FLORIDA ATLANTIC UNIVERSITY

Department of Mathematics and Statistics
Florida Atlantic University

PhD Dissertation Defense

Dominic Gold

Privacy-Preserving Topological Data Analysis Using Homomorphic Encryption

Tuesday, April 23, 11:00am in SE 215

Advisor: Dr. Francis Motta

Computational tools grounded in algebraic topology, known collectively as topological data analysis (TDA), have been used for dimensionality-reduction to preserve salient and discriminating features in data. This faithful but compressed representation of data through TDA's flagship method, persistent homology (PH), motivates its use to address the complexity, depth, and inefficiency issues present in privacy-preserving, homomorphic encryption (HE)-based machine learning (ML) models, which permit a data provider (often referred to as the Client) to outsource computational tasks on their encrypted data to a computationally-superior but semi-honest party (the Server). This work introduces efforts to adapt the well-established TDA-ML pipeline on encrypted data to realize the benefits TDA can provide to HE's computational limitations as well as provide HE's provable security on the sensitive data domains in which TDA has found success in (e.g., sequence, gene expression, imaging). The privacy-protecting technologies which could emerge from this foundational work will lead to direct improvements to the accessibility and equitability of health care systems. ML promises to reduce biases and improve accuracies of diagnoses, and enabling such models to act on sensitive biomedical data without exposing it will improve trustworthiness of these systems.

To adapt the beginning steps of the TDA-ML pipeline, we create an HE-compatible arithmetic circuit of the fundamental map to compute PH on an encrypted boundary matrix for further use in downstream model development (with a complete construction, parameter selection guarantees, and error analysis). We achieve this by modifying the logical structure of the map and by developing new arithmetic circuits to replace its computational and conditional statements. We also show work in adapting the terminal steps of the TDA-ML pipeline to realize the boons TDA affords HE-ML models on the MNIST digits dataset using a logistic regression (LR) classifier. We demonstrated that the TDA methods chosen improve encrypted model inference with a 10-25 fold reduction in amortized time while improving model accuracy up to 1.4% compared to naive reductions that used downscaling/resizing, and we show the first steps in realizing these same improvements on encrypted model training.

Please contact Dr. Hongwei Long (hlong@fau.edu) for an electronic copy of the dissertation.
ALL ARE CORDIALLY INVITED