Fluid Trusted Whitepaper

Introduction

The rates of innovation and adoption of AI in businesses, academia and government are increasing exponentially, as they are beginning to learn the immense value that can be derived from AI. The two most important requisites for AI are data and the algorithm. The AI algorithms can derive useful patterns and insights from the data, due to which both the data and the algorithm are highly valuable digital assets. As the paradigm requires the model and the data to be co-located on a common platform, there arises a need for trust and privacy when these digital assets are owned by different parties.

The traditional AI paradigm does not accommodate this need for privacy of the digital assets and trust between multiple parties, which are paramount drivers of new collaborations. The lack of a framework that allows for effective collaboration between such parties, while completely preserving the privacy and ownership of their digital assets, leads to a fragmentary use of economies of scale.

The solution

To enable secure collaboration between the data owners and the AI service providers, many aspects of cryptography need to be baked into the existing AI software paradigm. Such a secure framework can be achieved through the use of cryptographical methods such as homomorphic encryption, secure multiparty computation or hardware based trusted/isolated execution environments, where the data and the AI algorithm remain secure and private throughout the process of execution.

While homomorphic encryption and secure multiparty computation (MPC) can be used to keep both data and the AI models private, they incur large computational costs and lack efficiency, especially for more complex AI models containing millions of parameters. In contrast, trusted

execution environments provide a more feasible way to achieve the same as they are several orders of magnitude faster and support general-purpose computation i.e. not just arithmetic operations as in the case of MPC.

What is Fluid trusted and how does it work?

Fluid trusted is such a framework that enables privacy preserving AI through the use of trusted execution environments (TEE) in the cloud & hybrid cloud platforms. Fluid trusted currently supports Intel SGX, which uses the TEE in the Intel processors.

Trusted Execution Environments

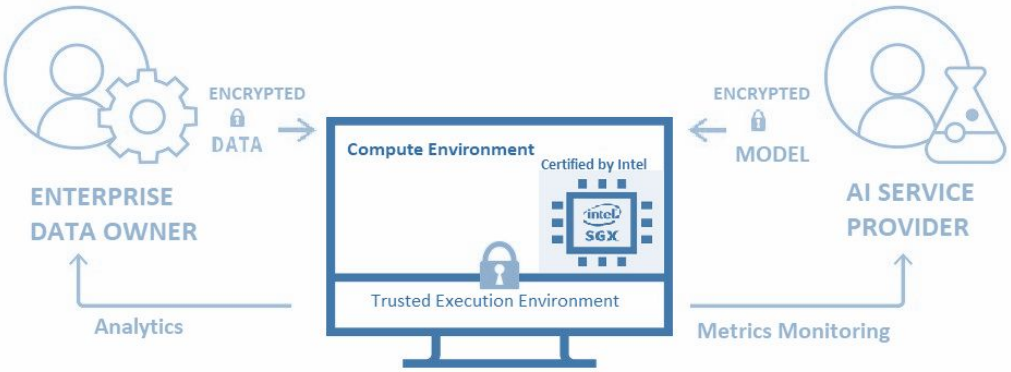
A Trusted Execution Environment (TEE) is an environment in the processor for executing computational logic, in which those executing the code can have high levels of trust in the environment which is separate from the main operating system and is not accessible from host OS's BIOS or hypervisor once initialized. It ensures that data is stored, processed and protected in a completely secure manner.

Intel SGX

Intel's Software Guard Extensions (SGX) is a set of extensions to the Intel architecture that aims to provide integrity and confidentiality guarantees for secure computation by creating enclaves* Enclaves are stored in a hardware guarded memory called Enclave Page Cache (EPC), which is currently limited to 128MB with only 90 MB for the application. They are designed to execute programs and handle secrets in the trusted/isolated execution environment without exposing data to any malicious agent, even if they may have privileged access to the machine ,such as through BIOS or Hypervisor.

SGX enables this trust between the engaging parties through the use of remote attestation. Attestation is the process of demonstrating that a software component that it is running on top of a trusted hardware platform with legitimate code and data. Attestation can be both local and remote. We use Intel Attestation Service to attest a service providers application to ensure it is running on a verified and trusted machine.

Overall workflow



Data Owner user flow (Enterprise):

1. Fluid is installed on Intel SGX enabled machine within enterprise infrastructure (on-prem or hybrid cloud)

2. Enterprise user creates projects, the user then requests the application to be sent from the service provider to the machine.

3. Then the user verifies the integrity of the application and the hardware, through remote attestation.

4. User then connects relevant (encrypted) data pipes and shares the key securely to the enclave which, using which the input data is decrypted in the enclave.

5. The user gets the encrypted output from AI model.

Model Owner user flow (Service Provider):

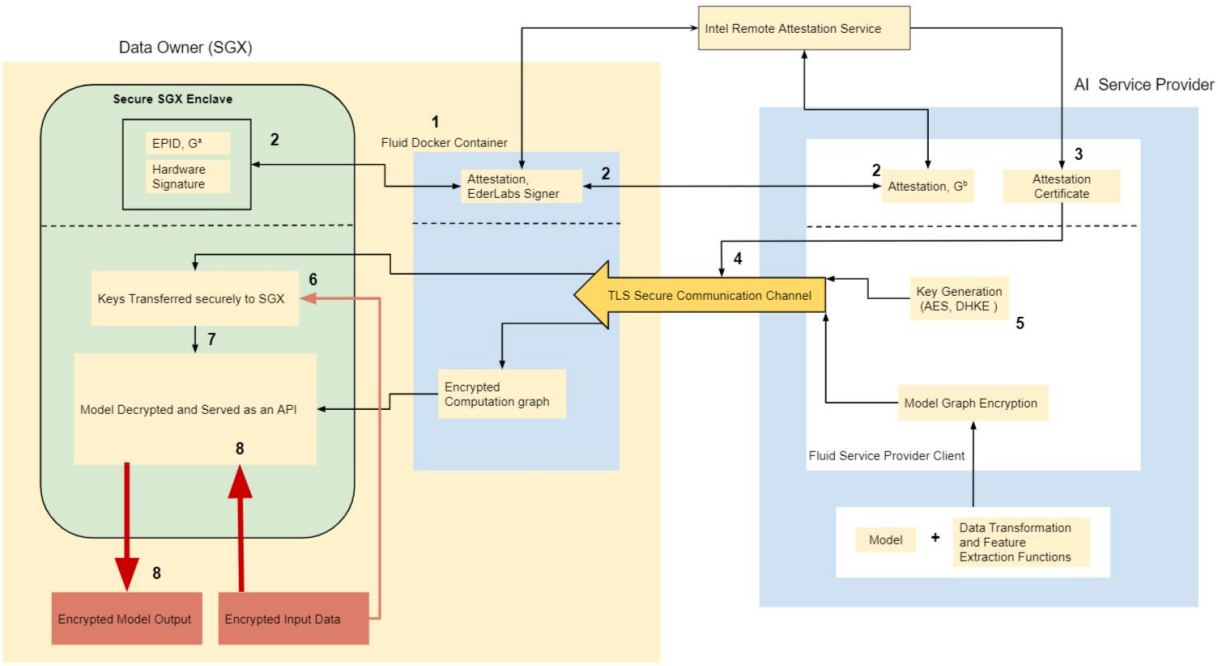
1. The ML application is first defined. The final model to be deployed is encrypted with a locally generated AES-GCM symmetric key.

2. The application is deployed to the client endpoint, once attested, a secure TLS channel is created.

3. Upload the encrypted model and share the symmetric key used to encrypt the model securely to the enclave.

    a. Obtain metrics (pre-decided in the engagement) about the performance of the model from the client infra.

Detailed process flow of the engagement using fluid



1. Eder's docker container is downloaded on the machine and run. It will start SGX services and initialise the Enclave.

2. Enclave will be attested remotely by Intel Attestation Services

3. Intel will generate an Attestation Certification verifying the integrity of the hardware.

4. After Attestation, a secure TLS channel is created.

5. Keys are generated and shared between the Enclave and Model Provider over TLS (AES, SE)

6. Keys are also shared between the Data Owner and the Enclave over TLS

7. Model graph is decrypted inside the enclave using the above keys. Fluid loads the model into SGX runtime and opens a port.

8. Encrypted Data goes into the enclave through the port and prediction is given out encrypted. This can only be decrypted by Data Owner

A few use cases across industries:

Healthcare

There exists large amounts of data in the healthcare industry owned by hospitals and Pharmaceutical companies, corresponding to patient health records and the human genome experiments, which are highly sensitive. Traditionally such data is kept private although they possess a great value as important insights can be derived from the data if multiple sources can work with a common group of models without compromising data privacy.

Automobile

In the case of the automotive industry, data security has become a pressing priority for original equipment manufacturers (OEMs) and tech companies as connectivity becomes ever more present in the vehicles of today and due to the lack of a framework that enables insight generation/information extraction on private data there exists no incentive for OEMs to share their private data.

Airline

For predictive maintenance tools to be used in real time based on Aircraft data from sensors in various components of an aircraft, there exists a need for a framework which can perform machine learning/ AI to predict and avoid failures without losing control and privacy of data obtained from aircrafts.

Finance

Compliance is an area which is particularly well suited for automation through machine learning. Banks can use machine learning to develop predictive models which are more accurate at detecting problematic transactions than the current heuristics-based methods, dramatically reducing the false-positive rate, but an impedance with the traditional methods is the lack of security and privacy of data. Problems such as credit risk assessment and VAT fraud can overcome compliance and legal issues using secure computation

Annexure:

*__Enclaves__ : enclaves are isolated regions of memory which are protected from processes running at any privilege level, including the operating system.

This document will be updated regularly. If you would like additional information, please write to sharat@eder.io