

Security and Privacy of Cloud Based Big Data Systems as Platform-as-a-Service (PaaS) for IoT Applications

Yixi Luo

Advisor: Dr. Sunnie Chung



Washkewicz College
of Engineering

Introduction

Big data is machine captured interactions, transactions and observations generated by mass population. Big data analytics systems on cloud is zero trust because of not only interceptor and hacker attacks but also privilege abuse by untrusty third-party provided system administrators.

We improved two Anti-Tamper schemes to compute database queries and deep learning algorithms directly over encrypted data on cloud servers. Anti-Tamper database is to protect privacy in Database-as-a-Service with Semantic Preserving Encryptions along with Anti-Tamper Cloud to protect Privacy in Deep learning Analytics server with Database Server (Platform-as-a-Service) with Homomorphic Encryption. Both schemes keep encrypted data only on cloud servers.

Stronger Semantic Preserving Encryption Algorithms are used to improve computation performance on cloud servers, especially Fully Homomorphic Encryption.

Problems of Security and Privacy of Cloud

Threat Categorization

Cyber Threats Categorized into Two Types.

- Threat 1: Privilege abuse of administrators is a threat to cloud servers. Untrusty cloud system administrators can easily cover their tracks to steal or misuse privacy data. Malicious actions of cloud system administrators can be hard to detect.
- Threat 2: Interceptors and hackers exploit vulnerabilities to gain unauthorized access to servers. Attackers with physical access to servers can tamper all data on disk and in memory.

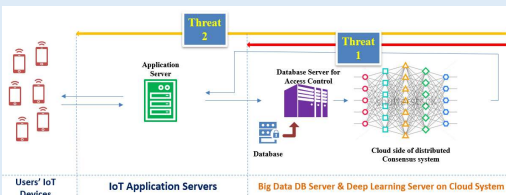


Figure 1: Two Types of Threats of IoT Applications

Limitation of Common Encryption System

- The encrypted data need to be decrypted to be processed to get the results on cloud site.
- Treat 1 has all the permissions to keys and passwords to decrypt.
- Cloud system administrators are from the third party, they are not trusty.

Our Approaches

- All data have been encrypted by Semantic Preserving Encryption.
- Keep Encrypted Data ONLY on Cloud Servers.
- Compute Database Query and Deep Learning Algorithm Directly over the Encrypted Data.
- No Point of Decryption During the Entire Communications between User and Cloud Servers.

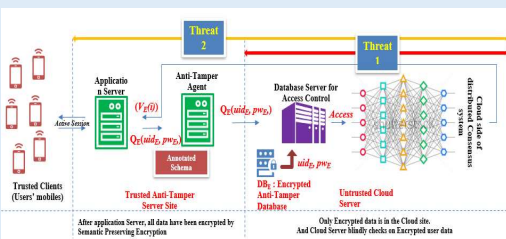


Figure 2: No Decryption Scheme

Privacy of Database-as-a-Service on Cloud

Challenge of Anti-Tamper Database-as-a-Service on Cloud

Q: How to Compute database retrieval queries over the encrypted data without decrypting?

A: Semantic Preserving Encryption Algorithm.

Semantic Preserving Encryption Algorithms

Order Preserving Encryption (OPE).

- Encryption function preserves numerical ordering of the plaintexts.

Homomorphic Encryption (HME)

- Encryption function preserves computation

Deterministic Scheme (DET)

- Given a particular input, will always produce the same output

Figure 3: Three Kinds of Semantic Preserving Encryption

Stronger Semantic Preserving Encryption

- Order Revealing Encryption (ORE)
- Fully Homomorphic Encryption (FHE)

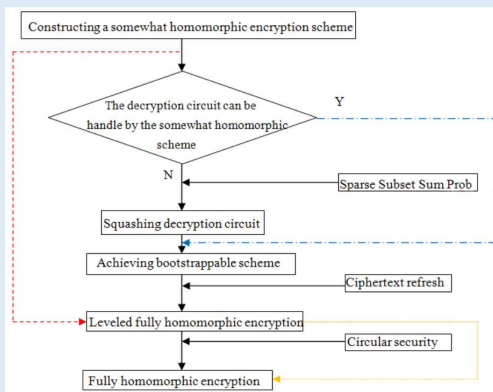


Figure 4: A General Interpretation of How to Construct FHE Scheme

Privacy Deep Learning-as-a-Service on Cloud Platform

Our Distributed Deep Learning System on Cloud

- Store the users' data in local database (smart phone).
- Transfer parameters to server to update the network structure.

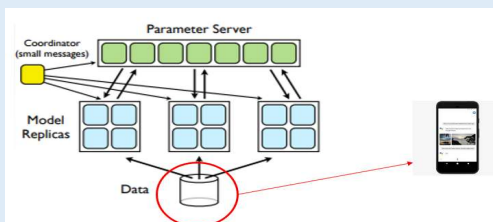


Figure 5: Distribute Deep Learning System on Cloud

Problem about Machine Learning Computation on Cloud

Q: How to Perform Machine Learning Computation over the Homomorphic Encrypted Data without Decryption?

A: Fully Homomorphic Encryption with Practical Computation Method for each Layer.

Five Steps of Weight Parameters Updating

- Client sends trained result to Anti-Tamper Agent.
- Anti-Tamper Agent rewrites queries to get access to updating the cloud model.
- Database Server sends encrypted weights to cloud server.
- Cloud server sends newest parameters to Anti-tamper Agent.
- Anti-Tamper Agent sends parameters to client.

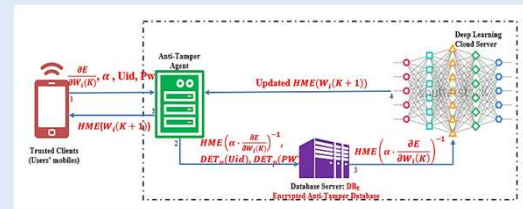


Figure 6: Weight Parameters Updating

Challenges on Encrypted Data Computation on Deep Learning Service on Cloud

- Performance of Computation over Homomorphic Encrypted Data
- Accuracy of Encrypted Data Processing and Computations

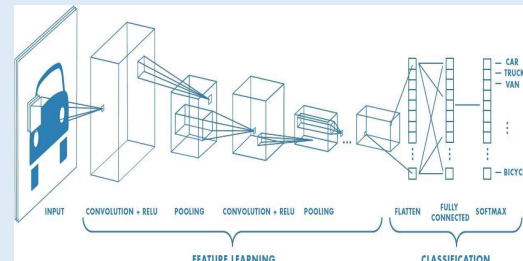


Figure 7: Typical Deep Learning Convolutional Neural Network

Our Approaches

All data computing in our deep learning convolutional neural network model are Encrypted Data

- Convolution Layer: Using rotated method to calculate convolution matrix.
- Pooling Layer: Average Pooling Layer
- Activation Layer (Algorithm): Using Polynomial Approaches typical activation function.

$$\begin{aligned} & x_1, x_2, x_3, \dots, x_{n-1}, x_n \\ & \downarrow \text{Rotate 1} \\ & x_n, x_1, x_2, \dots, x_{n-2}, x_{n-1} \\ & \downarrow \text{Sum} \\ & x_1 + x_n, x_2 + x_1, x_3 + x_2, \dots, x_{n-3} + x_{n-2}, x_{n-2} + x_{n-1} \\ & \downarrow \text{Rotate 2} \\ & x_{n-3} + x_{n-2}, x_{n-2} + x_{n-1}, x_1 + x_n, \dots, x_{n-5} + x_{n-4} \\ & \downarrow \text{Sum} \\ & \sum_{i=1}^n x_i, \sum_{i=1}^n x_i, \sum_{i=1}^n x_i, \dots, \sum_{i=1}^n x_i, \sum_{i=1}^n x_i \end{aligned}$$

Figure 8: Rotated Method

References

- Craig Gentry. 2009. A fully homomorphic encryption scheme. Ph.D. Dissertation. Stanford University, Stanford, CA, USA.
- Özsoyoglu, G., Singer, D.A., & Chung, S.S. (2003). Anti-Tamper Databases: Querying Encrypted Databases. DBSec.
- Farzad Firozabadi, Wael J. Obidallah, and Bijan Raahemi. 2017. Cloud computing service discovery framework for IaaS and PaaS models. In Proceedings of the Second International Conference on Internet of things, Data and Cloud Computing (ICC '17). Association for Computing Machinery, New York, NY, USA, Article 82, 1–6.
- Dmytro Bogatov, George Kollios, and Leonid Reyzin. 2019. A comparative evaluation of order-revealing encryption schemes and secure range-query protocols. Proc. VLDB Endow. 12, 8 (April 2019), 933–947.
- C. Borrego, M. Amadeo, A. Molinaro and R. H. Jhaveri, "Privacy-Preserving Forwarding Using Homomorphic Encryption for Information-Centric Wireless Ad Hoc Networks," in IEEE Communications Letters, vol. 23, no. 10, pp. 1708–1711, Oct. 2019.