Recent Advances in Information-Theoretically Secure Data Outsourcing*

Sharad Mehrotra and Shantanu Sharma
University of California, Irvine, USA.

ABSTRACT

Despite extensive research on cryptography, secure and efficient query processing over outsourced data remains an open challenge. This tutorial focuses on secret-sharing-based techniques that allow data outsourcing and provides a review of recent secret-sharing-based techniques based on the security they offer (revealing or hiding access-patterns). Then, we discuss database systems developed using such techniques and, particularly, discuss a recent secret-sharing based system, namely Obscure.

ACM Reference Format:
Sharad Mehrotra and Shantanu Sharma. 2020. Recent Advances in Information-Theoretically Secure Data Outsourcing. In *Proceedings of the Sixth International Workshop on Security and Privacy Analytics (IWSPA '20), March 18, 2020, New Orleans, LA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3375708.3382736

OVERVIEW OF THE TUTORIAL

Over the last two decades, several cryptographic techniques (*e.g.*, [2, 11, 12, 14–16]) have been developed for outsourcing the data to untrusted servers. These techniques may be broadly classified based on cryptographic security into two categories:

Computationally secure techniques that assume the adversary lacks adequate computational capabilities to break the underlying cryptographic mechanism in polynomial (practical amount of) time. Example of such techniques are non-deterministic encryption [12], homomorphic encryption [11], order-preserving encryption [2], and searchable-encryption [16].

Information-theoretically secure techniques that are unconditionally secure and independent of the adversary's computational capabilities. Shamir's secret-sharing (SSS) [15] is a well-known information-theoretically secure protocol. In SSS, multiple (secure) shares of a dataset are kept at mutually suspicious servers, such that a single server cannot learn anything about the data. Secret-sharing-based techniques are secure under the assumption that a majority of the servers (equal to the threshold of the secret-sharing mechanism) do not collude. Secret-sharing mechanisms also have applications in other areas such as Byzantine agreement, secure multiparty computations (MPC), and threshold cryptography, as discussed in [5].

While much of secure data outsourcing techniques have been built around computationally secure cryptographic mechanisms, recent works, both in academia and industries/startups have begun to explore information-theoretically secure techniques. In this tutorial, we classify the secret-sharing-based techniques/systems, based on the security they provide in terms of hiding or revealing access-patterns (*i.e.*, the identity of the tuples that satisfy the query), as follows:

- **Access-pattern revealing systems**: are those that reveal the identity of the tuples that satisfy the query. These systems require the database owner to significantly involve in executing a query. In particular, for answering a query, the database owner needs to always retain all the polynomials that were used at the time of database outsourcing. When a query arrives at the database owner, they use the same polynomial for the query keyword that was used when outsourcing the data. Systems such as [9, 10, 17] follow this technique. However, such techniques suffer from several drawbacks, *e.g.*, weak security guarantees such as leakage of access patterns, a significant overhead of maintaining polynomials for generating shares at the database owner, no support for third-party query execution on the secret-shared outsourced database.
- **Access-pattern hiding systems**: are those that completely hide the identity of the qualifying tuples that satisfy a query. Recently, [7] proposed an access-pattern hiding technique when searching over the secret-shared data, while also not overburdening the database owner by retaining all polynomials that were used to outsource the database. Such a technique is used to develop secret-sharing-based systems, such as Obscure [13] for supporting verifiable aggregation queries on secret-shared data, [8] for supporting selection and join queries on secret-shared data, and [4] for searching and count queries.

Secret-sharing-based techniques have not only been used by academia to develop secure systems, but several industries/startups are also focusing on secret-sharing-based database systems. Examples of such industries/startups are Jana [3] by Galois, Pulsar [1] by Stealth Software, Sharemind [6] by Cybernetica, Unbound Tech., Partisia, Secret Double Octopus, SecretSkyDB Ltd. We will discuss these industry-based systems.

OBSCURE

Obscure [13] is a secret-sharing based system and provides communication-efficient and information-theoretically secure algorithms for aggregation queries. Obscure comes with several advantages, as follows:

*This material is based on research sponsored by DARPA under agreement number FA8750-16-2-0021. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA or the U.S. Government. This work is partially supported by NSF grants 1527536 and 1545071.

IWSPA '20, March 18, 2020, New Orleans, LA, USA
© 2020 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-7115-5/20/03.
https://doi.org/10.1145/3375708.3382736

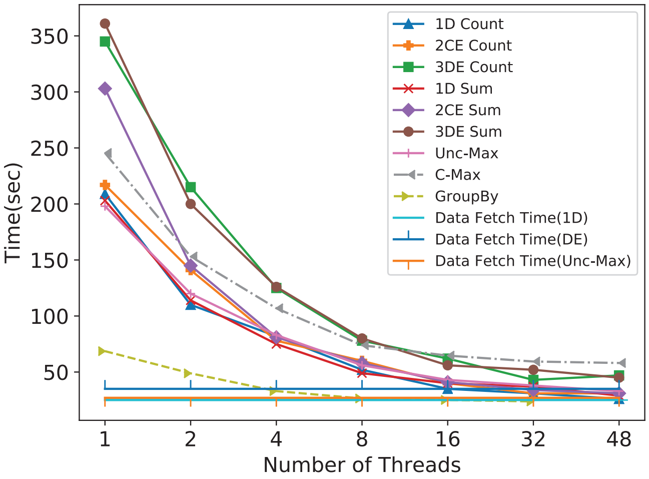

Figure 1: Obscure performance on 6M rows.

(1) Deals with honest but curious, as well as, malicious adversaries (which could deviate from the algorithm and delete tuples from the relation).
(2) Does not overburden the database owner by storing enough data related to polynomials and fully participating in query execution.
(3) Does not reveal access-patterns, while supporting selection predicate search over secret-shared data.
(4) Uses minimal communication rounds between the user and each server, (when having enough shares). Specifically, count, sum, average, and their verification algorithms require at most two rounds between each server and the user. However, maximum/minimum finding algorithms require at most four communication rounds. Also, Obscure achieves the minimum communication cost for aggregate queries, especially, for count, sum, and average queries, by aggregating data locally at each server.
(5) Neither involves the database owner to verify the results nor requires a trusted-third-party verifier.

The experimental results of Obscure show its practicality over a moderate-sized dataset. Particularly, several aggregation queries (such as count, sum, unconditional and conditional maximum, and group-by queries) using Obscure were evaluated on four columns (Orderkey, Partkey, Linenumber, and Suppkey) of 6M rows of LineItem table of TPC-H benchmark, with the help of AWS servers of 144GB RAM, 3.0GHz Intel Xeon CPU with 72 cores. The evaluation result of Obscure is shown in Figure 1 for one-dimensional (1D) count/sum, two/three-dimensional conjunctive-equality (2CE/3CE) count/sum, and two/three-dimensional disjunctive-equality (2DE/3DE) count/sum, unconditional maximum, and group-by queries.

BIOGRAPHIES

Sharad Mehrotra received the PhD degree in computer science from the University of Texas, Austin, in 1993. He is currently a professor in the Department of Computer Science, University of California, Irvine. Previously, he was a professor with the University of Illinois at Urbana Champaign. He has received numerous awards and honors, including the 2011 SIGMOD Best Paper Award, 2007 DASFAA Best Paper Award, SIGMOD test of time award, 2012, DASFAA ten year best paper awards for 2013 and 2014, 1998 CAREER Award from the US National Science Foundation (NSF), and ACM ICMR best paper award for 2013. His primary research interests include the area of database management, distributed systems, secure databases, and Internet of Things.

Shantanu Sharma received his Ph.D. in Computer Science in 2016 from Ben-Gurion University, Israel. During his Ph.D., he worked with Prof. Shlomi Dolev and Prof. Jeffrey Ullman. He obtained his Master of Technology (M.Tech.) degree in Computer Science from National Institute of Technology, Kurukshetra, India, in 2011. He was awarded a gold medal for the first position in his M.Tech. degree. Currently, he is pursuing his Post Doc at the University of California, Irvine, USA, assisted by Prof. Sharad Mehrotra. His research interests include data security and privacy, building secure and privacy-preserving systems on sensor data for smart buildings, designing models for MapReduce computations, distributed algorithms, mobile computing, and wireless communication.

REFERENCES

[1] Stealth Pulsar, available at:http://www.stealthsoftwareinc.com/.
[2] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu. Order-preserving encryption for numeric data. In *Proceedings of the ACM SIGMOD International Conference on Management of Data, Paris, France, June 13-18, 2004*, pages 563–574. ACM, 2004.
[3] D. W. Archer, D. Bogdanov, Y. Lindell, L. Kamm, K. Nielsen, J. I. Pagter, N. P. Smart, and R. N. Wright. From keys to databases - real-world applications of secure multi-party computation. *IACR Cryptology ePrint Archive*, 2018:450, 2018.
[4] H. Avni, S. Dolev, N. Gilboa, and X. Li. SSSDB: database with private information search. In *Algorithmic Aspects of Cloud Computing - First International Workshop, ALGOCLOUD 2015, Patras, Greece, September 14-15, 2015. Revised Selected Papers*, pages 49–61, 2015.
[5] A. Beimel. Secret-sharing schemes: A survey. In *Coding and Cryptology - Third International Workshop, IWCC 2011, Qingdao, China, May 30-June 3, 2011. Proceedings*, pages 11–46, 2011.
[6] D. Bogdanov, S. Laur, and J. Willemson. Sharemind: A framework for fast privacy-preserving computations. In S. Jajodia and J. López, editors, *Computer Security - ESORICS 2008, 13th European Symposium on Research in Computer Security, Málaga, Spain, October 6-8, 2008. Proceedings*, volume 5283 of *Lecture Notes in Computer Science*, pages 192–206. Springer, 2008.
[7] S. Dolev, N. Gilboa, and X. Li. Accumulating automata and cascaded equations automata for communicationless information theoretically secure multi-party computation. *Theor. Comput. Sci.*, 795:81–99, 2019.
[8] S. Dolev, P. Gupta, Y. Li, S. Mehrotra, and S. Sharma. Privacy-preserving secret shared computations using mapreduce. *IEEE Transactions on Dependable and Secure Computing*, pages 1–1, 2019.
[9] F. Emekçi, D. Agrawal, A. El Abbadi, and A. Gulbeden. Privacy preserving query processing using third parties. In *Proceedings of the 22nd International Conference on Data Engineering, ICDE 2006, 3-8 April 2006, Atlanta, GA, USA*, page 27, 2006.
[10] F. Emekçi, A. Metwally, D. Agrawal, and A. El Abbadi. Dividing secrets to secure data outsourcing. *Inf. Sci.*, 263:198–210, 2014.
[11] C. Gentry. *A fully homomorphic encryption scheme.* PhD thesis, Stanford University, 2009.
[12] S. Goldwasser and S. Micali. Probabilistic encryption. *J. Comput. Syst. Sci.*, 28(2):270–299, 1984.
[13] P. Gupta, Y. Li, S. Mehrotra, N. Panwar, S. Sharma, and S. Almanee. Obscure: Information-theoretic oblivious and verifiable aggregation queries. *PVLDB*, 12(9):1030–1043, 2019.
[14] H. Hacigümüs, B. R. Iyer, C. Li, and S. Mehrotra. Executing SQL over encrypted data in the database-service-provider model. In *Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data, Madison, Wisconsin, USA, June 3-6, 2002*, pages 216–227, 2002.
[15] A. Shamir. How to share a secret. *Commun. ACM*, 22(11):612–613, 1979.
[16] D. X. Song, D. A. Wagner, and A. Perrig. Practical techniques for searches on encrypted data. In *IEEE Symposium on Security and Privacy*, pages 44–55. IEEE Computer Society, 2000.
[17] T. Xiang, X. Li, F. Chen, S. Guo, and Y. Yang. Processing secure, verifiable and efficient SQL over outsourced database. *Inf. Sci.*, 348:163–178, 2016.