

# A Hybrid Approach for Privacy-preserving Record Linkage

Murat Kantarcioglu  
Department of Computer Science  
The University of Texas at Dallas  
Richardson, TX 75080, USA  
muratk@utdallas.edu

## ABSTRACT

The integration of information dispersed among multiple repositories is a crucial step for accurate data analysis in various domains. In support of this goal, it is critical to devise procedures for identifying similar records across distinct data sources. At the same time, to adhere to privacy regulations and policies, such procedures should protect the confidentiality of the individuals to whom the information corresponds. Various private record linkage (PRL) protocols have been proposed to achieve this goal, involving secure multi-party computation (SMC) and similarity preserving data transformation techniques. SMC methods provide secure and accurate solutions to the PRL problem, but are prohibitively expensive in practice for large data sets, mainly due to excessive computational requirements. Data transformation techniques offer more practical solutions, but incur the cost of information leakage and false matches.

In this talk, we discuss how the performance of SMC based PRL techniques could be significantly improved by combining them with data sanitization techniques without incurring the cost of information leakage and false matches. Furthermore, we discuss how to efficiently handle typographical errors exist in data during the PRL protocol execution.

## BIO

Dr. Murat Kantarcioglu is an Associate Professor in the Computer Science Department and Director of the UTD Data Security and Privacy Lab at the University of Texas at Dallas. He holds a B.S. in Computer Engineering from Middle East Technical University, and M.S. and Ph.D degrees in Computer Science from Purdue University. He is a recipient of NSF CAREER award and Purdue CERIAS Diamond Award for Academic excellence. Currently, he is a visiting scholar at Harvard Data Privacy Lab.

Dr. Kantarcioglu's research focuses on creating technologies that can efficiently extract useful information from any data without sacrificing privacy or security. His research has been supported by grants from NSF, AFOSR, ONR, NSA, and NIH. He has published over 100 peer reviewed papers. Some of his research work has been covered by the media outlets such as Boston Globe, ABC News etc. and has received two best paper awards.

(c) 2014, Copyright is with the authors. Published in the Workshop Proceedings of the EDBT/ICDT 2014 Joint Conference (March 28, 2014, Athens, Greece) on CEUR-WS.org (ISSN 1613-0073). Distribution of this paper is permitted under the terms of the Creative Commons license CC-by-nc-nd 4.0

*7th International Workshop on Privacy and Anonymity in the Information Society (PAIS'14)* March 28, 2014, Athens, Greece