Trusted Execution Environment (TEE)

Central to our project's commitment to privacy and security in the evaluation of AI models is the integration of TEEs. TEEs provide a secure area within a main processor, ensuring that sensitive data and operations are insulated from the rest of the system. Within our project, this secure environment is utilized to perform computations on encrypted private user data during the AI inference process. By leveraging TEEs, we ensure that the data, although processed by the inference committee, remains confidential and tamper-proof throughout the entire inference lifecycle.

This approach means that none of the sensitive information is exposed to any of the inference nodes to preserve the integrity and secrecy of the data. The inference committee, composed of a pre-selected set of nodes, is responsible for collaboratively conducting AI model inference without having direct access to unencrypted data. This creates a robust framework that facilitates user privacy while enabling secure and reliable model evaluations in a decentralized environment, turning our vision of secure and private AI computation into a tangible reality.

Building upon the concept of the TEE, there are several implementations of TEE technologies designed to cater to different types of processing units and their respective architectures. In the CPU domain, prominent players have advanced their offerings to provide robust security solutions:

- Intel'sTrustedDomainExtensions(TDX) is designed to enhance the security of virtual machines by providing hardware-level isolation capabilities. TDX creates private regions of memory, known as Trusted Domains, which help to protect code and data from external threats and unauthorized system software.
- AMDCPUscountersecuritythreatswithSecureEncryptedVirtualizationwith Secure Nested Paging (SEV-SNP). This technology adds strong memory integrity protection capabilities to the already existing SEV technology, further fortifying virtual machine isolation and helping to prevent malicious hypervisor-based attacks.
- ARMCPUsintroducetheConfidentialComputeArchitecture(CCA), which aims to fortify application security. CCA provides a secure environment for computation, ensuring that

sensitive data can be processed without exposure to the risk of interception or tampering by other software, including the operating system.

In the GPU landscape, NVIDIA has made significant strides with their Hopper H100 GPU architecture which supports confidential computing. The H100 GPU integrates with the aforementioned CPU TEE technologies, ensuring a secure and seamless interaction between the processing units. This integration allows for the extension of TEE's security benefits into the realm of high-performance computing, making it possible to securely process complex AI and machine learning workloads that require the parallel processing power of GPUs.

These TEE technologies form a multi-layered defense strategy, providing a secure computing backbone for models deployed on Nesa. By leveraging the strengths of each technology, we create a hybrid and interoperable secure environment capable of handling a diverse array of computingcompute demands while maintaining a stringent security posture for confidential computing.

Previous
Privacy Technology

Next
Secure Multi-Party Computation (MPC)

Last updated 1 month ago