UNIVERSITY of GUELPH

COLLEGE of ENGINEERING AND PHYSICAL SCIENCES

SCHOOL OF COMPUTER SCIENCE

PhD Seminar 2

Monday April 17, 2023 at 10am via Zoom [Remote]

Wenjing Zhang

Protecting Spatial-Temporal Density Data with Generative Adversarial Privacy

Advisor: Dr. Xiaodong Lin
Advisory: Dr. Ali Dehghantanha
Advisory: Dr. Lei Lei [SoE]

Abstract:

Numerous applications of spatial-temporal density, i.e., the number of individuals on a map at a given period of time, offer indisputable benefits in understanding complex processes, such as the spread of viruses, building better intelligent transport systems, preventing traffic congestion, identifying areas where to install new businesses or building new infrastructures. Unfortunately, the direct release of spatial-temporal density data poses a considerable threat to individuals' location privacy because their mobility patterns can be recognized and their traces can be reconstructed with very high accuracy. Therefore, it is essential to ensure that the density data is released in a privacy-preserving manner to mitigate such privacy breaches.

Most existing works on privacy protection of mobility data rely on the knowledge of underlying data statistics, which is hard to obtain in practice due to missing or inaccurate data. To address this issue, we propose a data-driven approach for privacy-preserving spatial-temporal density release with limited data samples. Specifically, we design a conditional generative adversarial training framework taking individuals' traces as training samples to learn a privacy mechanism that generates a perturbed version of density data while maintaining certain data utility.

In particular, the generator's loss function is defined as the mutual information between an individual's trace and the perturbed density with the goal of minimizing the fundamental information leakage on her traces. One challenge in our work is how to measure privacy leakage and ensure a rigorous privacy guarantee on individuals' traces without access to data probability distributions. We address this challenge by leveraging a new class of information measure built on neural networks to approximate the true mutual information and giving a bound on the privacy leakage related to sample complexity regarding the use of empirical measures. Experimental results show that the generated density data can achieve a better privacy-utility trade-off than a state-of-the-art differentially private mechanism.