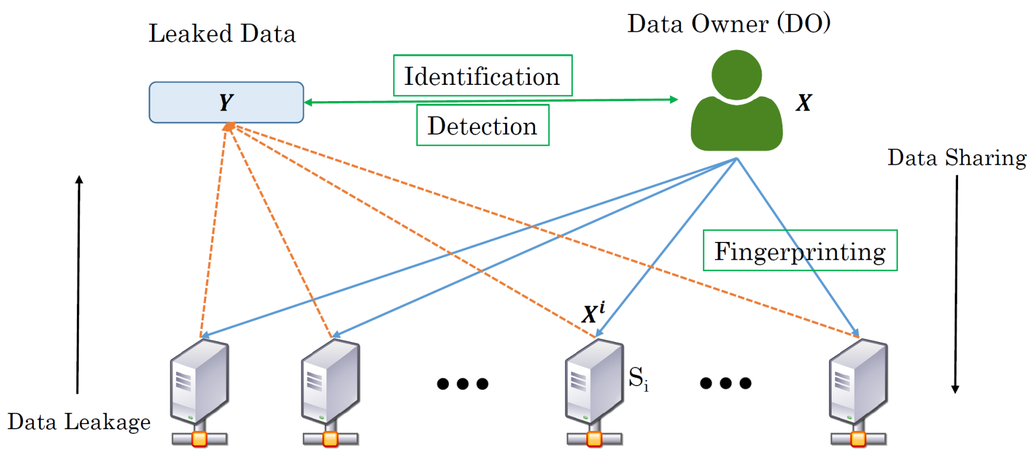
Robust, Privacy- and Utility-Preserving Fingerprinting Schemes for Correlated Data

PI: Erman Ayday (Case Western Reserve University), Co-PI: Emre Yilmaz (University of Houston - Downtown)

https://engineering.case.edu/research/labs/SPiD/projects/p2_satc

Challenge:

- Data owners share sensitive information with a wide-range of service providers (SPs). While doing so, they hope that (i) SPs will comply with the data usage agreements and not engage in unauthorized sharing of their data (ii) the privacy of their data is preserved, and (iii) the utility of the shared data is high
- Fingerprinting is a well-known technique to detect unauthorized distribution of sensitive data
- Existing fingerprinting techniques, mainly the ones for multimedia, rely on the high amount of redundancy in the data and they do not consider a wide-variety of attacks
- Simultaneously providing robust fingerprinting, privacy guarantees, and high data utility has unique challenges
- New research is needed to focus on algorithms which meet these challenges in the face of such robustness, data privacy, and data utility and yet maintain computational efficiency



Proposed system model

Solution:

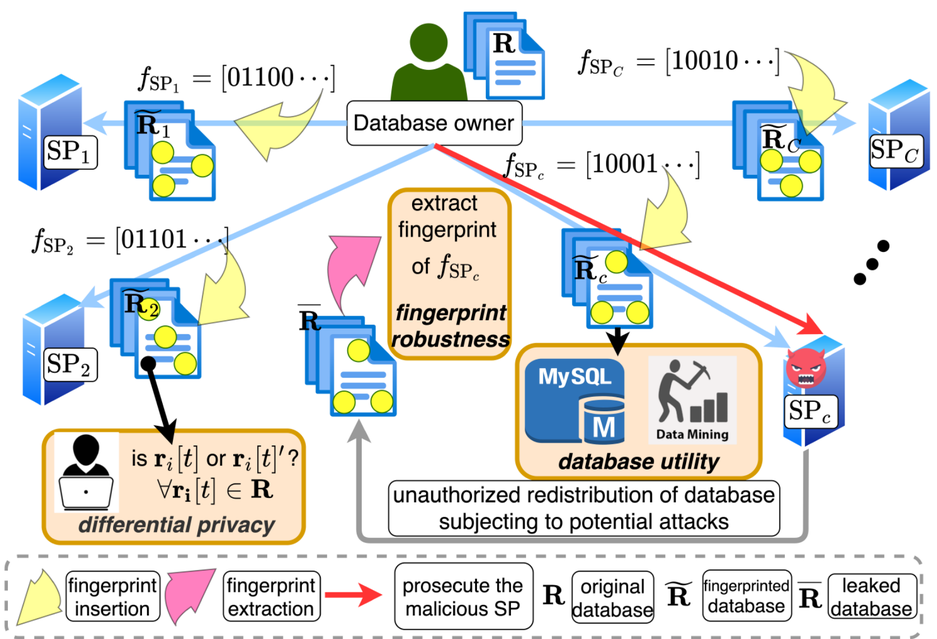
- We develop probabilistic fingerprinting schemes for correlated data that detect (and hence prevent) unauthorized sharing of data
- Developed schemes are robust against collusion and correlation attacks against a fingerprinting scheme
- We integrate our efforts on privacy-preserving data sharing into the proposed fingerprinting algorithms in order to address liability and privacy together
- For the first time, we propose differentially-private fingerprinting schemes
- We develop algorithms to find the optimal order of data

Impact on Society:

- We provide tools that identify the sources of unauthorized data leakages with high probability. This will deter malicious service providers (SPs) from unauthorized sharing of their users' data
- Data owners, knowing they have stronger control on how their data will be used and shared, will be more willing to share their data with the SPs

Scientific Impact:

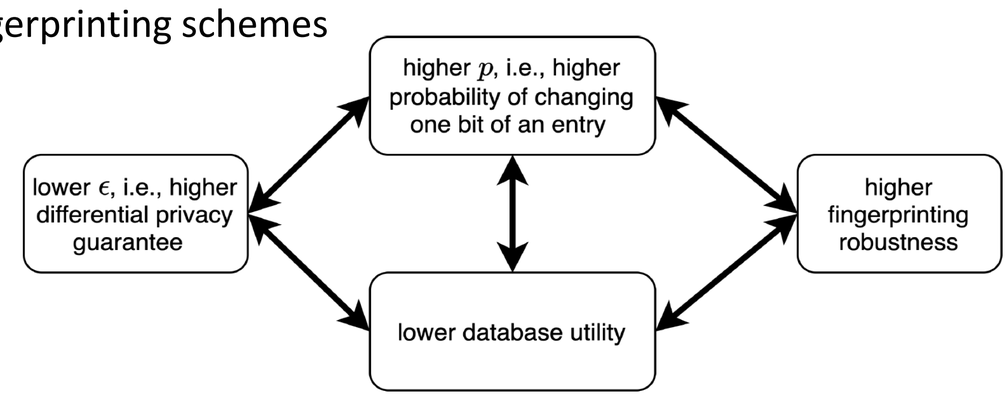
- We show the vulnerability of existing fingerprinting schemes to various attacks, especially the ones exploiting the correlations in the data
- We develop probabilistic fingerprinting algorithms that provide robustness against a wide-variety of attacks
- For the first time, we provide privacy guarantees in a fingerprinting algorithm
- We develop the proposed algorithms for sharing of personal correlated data, databases, and graphs
- We show the applications of the proposed algorithms for different domains, such as medical data, location data, financial data, and demographic information



Differentially-private fingerprinting

Solution:

- We develop different algorithms for sharing of different types of information: personal data, databases, and graph data
- Developed schemes can be used as an add-on to existing fingerprinting schemes



Relationship between fingerprint robustness, privacy, and data utility

Education and Outreach:

- We train graduate, undergraduate, and high school students in this project
- We are committed to recruiting women and minorities. This commitment will also continue throughout this project

Potential Impact:

Technologies resulting from this research that simultaneously consider data liability, privacy, and utility will be instrumental for various data sharing platforms

Video

References

1. T. Ji, E. Yilmaz, E. Ayday, and P.Li, "The Curse of Correlations for Robust Fingerprinting of Relational Databases," Proceedings of International Symposium on Research in Attacks, Intrusions and Defenses (RAID), 2021.
2. T. Ji, E. Ayday, E. Yilmaz, and P. Li. Robust Fingerprinting of Genomic Databases. 30th International Conference on Intelligent Systems for Molecular Biology (ISMB), 2022.
3. T. Ji, E. Ayday, E. Yilmaz, and P. Li. Differentially-private fingerprinting of relational databases. arXiv:2109.02768, 2021.
4. A. C. Oksuz, E. Ayday, and U. Gudukbay "Privacy-Preserving and Robust Watermarking on Sequential Genome Data using Belief Propagation and Local Differential Privacy", Bioinformatics, 2021.
5. E. Ayday, E. Yilmaz, and A. Yilmaz, "Robust optimization-based watermarking scheme for sequential data", Proceedings of International Symposium on Research in Attacks, Intrusions and Defenses (RAID), 2019.