# Database Forensics in the Service of Information Accountability

Kyriacos E Pavlou
Doctoral Adviser: Prof. Richard T Snodgrass
The University of Arizona Department of Computer Science

## Motivation

### Motivation

Corporate abuses by Enron and WorldCom have given rise to recent regulations which require many corporations to ensure trustworthy long-term retention of their routine business documents.

- Health Insurance Portability and Accountability Act: HIPAA (1996)
- Sarbanes-Oxley Act (2002)
- U.S. Food and Drug Administration regulation "21 CFR Part 11" (2003)

Due to widespread news coverage of collusion between auditors and the companies they audit, and a lack of tools to address such corruption, there has been interest within the file systems and database communities in built-in mechanisms to detect or even prevent tampering.

*Compliant records* are those required by law to follow certain "processes by which they are created, stored, accessed, maintained, and retained." It is common to use Write-Once-Read-Many (WORM) storage devices to preserve such records.

### Information Accountability vs Restriction

*Information restriction* entails rendering retained records immutable and controlling access to them. This approach appears to be the prevailing viewpoint for achieving privacy and security.

*Information accountability* assumes that information should be transparent so as to easily determine whether a particular use is appropriate under a given set of rules.


Information accountability has been tried and tested successfully since ancient times.

Fig. 1. Modern Tamper-Indicating Seals *(left)*. *Bulla*, 14th c. Byzantium *(top)*. *American Scientist*, 94(6):515–524, Nov–Dec 2006.

**Fair Credit Reporting Act**



## Thesis Statement

*A shift towards information accountability presents valuable advantages over information restriction in the particular area of correct storage, use, and maintenance of databases.*

*An information accountability approach to database security is cheaper, can protect against a variety of threats (including insider threats), can successfully deal with the aftermath of information restriction failure and can render complex security problems tractable.*

*We are working to show information accountability can effectively realize appropriate use (i.e., guarantee no unauthorized modifications—insertions, deletions, updates) in high-performance databases.*

We will achieve this by:
- developing a tamper detection approach,
- accommodating shredding and litigation holds,
- developing a taxonomy of corruption types,
- designing forensic analysis algorithms and associated techniques, and
- implementing and evaluating a prototype system.

Funded by NSF grants IIS-0415101 and IIS-0803229 and a grant from Surety, LLC.

## DRAGOON

### Database foRensic Analysis safeGuard Of arizONa

DRAGOON is a prototype *continuous assurance* auditing system that is highly customizable in terms of offering a tunable trade-off between level of security and forensic cost. A beta version of DRAGOON is already available at:
`http://www.cs.arizona.edu/projects/tau/dragoon/`
It is lightweight and scalable and hence is able to adequately address aspects of information accountability.

We intend to expand our prototype to an enterprise-wide information accountability solution that can effectively realize appropriate use (i.e., guarantee no unauthorized modifications–insertions, deletions, updates even by insiders) in high-performance databases.

### Reference Architecture & Execution Phases

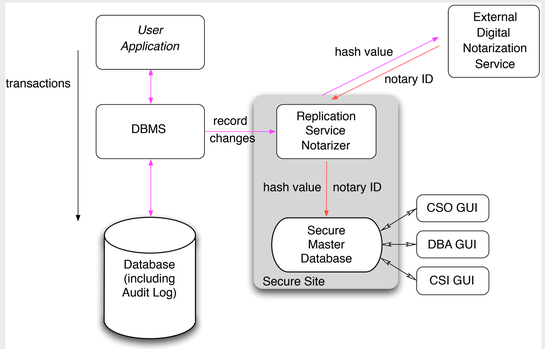



Fig. 2. The Normal Processing Phase.

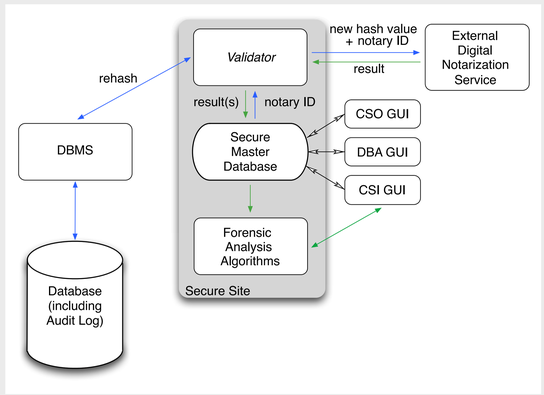



Fig. 3. The Tamper Detection and Forensic Analysis Phases.

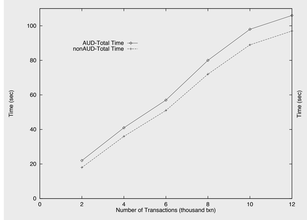



Fig. 4. Changing the Number of Transactions per Application.

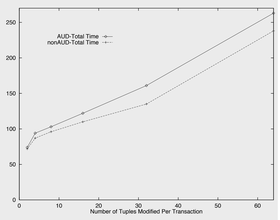



Fig. 5. Changing the Transaction Size.

## Forensic Analysis Algorithms

### Corruption Diagrams

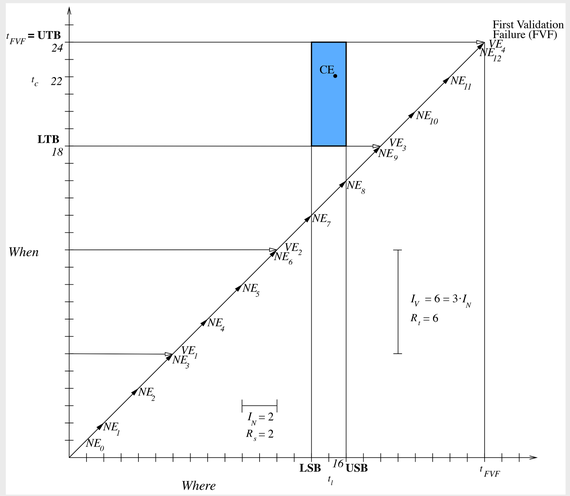



Fig. 6. The Monochromatic Forensic Analysis Algorithm.

| Symbol | Name | Definition |
|---|---|---|
| CE | Corruption event | An event that compromises the database |
| $VE$ | Validation event | The validation of the audit log by the notarization service |
| $NE$ | Notarization event | The notarization of a document (hash value) by the notarization service |
| $I_V$ | Validation interval | The time between two successive $VE$s |
| $I_N$ | Notarization interval | The time between two successive $NE$s |
| $R_t$ | Temporal detection resolution | Finest granularity chosen to express temporal bounds uncertainty of a CE |
| $R_s$ | Spatial detection resolution | Finest granularity chosen to express spatial bounds uncertainty of a CE |
| $t_{FVF}$ | Time of first validation failure | Time instant at which the CE is first detected |
| $USB$ | Upper spatial bound | Upper bound of the spatial uncertainty of the corruption region |
| $LSB$ | Lower spatial bound | Lower bound of the spatial uncertainty of the corruption region |
| $UTB$ | Upper temporal bound | Upper bound of the temporal uncertainty of the corruption region |
| $LTB$ | Lower temporal bound | Lower bound of the temporal uncertainty of the corruption region |

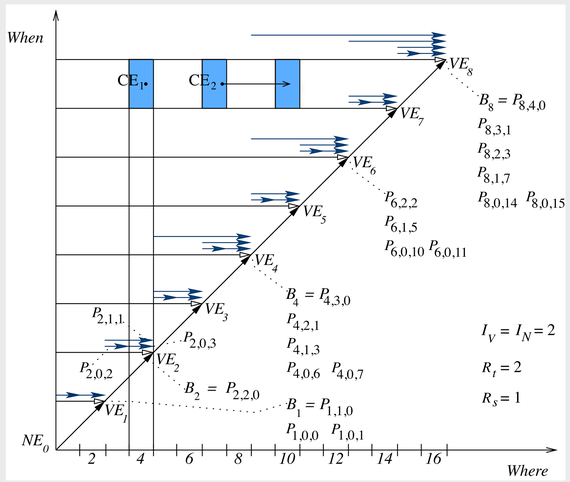
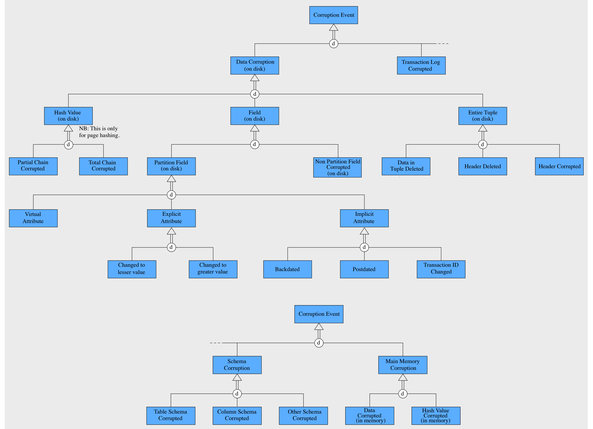



Fig. 7. The a3D Forensic Analysis Algorithm.

## Taxonomy of Corruption Types




## Contributions

This conceptual framework on information accountability architecture with advanced capabilities, forensic analysis tools and their evaluation will be extremely valuable and applicable to a variety of sectors. They can:

- ensure record compliance for financial and medical institutions,
- serve as an unbiased witness to databases storing sensitive information, e.g., court-submitted data from police databases,
- ensure non-deviation from standard operating procedures in biosciences labs (provenance of results),
- detect bugs silently corrupting databases,
- detect corruption shortly after tampering,
- automate some of the forensic work required in the aftermath of a database corruption saving both time and money,
- have advantages over information restriction approaches relying on hardware (prohibitive costs for small institutions, limited shelf-life, relatively complex), and
- mirror the relationship between the law and human behavior more closely.

## References

K. E. Pavlou and R. T. Snodgrass. Forensic Analysis of Database Tampering. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pages 109–120, June 2006.

K. E. Pavlou and R. T. Snodgrass. Forensic Analysis of Database Tampering. *ACM Transactions on Database Systems*, 33(4):1–47, November 2008.

K. E. Pavlou and R. T. Snodgrass. The Tiled Bitmap Forensic Analysis Algorithm. *IEEE Transactions on Knowledge and Data Engineering*, 22(4):590–601, April 2010.

R. T. Snodgrass, S. S. Yao, and C. Collberg. Tamper Detection in Audit Logs. In *Proceedings of the International Conference on Very Large Databases*, pages 504–515, September 2004.