Securing Big Data: New Access Control Challenges and Approaches

Murat Kantarcioglu
muratk@utdallas.edu
University of Texas at Dallas
Richardson, Texas

ABSTRACT

Recent cyber attacks have shown that the leakage/stealing of big data may result in enormous monetary loss and damage to organizational reputation, and increased identity theft risks for individuals. Furthermore, in the age of big data, protecting the security and privacy of stored data is paramount for maintaining public trust, and getting the full value from the collected data.

In this talk, we first discuss the unique security and privacy challenges arise due to big data and the NoSQL systems designed to analyze big data. Also we discuss our proposed SecureDL system that is built on top of existing NoSQL databases such as Hadoop and Spark and designed as a data access broker where each request submitted by a user app is automatically captured. These captured requests are logged, analyzed and then modified (if needed) to conform with security and privacy policies (e.g., [5]), and submitted to underlying NoSQL database. Furthermore, SecureDL can allow organizations to audit their big data usage to prevent data misuse and comply with various privacy regulations [2]. SecureDL is totally transparent from the user point of view and does not require any change to the user's code and/or the underlying NoSQL database systems. Therefore, it can be deployed on existing NoSQL databases.

Later on, we discuss how to add additional security layer for protecting big data using encryption techniques (e.g., [1, 3, 4]). Especially, we discuss our work on leveraging the modern hardware based trusted execution environments (TEEs) such as Intel SGX for secure encrypted data processing. We also discuss how to provide a simple, secure and high level language based framework that is suitable for enabling generic data analytics for non-security experts who do not have security concepts such as "oblivious execution". Our proposed framework allows data scientists to perform the data analytic tasks with TEEs using a Python/Matlab like high level language; and automatically compiles programs written in our language to optimal execution code by managing issues such as optimal data block sizes for I/O, vectorized computations to simplify much of the data processing, and optimal ordering of operations for certain tasks. Using these design choices, we show how to provide guarantees for efficient and secure big data analytics over encrypted data.

SACMAT '19, June 3–6, 2019, Toronto, ON, Canada
© 2019 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-6753-0/19/06.
https://doi.org/10.1145/3322431.3326330

CCS CONCEPTS

• **Security and privacy** → **Data anonymization and sanitization**; **Management and querying of encrypted data**; **Information accountability and usage control**; **Database activity monitoring**.

KEYWORDS

NoSQL databases; access control ; security ; privacy ; intrusion detection; encrypted data processing

ACM Reference Format:
Murat Kantarcioglu. 2019. Securing Big Data: New Access Control Challenges and Approaches. In *The 24th ACM Symposium on Access Control Models and Technologies (SACMAT '19), June 3–6, 2019, Toronto, ON, Canada.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3322431.3326330

BIOGRAPHY

Dr. Murat Kantarcioglu is a Professor in the Computer Science Department and Director of the Data Security and Privacy Lab at The University of Texas at Dallas (UTD). He received a PhD in Computer Science from Purdue University in 2005 where he received the Purdue CERIAS Diamond Award for Academic excellence. He is also a visiting scholar at Harvard Data Privacy Lab since 2013. Dr. Kantarcioglu's research focuses on the integration of cyber security, data science and blockchains, creating technologies that can efficiently and securely process and share data.

His research has been supported by grants including from NSF, AFOSR, ARO, ONR, NSA, and NIH. He has published over 170 peer reviewed papers in top tier venues such as ACM KDD, SIGMOD, IEEE ICDM, ICDE, PVLDB, NDSS, USENIX Security and several IEEE/ACM Transactions as well as served as program chair for conferences such as ACM SACMAT, IEEE Cloud, ACM CODASPY. Some of his research work has been covered by the media outlets such as the Boston Globe, ABC News , PBS/KERA, DFW Television, and has received multiple best paper awards. He is the recipient of various awards including NSF CAREER award, the AMIA (American Medical Informatics Association) 2014 Homer R Warner Award and the IEEE ISI (Intelligence and Security Informatics) 2017 Technical Achievement Award presented jointly by IEEE SMC and IEEE ITS societies for his research in data security and privacy. He is also a Distinguished Scientist of ACM.

ACKNOWLEDGMENTS

The research reported herein was supported in part by NIH award 1R01HG006844, NSF awards CICI- 1547324, IIS-1633331, CNS-1837627, OAC-1828467, IIP-1758628 and ARO award W911NF-17-1-0356.

REFERENCES

[1] Mohammad Saiful Islam, Mehmet Kuzu, and Murat Kantarcioglu. 2012. Access Pattern disclosure on Searchable Encryption: Ramification, Attack and Mitigation. In *19th Annual Network and Distributed System Security Symposium, NDSS 2012, San Diego, California, USA, February 5-8, 2012.* https://www.ndss-symposium.org/ndss2012/access-pattern-disclosure-searchable-encryption-ramification-attack-and-mitigation

[2] Mohammad Saiful Islam, Mehmet Kuzu, and Murat Kantarcioglu. 2015. A Dynamic Approach to Detect Anomalous Queries on Relational Databases. In *Proceedings of the 5th ACM Conference on Data and Application Security and Privacy, CODASPY 2015, San Antonio, TX, USA, March 2-4, 2015.* 245–252. https://doi.org/10.1145/2699026.2699120

[3] Mehmet Kuzu, Mohammad Saiful Islam, and Murat Kantarcioglu. 2012. Efficient Similarity Search over Encrypted Data. In *IEEE 28th International Conference on Data Engineering (ICDE 2012), Washington, DC, USA (Arlington, Virginia), 1-5 April, 2012.* 1156–1167. https://doi.org/10.1109/ICDE.2012.23

[4] Fahad Shaon, Murat Kantarcioglu, Zhiqiang Lin, and Latifur Khan. 2017. SGX-BigMatrix: A Practical Encrypted Data Analytic Framework With Trusted Processors. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, CCS 2017, Dallas, TX, USA, October 30 - November 03, 2017.* 1211–1228. https://doi.org/10.1145/3133956.3134095

[5] Zhiyu Wan, Yevgeniy Vorobeychik, Weiyi Xia, Ellen Wright Clayton, Murat Kantarcioglu, Ranjit Ganta, Raymond Heatherly, and Bradley A. Malin. 2015. A Game Theoretic Framework for Analyzing Re-Identification Risk. *PLOS ONE* 10, 3 (03 2015), 1–24. https://doi.org/10.1371/journal.pone.0120592