March 13th, 2024

Seamless Attestation of Intel TDX and NVIDIA H100GPUs for Confidential AI

Raghu Yeluri, Intel Corp.

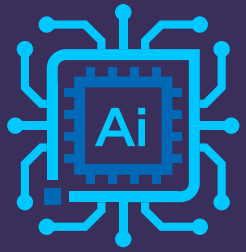
Michael O'Connor, NVIDIA

Agenda

- Confidential AI - Context
- Confidential Computing – Intel Technologies
  - Intel SGX and Intel TDX
  - Attestation: Intel Trust Authority
- Confidential Computing – NVIDIA Technologies
  - H100 GPUs
  - Attestation –SDK, NRAS and supporting services
- Seamless attestation w/ Intel Trust Authority
- Demo
- Summary

Confidential AI Helps Protect Data & Models In-Use

AI



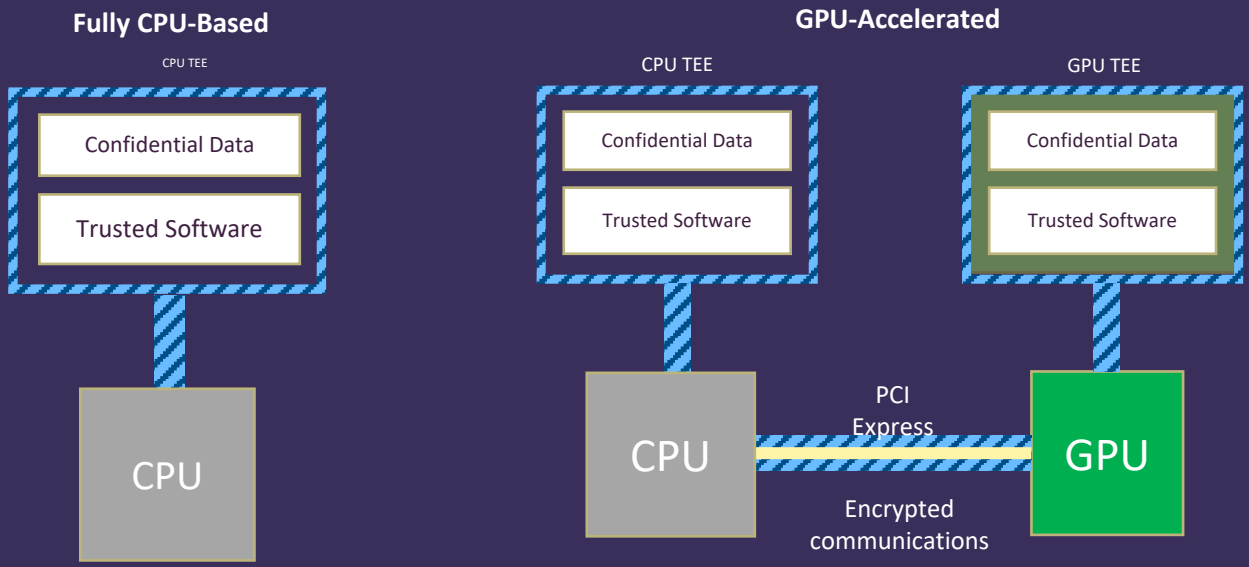
+

Confidential Computing

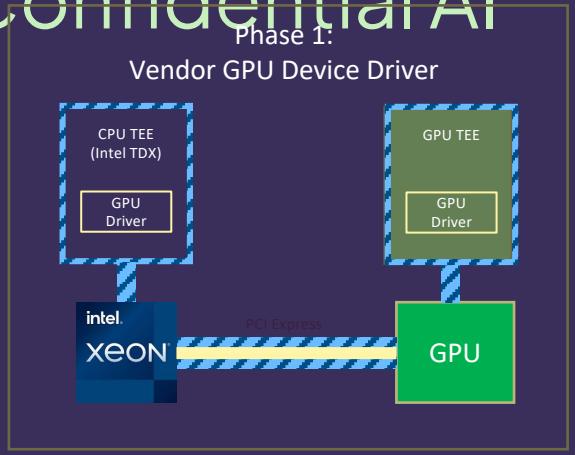
Hardware-Based Protection of Data In-Use

1. Trusted Execution Environment (TEE)
2. Encryption controlled by workload owner
3. Cryptographic attestation of TEE integrity

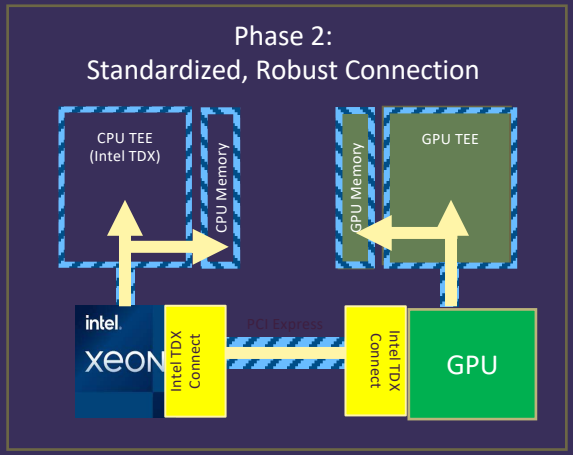
Two Types of Confidential AI



The Path to GPU-Accelerated Confidential AI



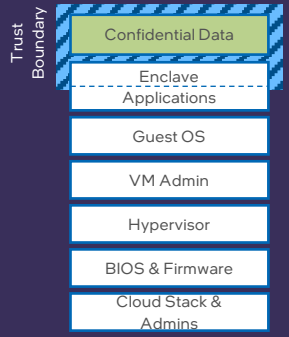
- ☑ Vendor-specific encrypted communication between TEEs



- ☑ Intel® TDX Connect
- ☑ Standards-based encrypted communication between TEEs
- ☑ Direct memory access across TEEs

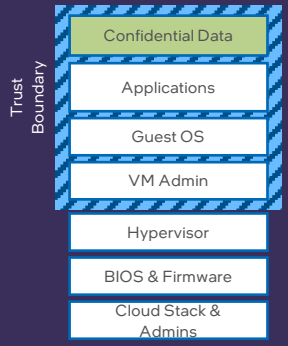
Intel's Confidential Computing Technologies

App Isolation
Intel® SGX



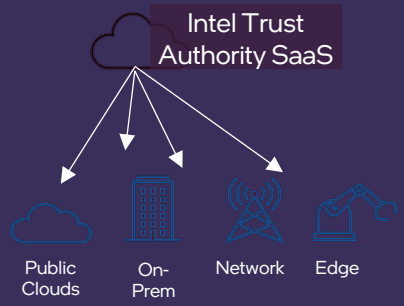
Smallest trust boundary for greatest data protection & code integrity

VM Isolation
Intel® TDX



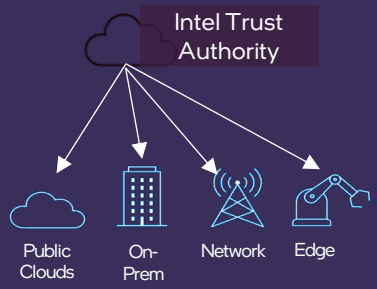
Most straightforward path to greater security, compliance & control for legacy apps

Trust Services
Intel® Trust Authority



Uniform, independent attestation of trustworthy environments

Intel® Trust Authority



Intel's independent, scalable, turnkey attestation service

Available for Intel SGX and Intel TDX

ISO:29001: 2022 certified and 99.95+ uptime SLA.

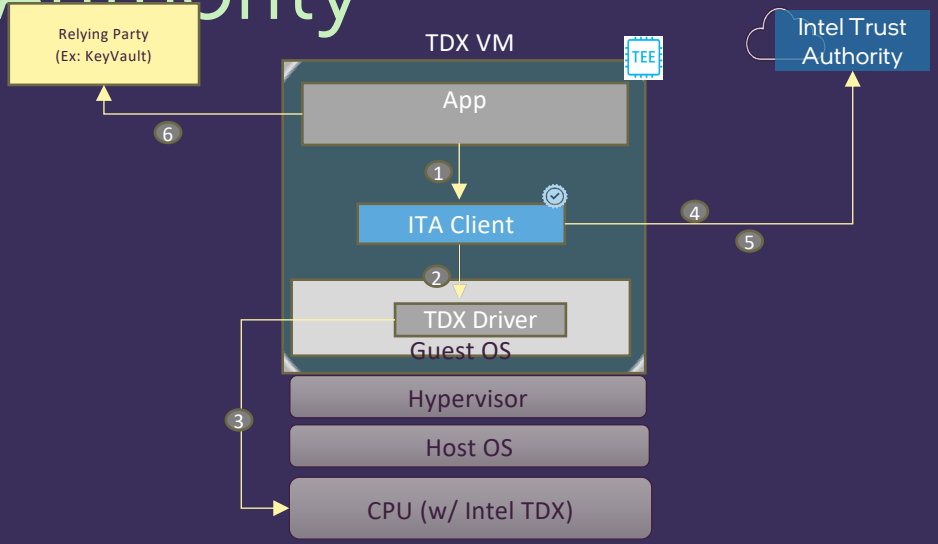
Why Use Intel Trust Authority?

1. Separates the provider of infrastructure from verifier of trust (Zero Trust principles)
2. Consistent attestation services across cloud, on-prem & edge deployments
3. Easiest solution for on-prem
4. Roadmap of growing capabilities (GPU TEEs, non-intel TEEs, platform attestation…)*

Learn more at:
Intel.com/trustauthority

*All product plans and roadmaps are subject to change without notice.

TDX Attestation with Intel Trust Authority



Steps

1. App calls ITA Client API – CollectCPUToken

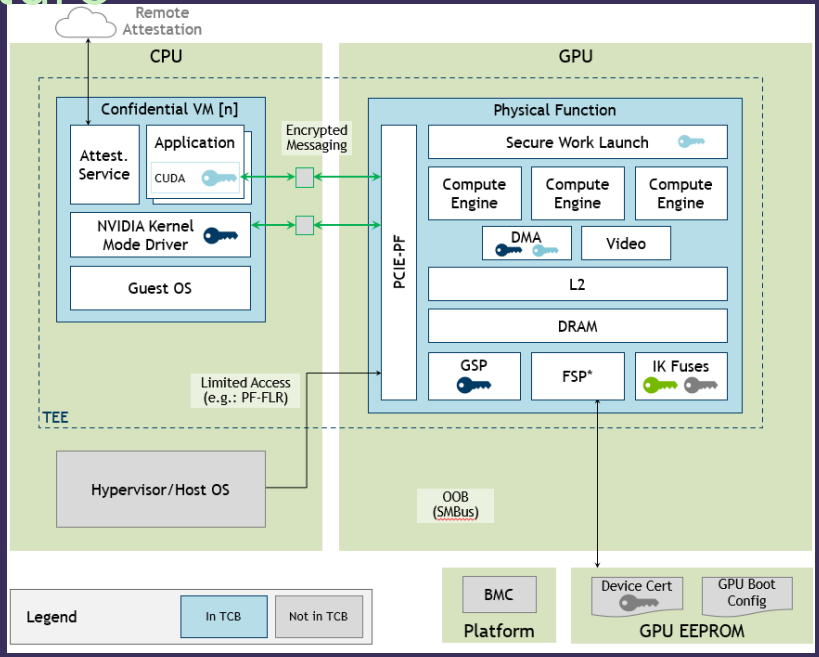
ITA Client encapsulates steps 2-5.

2. ITA Client calls TDX Driver for Evidence
3. TDX Driver gets Evidence from Hardware
4. ITA Client sends Evidence to Intel Trust Authority
5. ITA responds with Attestation Token

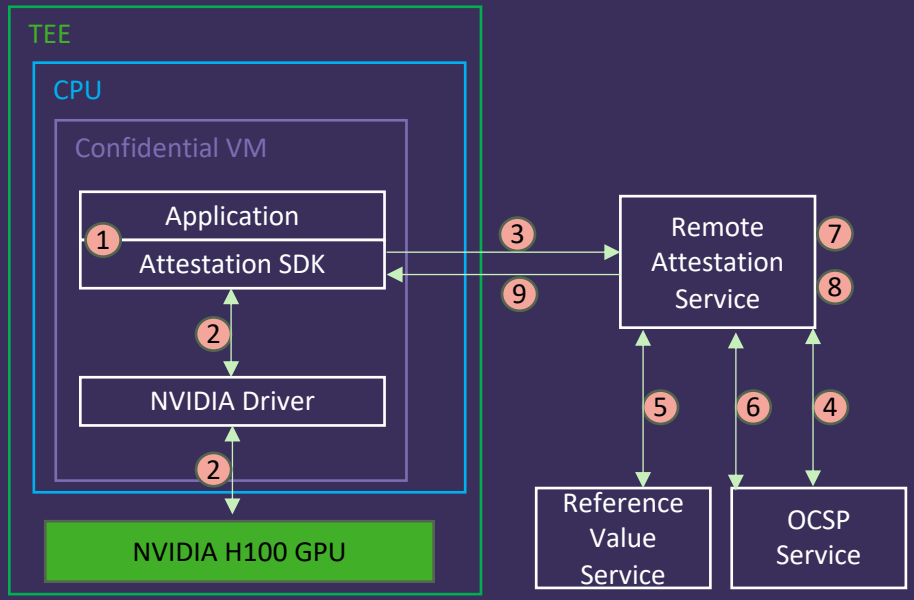
6. App provides Token to Relying Party

ITA – Intel Trust Authority

NVIDIA H100 Confidential Computing Architecture

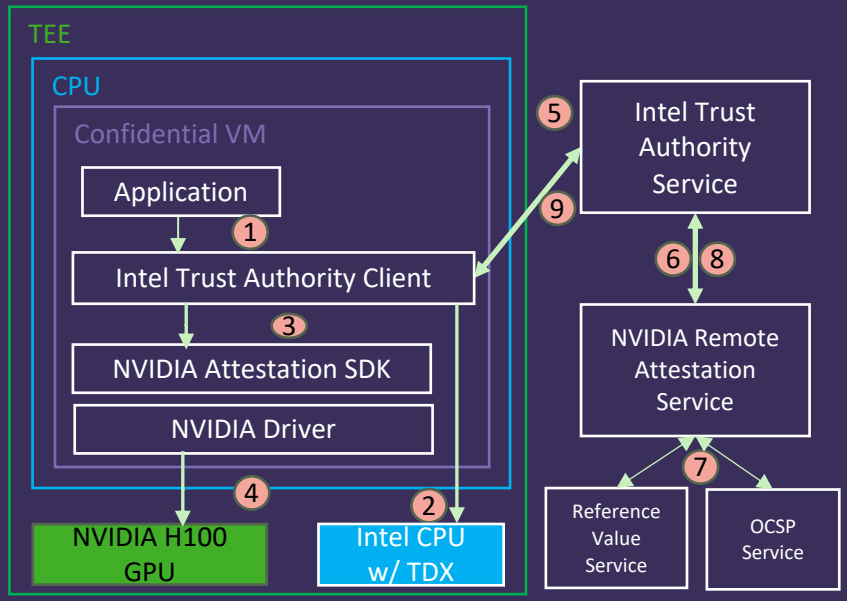


GPU Attestation Step-by-Step



1. Request attestation
2. Request signed GPU measurements
3. Send signed GPU measurements
4. Verify certificate of GPU measurements
5. Request golden measurements
6. Verify certificate of golden measurements
7. Compare measurements and generate report
8. Compare report against policy and generate token
9. Return token

Attestation Integration



1. Application requests attestation
2. Trust Authority client requests CPU report
3. Trust Authority client requests GPU report
4. GPU Driver requests report from GPU
5. Trust Authority client asks service to verify report
6. Trust Authority delegates GPU Verification to NRAS
7. GPU Service gets reference values and validates certificate chain
8. GPU Service verifies GPU report
9. Trust Authority Service generates composite token

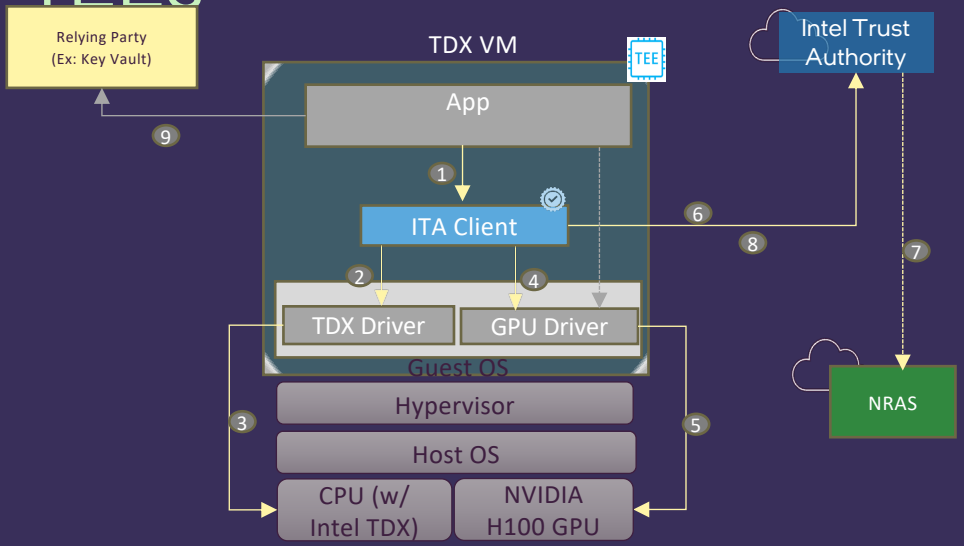
NVIDIA, Intel CC Summit 2023 announcement & Joint Blog (Q3 2023)

Intel and NVIDIA deliver Confidential Computing technologies that establish independent TEEs on the CPU and GPU, respectively. For a customer, this presents an attestation challenge, requiring attestation from two different services to gather the evidence needed to verify the trustworthiness of the CPU and GPU TEEs.

Through this collaboration, Intel and NVIDIA are providing a unified attestation solution for customers to verify the trustworthiness of the CPU and GPU TEEs for Confidential Computing based on Intel Xeon processors with Intel Trust Domain Extensions (Intel TDX) and NVIDIA Tensor Core H100 GPUs.

https://community.intel.com/t5/Blogs/Products-and-Solutions/Security/Seamless-Attestation-of-Intel-TDX-and-NVIDIA-H100-TEEs-with/post/1525587

Seamless Attestation of TDX and H100 TEEs



Highlights

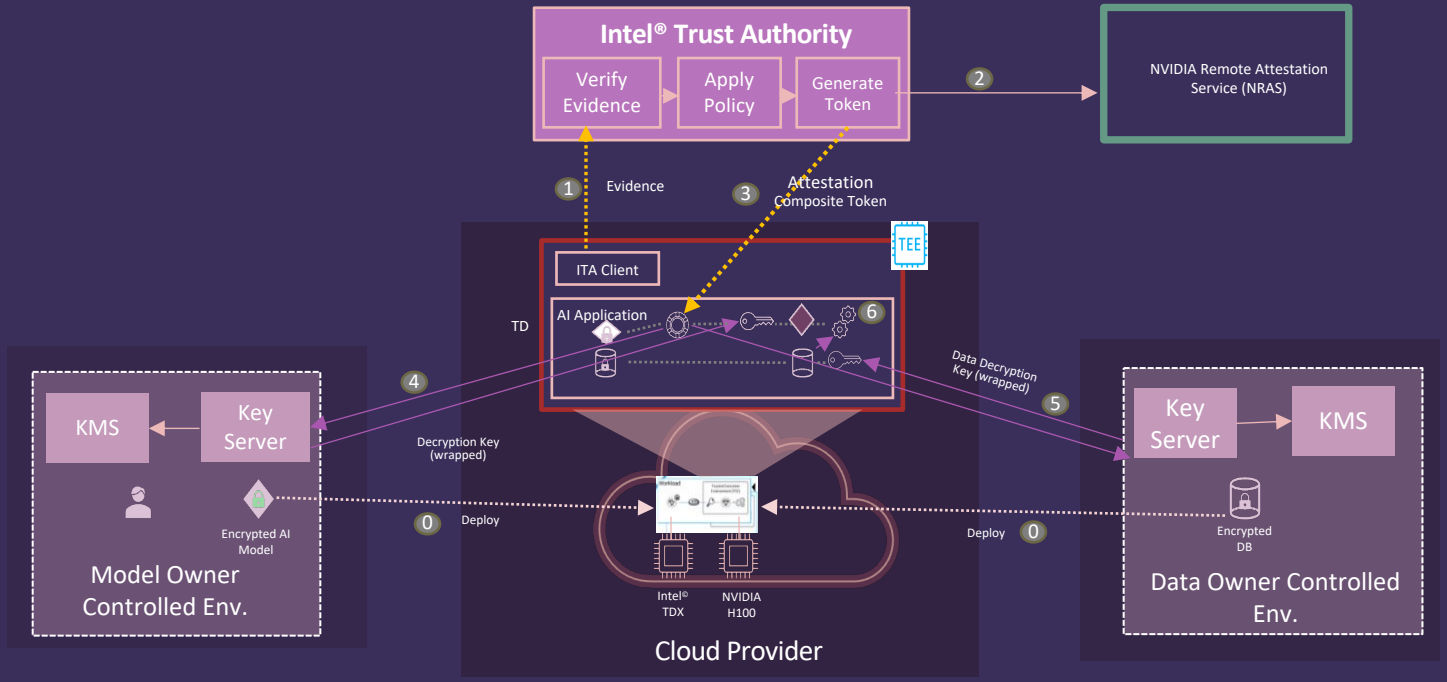
- Integrated attestation of TDX and H100 TEEs.
- One call to ITA Client
  - CollectCPUToken, CollectGPUToken, CollectCompositeToken
- Steps 2-8 done transparently by ITA Client.
- ITA Client - CLI or SDK (your choice)
- Declarative Attestation Appraisal Policies with ITA

ITA – Intel Trust Authority

Beta availability Q2 2024; GA and CSP specific availability planned for H2 2024

Confidential AI Example

with Intel TDX, NVIDIA H100, and Intel Trust Authority



Demo

Summary

- Confidential AI protects data and models using Intel and NVIDIA Confidential Computing technologies.
- Attestation provides irrefutable verification of trust worthiness of confidential computing environment.
- Seamless Attestation of TDX and H100 TEEs, with Intel Trust Authority (ITA).
- Beta availability in Q2’24, with GA planned for 2H’24

Contact us!

info@oc3.dev

Notices & Disclaimers

- Intel technologies may require enabled hardware, software or service activation.
- All product plans and roadmaps are subject to change without notice.
- No product or component can be absolutely secure.
- Your costs and results may vary.
-