Nomination of Paper “Towards Robust Fingerprinting of Relational Databases by Mitigating Correlation Attacks” in IEEE TDSC’2022

I would like to nominate the paper titled “*Towards Robust Fingerprinting of Relational Databases by Mitigating Correlation Attacks*” published in IEEE TDSC 2022.

Database fingerprinting is steganographic technique that prevents unauthorized redistribution of shared databases. Fingerprinting schemes insert binary bit-string (customized for each database recipient) into databases before sharing. If there is a data leakage or a pirated copy, the database owner can extract the binary bit-string from the targeted copy and hold the traitors (malicious database recipients) responsible. Existing database fingerprinting schemes are robust against attacks like randomly changing database entries, removing rows or columns of database, and adding new data records. For example, even if about 80% entries are modified/deleted in a fingerprinted database, the database owner can still extract the inserted fingerprint bit-string.

The nominated TDSC paper proposed much more powerful attacks (called the correlation attacks) that can successfully compromise the inserted fingerprint bit-string, generate pirate copies, and mislead the database owner to accuse innocent database recipients. For example, by only modifying around 14% entries, such attack can destroy more than 65% of the fingerprint bit-string and cause more than 90% innocent recipients be mistakenly accused. Correlation attacks leverage the intrinsic correlations among the data records and the attributes to determine the potentially fingerprinted locations with very high probability.

To defend against the powerful correlation attacks, the nominated paper also developed mitigation techniques, which just change a few unfingerprinted entries to make the resulting database has similar correlation models before and after fingerprint insertion. The mitigation techniques can serve as an adds-on for all off-the-shelf database fingerprinting schemes to make them also robust against the correlation attacks. Besides, the mitigation techniques are also prior knowledge insensitive, which means that if the attackers use accurate correlation models (e.g., the joint distributions directly calculated from the original database) to launch the attacks, and the defender (database owner) uses inaccurate correlation models, the correlation attacks can still be mitigated effectively.

The techniques proposed in this TDSC paper can also be applied to protect other relational database in the course of data sharing. For example, human genomic database contains stronger correlation models due to kinship among family members and correlation among genomic entries at different locations. The techniques (developed in the nominated paper) have shown its success in the genomic domain and lead to a publication (titled “*Robust Fingerprinting of Genomic Databases*”) in one of the most prestigious venues in computational biology (i.e., *ISMB’2022*, 30th International Conference on Intelligent Systems for Molecular Biology). Moreover, it also points out the possibility of combing database fingerprinting with data privacy and leads to a publication (titled “*Privacy-Preserving Database Fingerprinting*”) in the 30th Network and Distributed System Security Symposium (*NDSS'2023*).

In conclusion, I highly recommend “*Towards Robust Fingerprinting of Relational Databases by Mitigating Correlation Attacks*” for NSA's best science of cybersecurity paper award.