Poster: A Framework Design for Privacy-Preserving Computation on Shared Data

Saffija Kasem-Madani (PhD student)
Rheinische Friedrich-Wilhelms-Universität Bonn
IT Security Working Group
Institute of Computer Science
Friedrich-Ebert-Allee 144, 53113 Bonn
Email: kasem@cs.uni-bonn.de

***Abstract*—In times of surveillance and data retention, sharing information among multiple parties often comes together with privacy concerns. To profit from the advantages of information sharing, including the analysis of data aggregated from multiple parties, there is need for a mechanism that ensures both the analyzability of data and the preservation of the data owners' privacy. We present an information sharing framework design that allows to perform homomorphic computations on encrypted shared data. The process of decryption makes the result of the data analysis available to the data analyzer and ensures the confidentiality of the shared information. It thus meets both the privacy and the utility requirements of the information sharing parties and can be used for offline data analysis.**

When information is shared among multiple parties for data analysis, the objective of the data holders in keeping cleartexts confidential for privacy collide with the objective of the data receiver in the highest achievable utility. Consider the following application scenario: Data holders want to share log files (e.g., see [1]) with a centralized analysis entity, called data analyzer. The data holders are interested in the results of the analysis performed by the analysis entity. The analysis entity profits from collecting log files from multiple sources because it gives him a broader view on the field, e.g. on the current situation regarding certain attacks in a business area. On the other hand, the data holders want to keep their data confidential. One reason is the privacy concerns of the data owners. This contradicts the data analyzer's interest in data utility. The described contradiction may prevent different parties from information sharing and hence, decreases the chances to profit from the analysis of data collected from multiple parties.

In this work, our goal is to overcome the described problem. We propose a solution that meets both the privacy and utility needs of the information sharing parties. We assume that the data holders and the data analyzer agrees on the utility required for data analysis in mutual agreements called *policies*. A policy is a set of conditions to be fulfilled by the exchanged data in order to meet all the utility options satisfying the analyzer's and the holders' requirements. This may include the ability to calculate overall sum of numerical data for statistical analysis, or to disclose IP addresses in a log file under certain conditions, or to check encrypted data for equality. The conditions stated in a policy must be selected very carefully to ensure that the privacy of the data owners will be preserved. For formulating such policies, we refer to [2].

Before sharing the data with the analyzer, the data holder must transform the data to a *data appearance* that meets the formulated policy. Depending on the use case and the sensitivity of the data under consideration, the selection of the provided utility must keep most of the information confidential so that an attacker would not be able to use the utility properties of the data appearances for attacks utilizing information linkage or correlation with external information.

In order to solve the problem described above, one approach is the utilization of homomorphic encryption. In homomorphic encryption, the cleartext data p_1, p_2 are encrypted with an encryption mechanism E to $E(p_1), E(p_2)$. E ensures that certain operations can be performed on $E(p_1), E(p_2)$ and are equivalent to operating on the plaintexts p_1 and p_2. That means, $E(p_1) * E(p_2) = E(p_1 + p_2)$ for appropriate operations $*$ and $+$. Note that the result of the encrypted computation is encrypted.

In the above-mentioned application scenario, the data holder would use an appropriate homomorphic encryption mechanism to encrypt the cleartext content of the log files to data appearances. Depending on the selected homomorphic encryption mechanism, this would allow the analyzer to perform certain operations on the received encrypted data. To disclose the result of an operation to the analyzer, it must be decrypted. This usually requires knowledge of the secret (private) key and comes together with the fact that having access to the private key enables the data analyzer to decrypt the content of the data appearance and hence, violating the privacy of the data owners.

In this work, we describe a framework design that ensures that only results of allowed computations can be disclosed (i.e. decrypted) to the analyzer. Our contributions are as follows:

1) **Information sharing framework:** We present a framework for sharing information in data appearances with a centralized analysis entity that can perform homomorphic computations on the encrypted data. For the reasons described above, the tasks required for the computation and decryption are distributed among five stakeholders: the data analyzer, the preprocessor, the activator, the private-key holder and the decryptor.

2) **Protocol for distribution of data:** We describe a protocol for data-appearance and key distribution.
3) **Protocol for privacy-preserving decryption:** We describe a protocol for the privacy-preserving decryption of computation results.

In the current design of the framework, we assume the honest-but-curious adversarial model [3]. We assume that for very sensitive data, e.g. healthcare data, stronger assumptions are required. This may be achieved by combining our approach with secret sharing mechanisms [4].

We describe the parties of the framework and the design assumptions made for ensuring the privacy preservation of the shared data. The framework is based on Paillier, a public-key probabilistic homomorphic encryption scheme [5]. With the prospected upcoming improvement of homomorphic encryption schemes[1], we aim at providing a feasible solution for the use of homomorphic computation methods for privacy preserving information sharing.

REFERENCES

[1] A. J. Slagell and W. Yurcik, "Sharing Computer Network Logs for Security and Privacy: A Motivation for New Methodologies of Anonymization," *CoRR*, vol. cs.CR/0409005, 2004. [Online]. Available: http://dblp.uni-trier.de/db/journals/corr/corr0409.htmlcs-CR-0409005

[2] S. Kasem-Madani and M. Meier, "Definition of Availability Policies for Data Pseudonymization Using XACML," Aug. 2015, IFIP Summer School on Privacy and Identity Management. [Online]. Available: http://www.ifip-summerschool.org/wp-content/uploads/2015/08/Summer-School-2015-PreProceedings.pdf

[3] A. Paverd, A. Martin, and I. Brown, "Modelling and Automatically Analysing Privacy Properties for Honest-but-Curious Adversaries," Tech. Rep. [Online]. Available: https://www.cs.ox.ac.uk/people/andrew.paverd/casper/casper-privacy-report.pdf

[4] A. Shamir, "How to share a secret," *Commun. ACM*, vol. 22, no. 11, pp. 612–613, Nov. 1979. [Online]. Available: http://doi.acm.org/10.1145/359168.359176

[5] P. Paillier, "Public-Key Cryptosystems Based on Composite Degree Residuosity Classes," in *Advances in Cryptology, EUROCRYPT 99*, ser. Lecture Notes in Computer Science, J. Stern, Ed. Springer Berlin Heidelberg, 1999, vol. 1592, pp. 223–238. [Online]. Available: http://dx.doi.org/10.1007/3-540-48910-$X_1$6

[1] See e.g. the HEAT project https://heat-project.eu/