A Fingerprinting Scheme for Relational Data Resistant to Collusion Attack

Daniel Cruz Paz
Advisor: PhD. Claudia Feregrino Uribe (INAOE)

Abstract

The use of relational databases as digital means for the storage and analysis of organizational data has increased significantly in the last years, originating the growth of threats that intend to infringe the data with illicit purposes, therefore it is essential to develop mechanisms that allow the ownership protection of these digital assets. In this work, we propose a robust fingerprinting scheme to collusion attacks for relational databases with the objective of providing a security mechanism for the identification of the guilty user against threats of illegal copying and distribution of unauthorized content.

Motivation

- Use of digital media to share information.
- Sale of databases and sensitive data used by outsourcing providers.
- The alteration and redistribution of digital media pose serious threat to both governmental security and commercial markets.
- Security mechanisms for the protection of digital media ownership.

Problem Statement

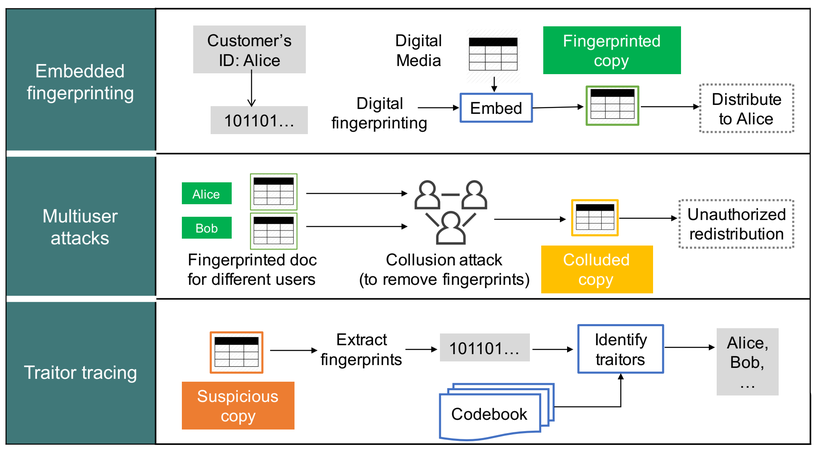


Fig. 1: Using embedded fingerprinting for tracing users, adapted from [11]

Watermarking Issues

- Usability
- Capacity
- Robustness

Related Work

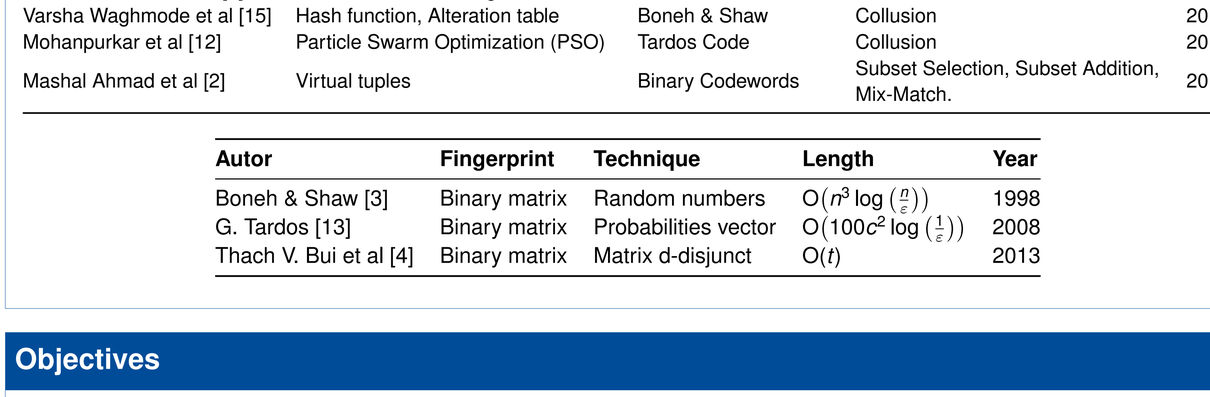
Autor	Techniques	Fingerprint	Attacks	Year
Agrawal & Kiernan [1]	Hash function and LSB	Binary Codewords		2002
Yingjiu Li et al [9]	MSB used as a virtual primary key	Binary Codewords		2003
Zhi & Xiao [16]	Image Based Watermarking (IBW)	Image	Subset Alteration, Subset Addition, Subset out-of-ordering	2004
Yingjiu Li et al [10]	Pseudorandom sequence generator	MD5 of serial number, Boneh & Shaw	Bit Flipping, Invertibility, Collusion	2005
Guo et al [5]	Twice-embedding with hash function	Binary Codewords	Subset Selection, Subset Addition, Subset Alteration	2006
Lafaye et al [8]	Pairing Algorithm	Tardos Code	Data Alteration, Subset Attacks, Collusion	2008
Uzun & Stephenson [14]	Permutations of attributes, Insertion of virtual tuples	Binary Codewords		2008
Thach V. Bui et al [4]	Anonymize attributes	Matrix d-disjunct	Collusion	2013
Saman Iftikhar et al [7]	Arithmetic operations	Secret signature	Subset Deletion, Subset Alteration, Collusion	2014
Namrata Gursale et al [6]	Twice Data Partitioning	Tardos Code	Collusion	2014
Varsha Waghmode et al [15]	Hash function, Alteration table	Boneh & Shaw	Collusion	2014
Mohanpurkar et al [12]	Particle Swarm Optimization (PSO)	Tardos Code	Collusion	2015
Mashal Ahmad et al [2]	Virtual tuples	Binary Codewords	Subset Selection, Subset Addition, Mix-Match.	2017

Autor	Fingerprint	Technique	Length	Year
Boneh & Shaw [3]	Binary matrix	Random numbers	$O(n^3 \log(\frac{n}{\varepsilon}))$	1998
G. Tardos [13]	Binary matrix	Probabilities vector	$O(100c^2 \log(\frac{1}{\varepsilon}))$	2008
Thach V. Bui et al [4]	Binary matrix	Matrix d-disjunct	$O(t)$	2013

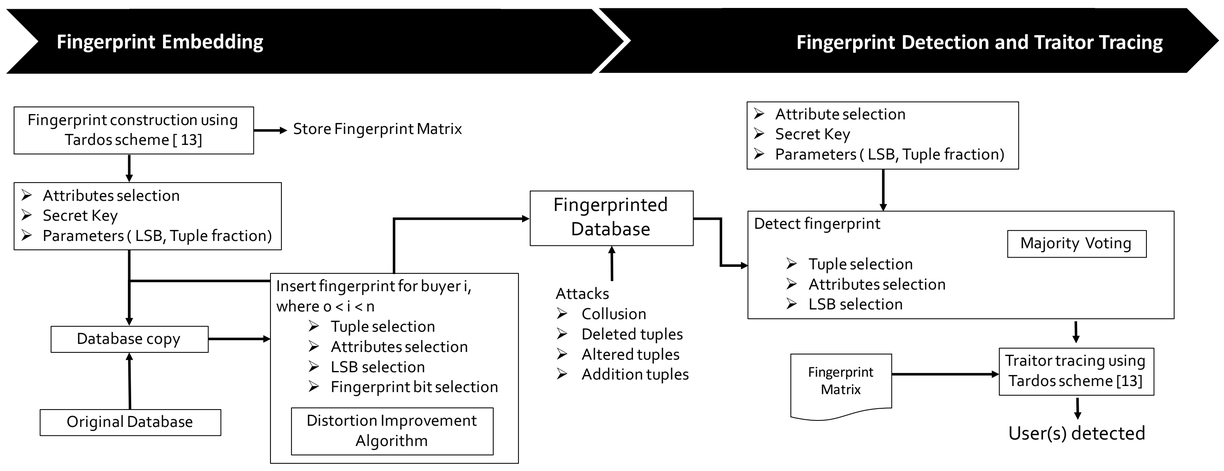
Objectives

- **Main objective**
  To design and implement a robust fingerprinting scheme to collusion attacks for protecting the ownership of relational databases vulnerable to illegal copying and distribution of unauthorized content.
- **Specific Objectives**
  - To define an adequate fingerprint scheme that have provided better results regarding capacity and robustness, which will be taken as a reference for the development of the proposed scheme.
  - To design an algorithm for fingerprint embedding and generation for relational databases with minimal data distortion and greater robustness to collusion attacks.
  - To design the detection and extraction fingerprint algorithms for the identification of the end user before the distribution of unauthorized content.
  - To evaluate the robustness of fingerprint embedding against collusion attacks and common operations to RDB.

Methodology



Proposed Solution



Preliminary Results

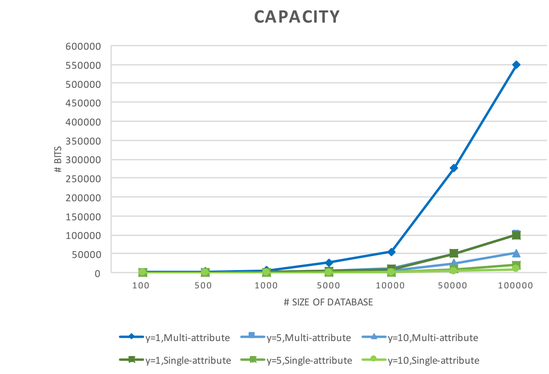


Fig. 2: Capacity analysis for different sizes of the database

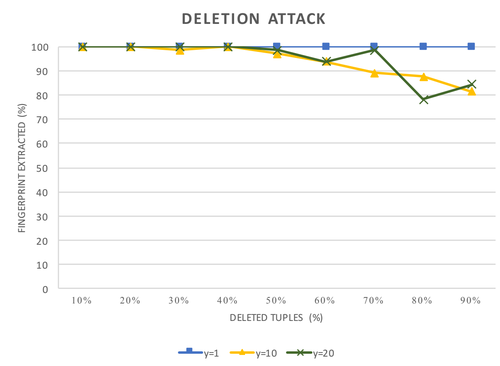


Fig. 3: Fingerprint recovery after a random tuple deletion attack

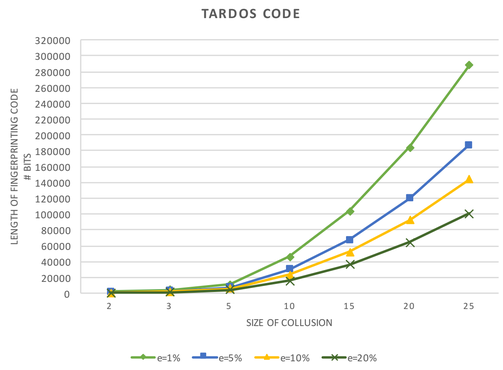


Fig. 4: Tardos code construction for N users

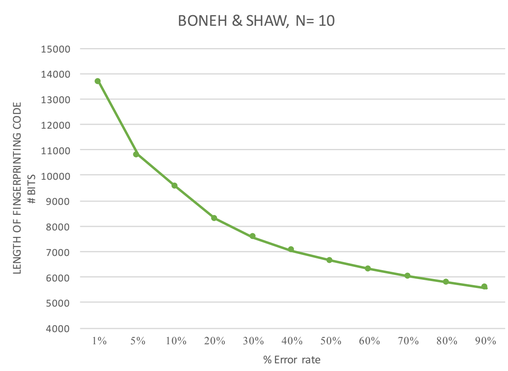


Fig. 5: Boneh & Shaw code construction with different error rate

Current and Future Work

- Base scheme implementation.
- Improvement on insertion capacity of the scheme.
- Distortion evaluation.
- Analysis of tuple deletion attack.
- Implementation of the Boneh & Shaw and G. Tardos scheme.
- Code evaluation and collusion attack.
- Insertion capacity analysis of the scheme.
- Incorporation of the Tardos scheme to the current fingerprinting scheme.

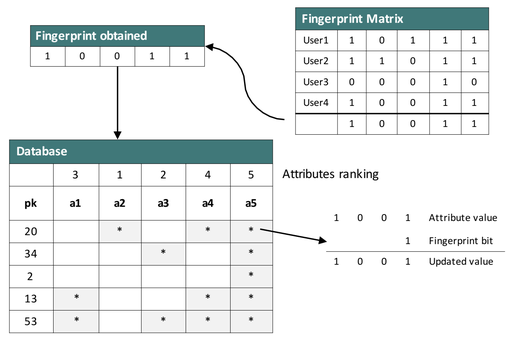
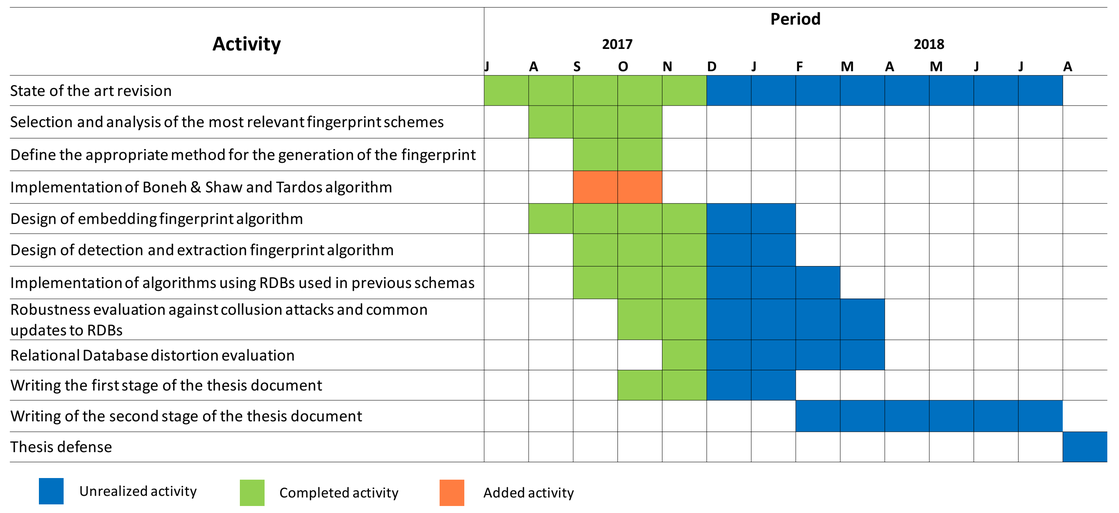


Fig. 6: Distortion Improvement Algorithm

Activity Schedule



References

[1] AGRAWAL, R., AND KIERNAN, J. Watermarking relational databases. In Proceedings of the 28th international conference on Very Large Data Bases (2002), VLDB Endowment, pp. 155–166.
[2] AHMAD, M., SHAHID, A., QADRI, M. Y., HUSSAIN, K., AND QADRI, N. N. Fingerprinting non-numeric datasets using row association and pattern generation. In Communication Technologies (ComTech), 2017 International Conference on (2017), IEEE, pp. 149–155.
[3] BONEH, D., AND SHAW, J. Collusion-secure fingerprinting for digital data. IEEE Transactions on Information Theory 44, 5 (1998), 1897–1905.
[4] BUI, T. V., NGUYEN, B. Q., NGUYEN, T. D., SONEHARA, N., AND ECHIZEN, I. Robust fingerprinting codes for database. In International Conference on Algorithms and Architectures for Parallel Processing (2013), Springer, pp. 167–176.
[5] GUO, F., WANG, J., AND LI, D. Fingerprinting relational databases. In Proceedings of the 2006 ACM symposium on Applied computing (2006), ACM, pp. 487–492.
[6] GURSALE, M. N. Ms. arti mohanpurkar,"a robust, distortion minimization fingerprinting technique for relational database". International Journal on Recent and Innovation Trends in Computing and Communication 2, 6 (2015), 1737–1741.
[7] IFTIKHAR, S., ANWAR, Z., AND KAMRAN, M. A novel and robust fingerprinting technique for digital data based on genetic algorithm. In High-capacity Optical Networks and Emerging/Enabling Technologies (HONET), 2014 11th Annual (2014), IEEE, pp. 173–177
[8] LAFAYE, J., GROSS-AMBLARD, D., CONSTANTIN, C., AND GUERROUANI, M. Watermill: An optimized fingerprinting system for databases under constraints. IEEE Transactions on Knowledge and Data Engineering 20, 4 (2008), 532–546.
[9] LI, Y., SWARUP, V., AND JAJODIA, S. Constructing a virtual primary key for fingerprinting relational data. In Proceedings of the 3rd ACM workshop on Digital rights management (2003), ACM, pp. 133–141.
[10] LI, Y., SWARUP, V., AND JAJODIA, S. Fingerprinting relational databases: Schemes and specialties. IEEE Transactions on Dependable and Secure Computing 2, 1 (2005), 34–45.
[11] LIU, K. R. Multimedia fingerprinting forensics for traitor tracing, vol. 4. Hindawi Publishing Corporation, 2005.
[12] MOHANPURKAR, A., AND JOSHI, M. A fingerprinting technique for numeric relational databases with distortion minimization. In Computing Communication Control and Automation (ICCUBEA), 2015 International Conference on (2015), IEEE, pp. 655–660.
[13] TARDOS, G. Optimal probabilistic fingerprint codes. Journal of the ACM (JACM) 55, 2 (2008), 10.
[14] UZUN, E., AND STEPHENSON, B. Security of relational databases in business outsourcing. HP Labs Report HPL-2008-168 (2008).
[15] WAGHMODE, M. V., AND MOHANPURKAR, M. A. Collusion avoidance in fingerprinting outsourced relational databases with knowledge preservation. International Journal on Recent and Innovation Trends in Computing and Communication 2, 5 (2014), 1332–1337.
[16] ZHANG, Z.-H., JIN, X.-M., WANG, J.-M., AND LI, D.-Y. Watermarking relational database using image. In Machine Learning and Cybernetics, 2004. Proceedings of 2004 International Conference on (2004), vol. 3, IEEE, pp. 1739–1744.