PRIVACY PRESERVING CLASSIFICATION OVER ENCRYPTED DATA USING FULLY HOMOMORPHIC ENCRYPTION TECHNIQUE

By

ABDULLAHI MONDAY JUBRIN *

VICTOR ONOMZA WAZIRI **

MUHAMMAD BASHIR ABDULLAHI ***

IDRIS ISMAILA ****

*,*** Department of Computer Science, Federal University of Technology, Minna, Nigeria, and Department of Computer Science, Veritas University, Abuja, Nigeria.

,** Department of Cyber Security Science, Federal University of Technology Minna, Nigeria.

Date Received: 11/01/2019 Date Revised: 25/01/2019 Date Accepted: 06/03/2019

ABSTRACT

Applying Machine Learning to a problem which involves medical, financial, or other types of sensitive data needs careful attention in order to maintaining data privacy and security. This paper presents a model for privacy preserving classification and demonstrated that, by using a decision tree classifier, it is possible to perform a privacy preserving classification operation on an encrypted data residing on an untrusted server using the technique of Fully Homomorphic Encryption. First, the paper presented a model for the design and implementation of privacy preserving decision tree classifier over encrypted data. Also, Fully Homomorphic Encryption technique was used to secretly carry out classification on ciphertext using decision tree model built out of confidential medical data. The classifier was implemented using the SEAL homomorphic library and evaluation was done using encrypted medical datasets. The experimental results demonstrated high accuracy of the ciphertext classifier (when compared to the plaintext data equivalent) and efficiency (compared to other classifier on similar tasks). It takes less than 5 seconds (depending on the depth) to perform classification over an encrypted hepatitis feature vector dataset.

Keywords: Privacy Preserving, Machine Learning, Algorithms, Helib, Homomorphic Encryption, Classification, Classifiers, RLWE, SEAL, Decision Tree.

INTRODUCTION

Data that is needed for vital mining evaluations are becoming more and more voluminous, and as a result, the data and computational tasks on them are outsourced and confided on untrusted third party service providers' servers in data centres and the cloud. As more business parties and individuals entrust their data and outsource computational tasks to facilities (servers) owned by third party, concerns for privacy of such information is being raised. These concerns are genuine since the data residing on the server of the third party can be perturbed and also computations on such data leaks the sensitive information as a result of data mining. While endless examples of areas for such privacy concerns exist, one area of utmost privacy concerns that deserve special attention is Medical Computing (Kocabas, Soyata & Aktas, 2016). In the medical sector, confidentiality of sensitive private records is mandatory as stated in the rules and regulations introduced by Health Insurance Portability and Accountability Act in the USA (HIPAA, 2014). According to HIPAA regulations, private medical information should be treated with utmost care and privacy. Traditionally, privacy of such information is only guaranteed if prior to being uploaded to a third party (cloud service) server, the data is encrypted by the owner (Bos, Lauter, & Naehrig, 2014). Through this process, only the rightful owner of the data should have connection to the data by the use of their secret key. Nevertheless encryption restricts the desire to delegate computations on the stored information because the data centre does not have the key to decrypt them since the secret key is needed to decrypt the data before any computation