman, S. Khanna, and W. C. Tan. Why and where: A characterization of data provenance. Lecture Notes in Computer Science, 1973:316–330, 2001.
10. B. W. Dearstyne. The archival enterprise: Modern archival principles, practices, and management techniques. American Library Association, 1993.
11. E. Deelman, G. Singh, M. Atkinson, A. Chervenak, N. C. Hong, C. Kesselman, S. Patil, L. Pearlman, , and M. Su. Grid-based metadata services. In SSDBM '04: Proceedings of the 16th International Conference on Scientific and Statistical Database Management (SSDBM'04), page 393, Washington, DC, USA, 2004. IEEE Computer Society.
12. I. T. Foster, J. Vockler, M. Wilde, and Y. Zhao. Chimera: A virtual data system for representing, querying, and automating data derivation. In SSDBM '02: Proceedings of the 14th International Conference on Scientific and Statistical Database Management, pages 37–46, Washington, DC, USA, 2002. IEEE Computer Society.

Refs: Provenance (2)

13. C. Goble. Position statement: Musings on provenance, workflow workflow and (semantic web) annotations for bioinformatics. In Workshop on Data Derivation and Provenance, Chicago, 2002.
14. J. Golbeck. Combining provenance with trust in social networks for semantic web content filtering. In Proceedings of the International Provenance and Annotation Workshop (IPAW), pages 101–108, 2006.
15. R. Hasan, S. Myagmar, A. J. Lee, and W. Yurcik. Toward a threat model for storage systems. In Proceedings of the first ACM workshop on Storage security and survivability (StorageSS), pages 94–102, Fairfax, VA, USA, 2005. ACM Press.
16. C. A. Lynch. When documents deceive: Trust and provenance as new factors for information retrieval in a tangled web. Journal of the American Society for Information Science and Technology, 52(1):12–17, 2001.
17. K.-K. Muniswamy-Reddy, D. A. Holland, U. Braun, and M. I. Seltzer. Provenance-aware storage systems. In USENIX Annual Technical Conference, General Track, pages 43–56, 2006.
18. J. D. Myers, T. C. Allison, S. Bittner, B. Didier, M. Frenklach, J. William H. Green, Y.-L. Ho, J. Hewson, W. Koegler, C. Lansing, D. Leahy, M. Lee, R. McCoy, M. Minkoff, S. Nijsure, G. von Laszewski, D. Montoya, C. Pancerella, R. Pinzon, W. Pitz, L. A. Rahn, B. Ruscic, K. Schuchardt, E. Stephan, A. Wagner, T. Windus, and C. Yang. A collaborative informatics infrastructure for multi-scale science. clade, 00:24, 2004.
19. C. Sar and P. Cao. Lineage file system. Online at http://crypto.stanford.edu/ cao/lineage.html, January 2005.
20. Y. L. Simmhan, B. Plale, and D. Gannon. A survey of data provenance in e-science. SIGMOD Rec., 34(3):31–36, September 2005.
21. M. Szomszor and L. Moreau. Recording and reasoning over data provenance in web and grid services. In International Conference on Ontologies, Databases and Applications of SEmantics (ODBASE), volume 2888 of Lecture Notes in Computer Science, pages 603–620, Catania, Italy, 2003.
22. N. N. Vijayakumar and B. Plale. Towards low overhead provenance tracking in near real-time stream filtering. In Proceedings of the International Provenance and Annotation Workshop (IPAW), pages 46–54, 2006.
23. J. Widom. Trio: A system for integrated management of data, accuracy, and lineage. In Proceedings of the Second Biennial Conference on Innovative Data Systems Research (CIDR), January 2005.
24. J. Zhao, C. A. Goble, R. Stevens, and S. Bechhofer. Semantically linking and browsing provenance logs for e-science. In ICSNW, pages 158–176, 2004.