Privacy for Big Data
SWITS 2016

Boel Nelson

May 2, 2016

My research concerns big data and security, and in particular I am researching privacy for big data. Most recently, I have been working with enforcing differential privacy for big data from cyber-physical systems.

Differential privacy is a privacy model with strong privacy guarantees. When querying a database in a differentially private manner, there is no added privacy risk for an individual to participate in the database. That is, from a privacy perspective, it does not matter for an individual if it belongs to the database or not.

To preserve privacy, one common technique to enforce differential privacy is to introduce controlled Laplacian noise to query answers. When processing big data, it is especially interesting to provide privacy using this technique, as it is independent of the size of the database. However, the downside of adding noise is that for some queries, the noise can ruin the utility of the answer.

My most recent research has been focused on how to release high accuracy, differentially private, answers when performing noisy queries such as sums. In order to do this, I have been investigating the trade-off between approximation errors for histograms, and the noise added by the differentially private release mechanism.

Furthermore, I am working on constructing a systematic survey of big data papers with a focus on security and privacy. In total, this survey involves 150 papers from top conferences. The goal is to categorize the papers into different topics, as well as identify what topics relate to each other. Thus, the survey both showcases the current state-of-the-art and also highlights possible areas for future work.