NEWS LETTER

16

July | 2025

CLOUD STARS

Cloud, AI, and Security: Challenges for Sustainable Computing

by Maximilian Meissner (University of Wuerzburg)

The integration of artificial intelligence (AI) and cloud computing offers tremendous potential, but also presents two major challenges: **efficiency** and **security**. While the cloud provides scalable, on-demand resources ideally suited for compute-intensive AI workloads, the handling of sensitive data remains a significant obstacle. This is especially true in domains such as healthcare and finance, where strict privacy regulations limit the use of cloud-based AI solutions. Traditionally, executing AI algorithms in the cloud requires data to be decrypted,

exposing it to potential security risks during processing. A promising solution to this dilemma is Fully Homomorphic Encryption (FHE), which enables computations to be performed directly on encrypted data, ensuring that sensitive information remains confidential, even during processing in untrusted environments. FHE could eliminate a major barrier to adopting AI in regulated fields. However, it incurs significant overhead in terms of performance and power consumption, making it impractical for most real-world applications today.

In collaboration with experts at the IBM Thomas J. Watson Research Center, we carried out a detailed performance and efficiency analysis of various fundamental FHE workloads, including a case study using a Neural Network benchmark for credit card fraud detection. As part of our study, we examined how different parameter settings impact performance, power consumption, and security level. The insights from this work can guide the optimization of FHE workloads and shape future developments in both hardware and software.

This research aligns with our group's broader focus on energy efficiency benchmarking, particularly within the SPEC RG Power Working Group, a collaborative effort between academia and industry. Our mission is to advance and promote methods and tools for evaluating and improving energy and resource efficiency of computing systems, addressing an essential concern for industrial stakeholders, academia, as well as regulatory institutions and policymakers. As AI applications continue to grow in both complexity and demand, developing next-generation efficiency benchmarks as well as mechanisms to improve cloud infrastructure efficiency with these workloads in mind becomes increasingly critical.

CLOUDSTARS project has received funding from the European Union's Horizon research and innovation programme under grant agreement No 101086248