###

Hope4Sec
Tallinn
contact@hope4sec.eu

Release date: 09/11/2024

Hope4Sec has taken a major step forward with a new class of cryptographic primitives enabling ultra-fast processing of encrypted data using conventional AI algorithms.

IA on non-sovereign cloud infrastructures is becoming a reality compatible with GDPR data privacy and security regulation, confidentiality requirements and carbon footprint reduction.

In this release, the "AI" term is used to describe all data analysis techniques (machine learning, deep learning) to the exclusion of LLM (generative AI).

Using a new cryptographic approach in data security, known as HDA (Homomorphic Data Analysis), Hope4Sec succeeded in carrying out homomorphic AI. These techniques analyze and process data in its cryptographically secure form, while using existing AI algorithms (supervised or unsupervised techniques, native TensorFlow libraries, Keras).

The few known approaches (CKKS and BFV schemes) need to rewrite the functions and algorithms used in data analysis. In real practice scenarios, all of these techniques are slow and demanding in computing and memory resources. In addition to significant computing overhead costs, they also suffer from precision loss and limitations.

Considering the above-mentioned weaknesses and limitations of existing homomorphic techniques, Hope4Sec developed powerful and disruptive homomorphic techniques, known as HDA techniques. These innovative techniques make possible ultra-fast processing/analysis of encrypted data using conventional, "off-the-shelf" AI algorithms. They offer two main benefits:

- HDA techniques **fully guarantee data security** and enable you to process/analyze data **with the same algorithms as those already used natively in AI tools and libraries** (only the parameter settings may change, depending on the desired performance).
- HDA techniques allow you to work on **data and models with reduced size (theoretical results confirmed by tests showed a reduction by a factor of 3).**

Having successfully tested this approach using traditional unsupervised and supervised learning techniques (clustering), a new milestone has now been reached with the use of neural networks (Multilayer Perceptron). The Fashion-MNIST reference dataset (developed by Zalendo Research) was cryptographically processed using HDA techniques and **then subjected as is to the algorithms conventionally used under TensorFlow (same parameters) for the learning and validation phase**. The result (see Figure below) shows a learning profile that is very similar to the learning profile for the unprotected version of the reference dataset.

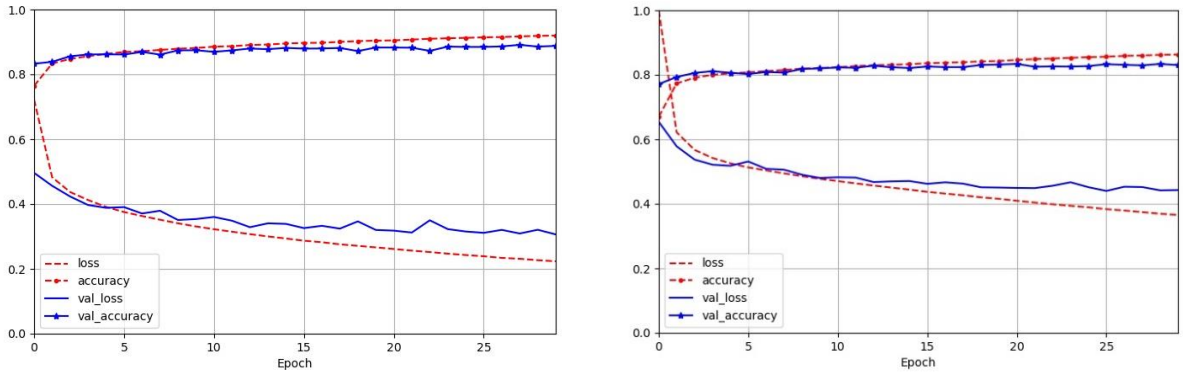
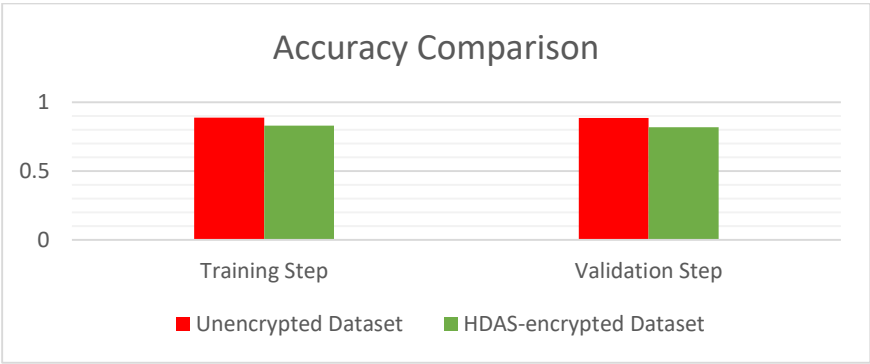


Figure 1.- Comparison of Learning Curves (left unencrypted dataset, right HDA-encrypted dataset)

Raw performance (training and validation) is already very significant, even if still a little below expectations.



On the other hand, the computing time is significantly reduced (18% for the training phase and 12% for the validation phase) once the data is pre-processed by HDA. This technique therefore offers a notable reduction in carbon footprint, and there is a strong likelihood that future versions of HDA will reduce computing time even further (the target is a reduction of at least 30%).

As a second stage, a data scientist worked blindly on the HDA-protected Fashion-MNIST dataset to improve the final validation accuracy. The aim was to test the full range of "off-the-shelf" Machine Learning and Deep Learning techniques and algorithms. **All these techniques classified the data correctly, with a validation accuracy of almost 89% overall. The best result was obtained with random forests reaching a final validation accuracy of 0.9534.**

The use-cases identified to date are as follows:

- IA on any cloud infrastructure, guaranteeing data and derived models confidentiality. Cloud service providers no longer have access to your information. All analysis processes can be performed (training, validation and test/request).
- Securing databases. Database queries can now be carried out very quickly using a homomorphic approach. The database can be hosted on any external cloud infrastructure (outsourcing services). In the event cyber-attacks resulting in data leaks, the information contained in the database remains protected, preventing access to data in its unprotected form.

In both cases, companies no longer have to develop and maintain on-premises infrastructure to ensure the confidentiality of their data. This can result in significant economies of scale (investment and operation) while contributing to an ecological approach.

Privacy is embedded in HDA techniques. Because data processing is carried out using encrypted data, companies do not need sophisticated and costly environments either to store databases or to process and secure data. As they are "privacy by-design" techniques, there is also no need to invest in additional data security measures to comply with GDPR or similar regulations. Privacy is thus integral to HDA techniques without diminishing functionality and performance.

Hope4Sec carried out a deep mathematical analysis of the security provided by HDA (using 256-bit secret key) against all the identified threat scenarios. As a result, it shows a very high level of security. Given both the current state of knowledge and the state of the art of quantum cryptanalysis techniques, HDA is Quantum-resistant.

A third use-case, Hope4Sec is working on, relates to AI in constrained environments (IoT, autonomous vehicles, robots). HDA technology is particularly well-suited to these environments, reducing storage size and computing time (and therefore memory and energy resources). In addition, as they are directly exposed to the adversary, models and data security is preserved in the event of theft.

It is Hope4Sec's ambition to further develop HDA techniques for AI in order to increase its performance while looking for partnerships including technology transfer opportunities to ensure its industrialization.

Quote 1: "*With these new cryptographic techniques, data processing location is no longer an issue as it both enables a high level of data security ("by design") and a greatly reduced carbon footprint. At a time when the EUCS project is likely to undermine European sovereignty in terms of cloud and data security. In this context, HAD technology is the perfect solution for ensuring data and AI protection. (Hope4Sec Data Privacy Officer and Ethical Officer)*."

Quote 2: "*Fashion-MNIST is a large freely available dataset of Zalando's article images—consisting of a training set of 60,000 examples and a test set of 10,000 examples. Each example is a 28x28 grayscale image, associated with a label from 10 classes. We intend Fashion-MNIST to serve as a direct* ***drop-in replacement*** *for the original MNIST dataset for benchmarking machine learning algorithms. It shares the same image size and structure of training and testing splits. (Source https://github.com/zalandoresearch/fashion-mnist)*"

Hope4Sec is a European collaborative R&D group of experts, researchers and engineers in the field of mathematical and algorithmic engineering. Hope4Sec's main field of application is information and systems security focused on respect for privacy, ethical and democratic values.

Its business model is based on transferring innovative technological solutions to stakeholders interested in developing them industrially or commercially.

Official website: https://hope4sec.eu/ Contact: contact@hope4sec.eu

###