

Privacy in Data Integration

Wei Jiang
Ph.D., Purdue University

Date: Feb 15th, 2008
Time: 3:00 PM
Venue: 256 Coates Hall

Abstract

Integration of data can reveal a full profile of an individual, placing personal privacy at risk. Anonymizing such datasets can mitigate risk by preventing re-identification of sensitive data. While algorithms exist for producing k-anonymous data, the model has been that of a single source wanting to publish data. What if the data is privately distributed? Integrating to produce the anonymized data poses a threat. This talk presents a multiparty framework that generates k-anonymous data from vertically partitioned sources, without disclosing data from one site to another. The framework is secure according to the definition of Secure Multiparty Computation.

Most privacy-preserving protocols in this area are only secure under the semi-honest model, which is rarely sufficient for practical applications since there is no way to monitor participating parties' behaviors. Is trusting parties to follow the protocol safer than trusting them with the data? Most existing secure protocols under the malicious model are very inefficient. To overcome these difficulties, this talk also presents the accountable-computing framework that allows "trust but verify"; that is, a party who correctly followed the protocol can be proven to have done so, and malicious behaviors can be detected. This leads to practical and efficient privacy-preserving protocols.