ETH zürich

SECTRS
Secure & Trustworthy Systems Group

Lessons from a Decade of Confidential Computing

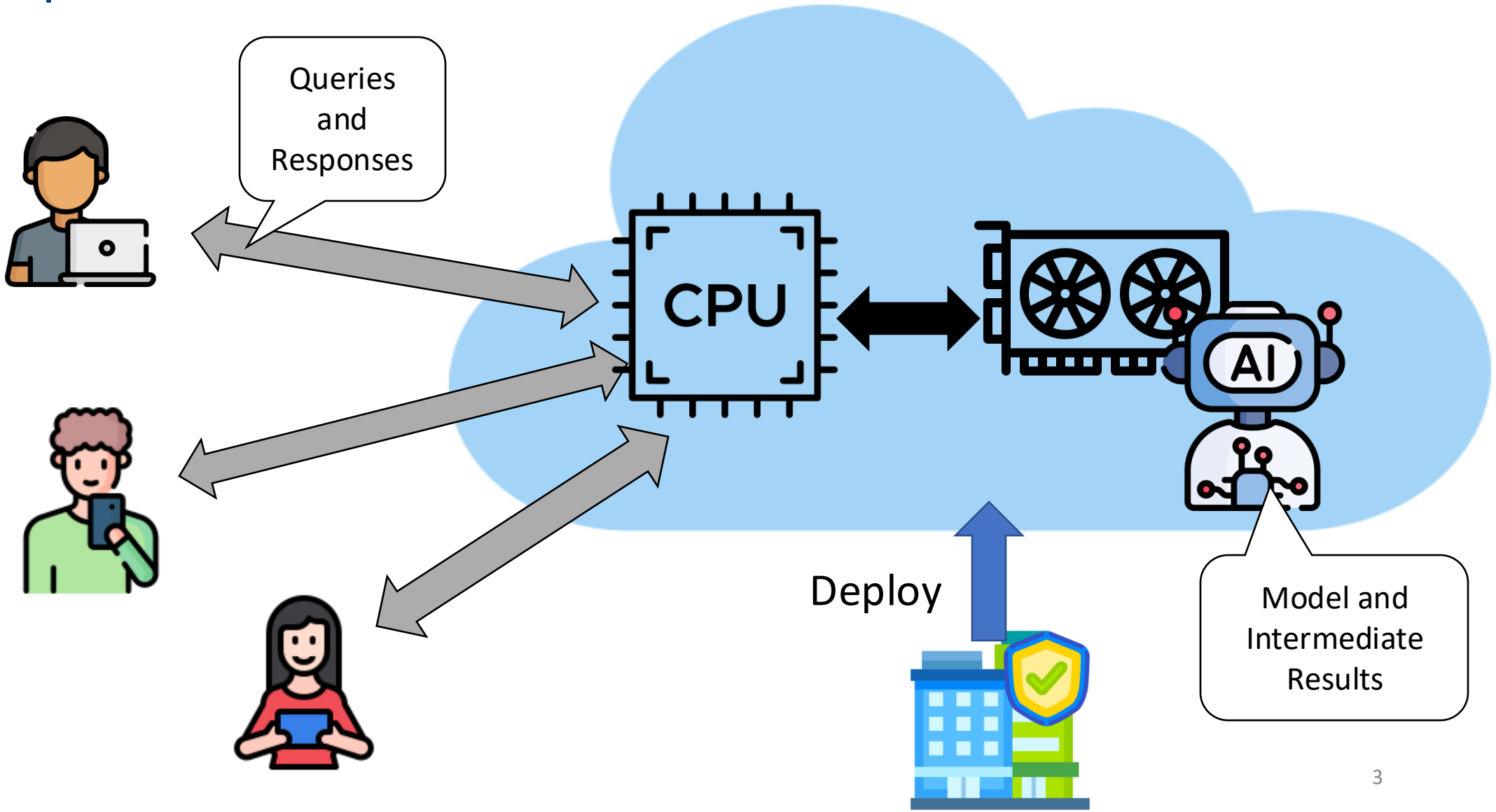
Shweta Shinde,
ETH Zurich

Secure & Trustworthy Systems Group

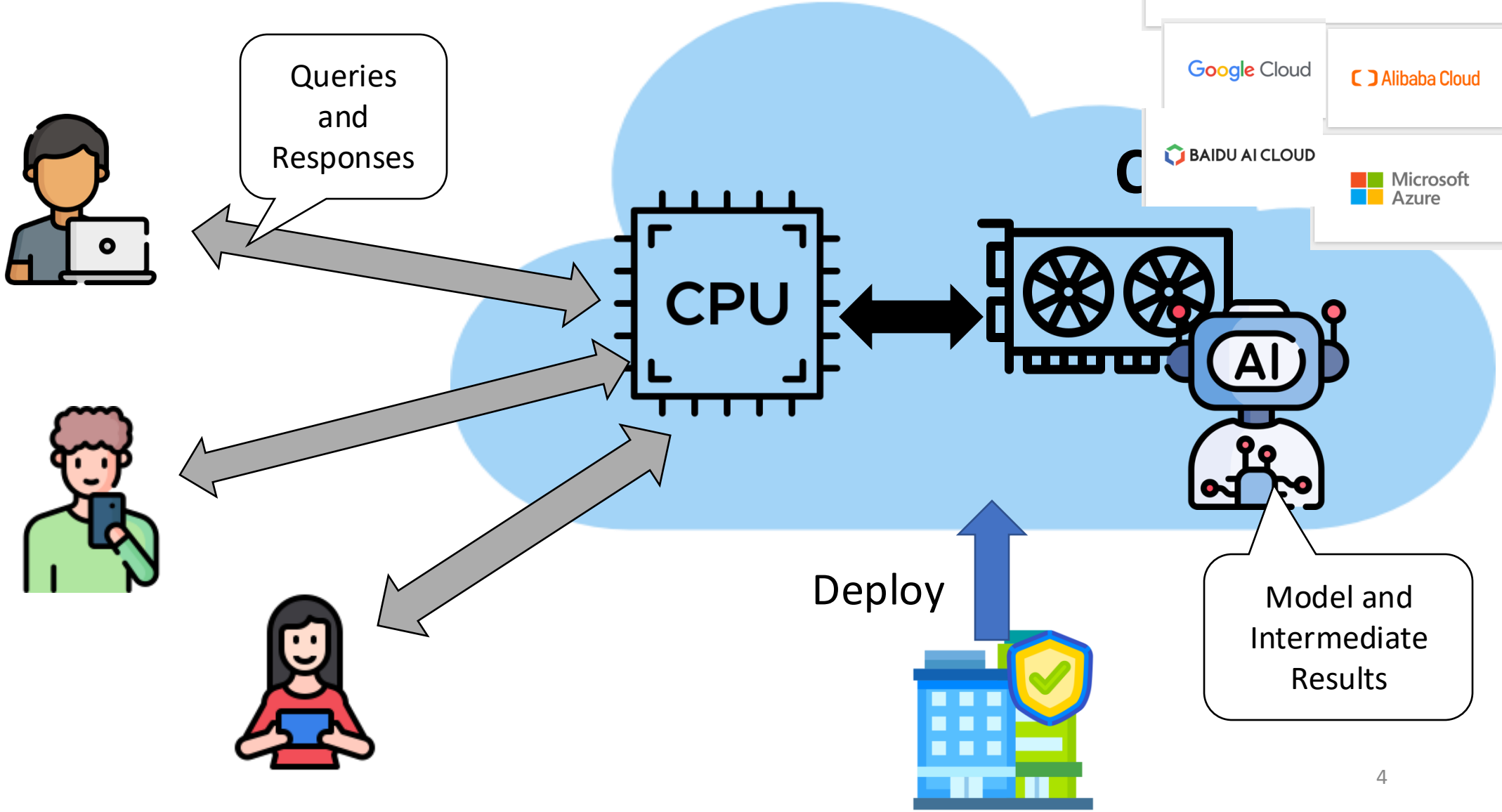
SECTRS
Secure & Trustworthy Systems Group

https://sectrs.ethz.ch/

Example: LLM Chatbot

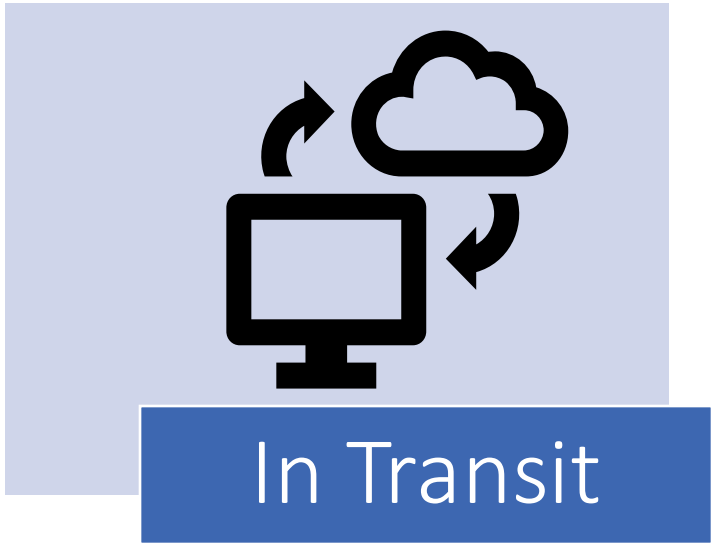


Confidentiality and Integrity

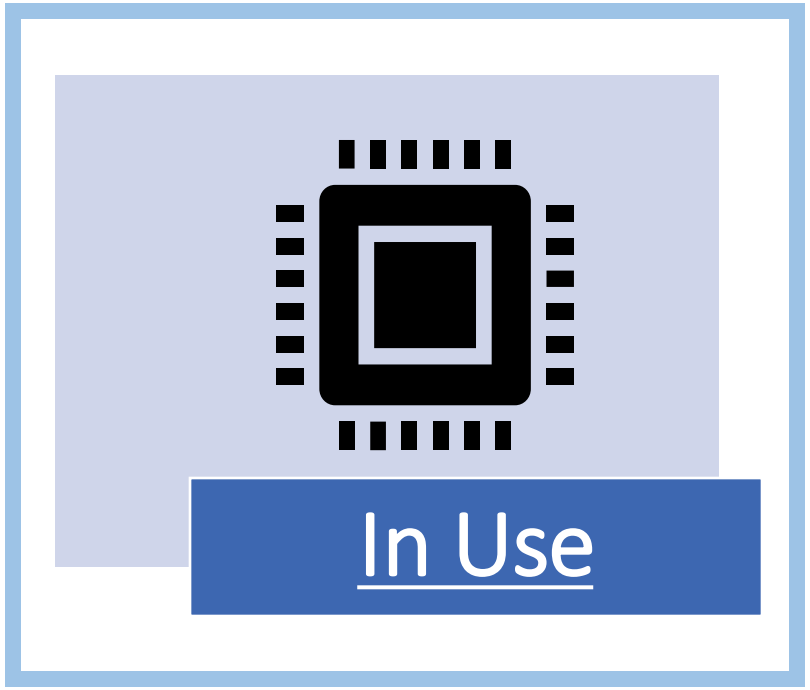


Attack surfaces for confidential data

At rest



In Transit

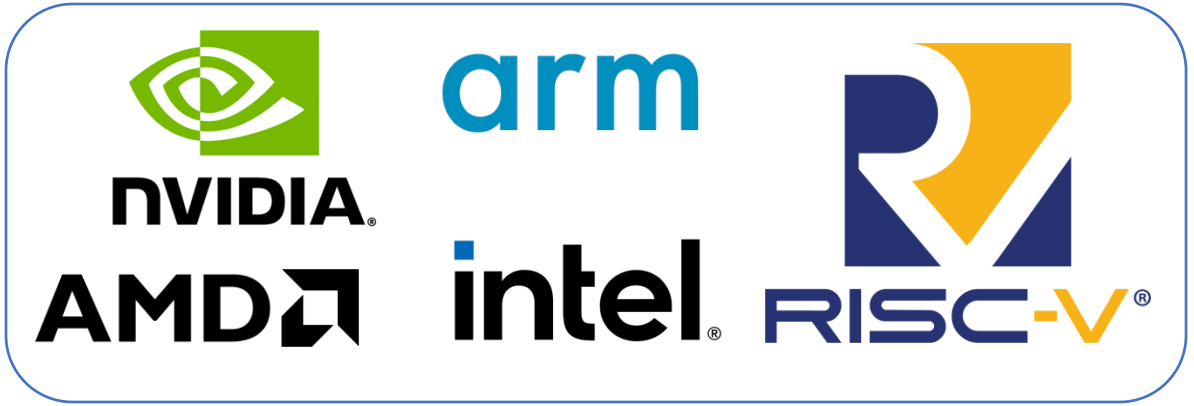


In Use

Trusted Execution Environments: Versatile Principle Applied to Real Systems

Sensitive Apps

Cloud Providers, Operating Systems



Servers, Mobiles, Sensor, GPUs, FPGAs, NICs

An incomplete history of the Evolution of Trusted Computing

ARM TRUSTZONE

AMD

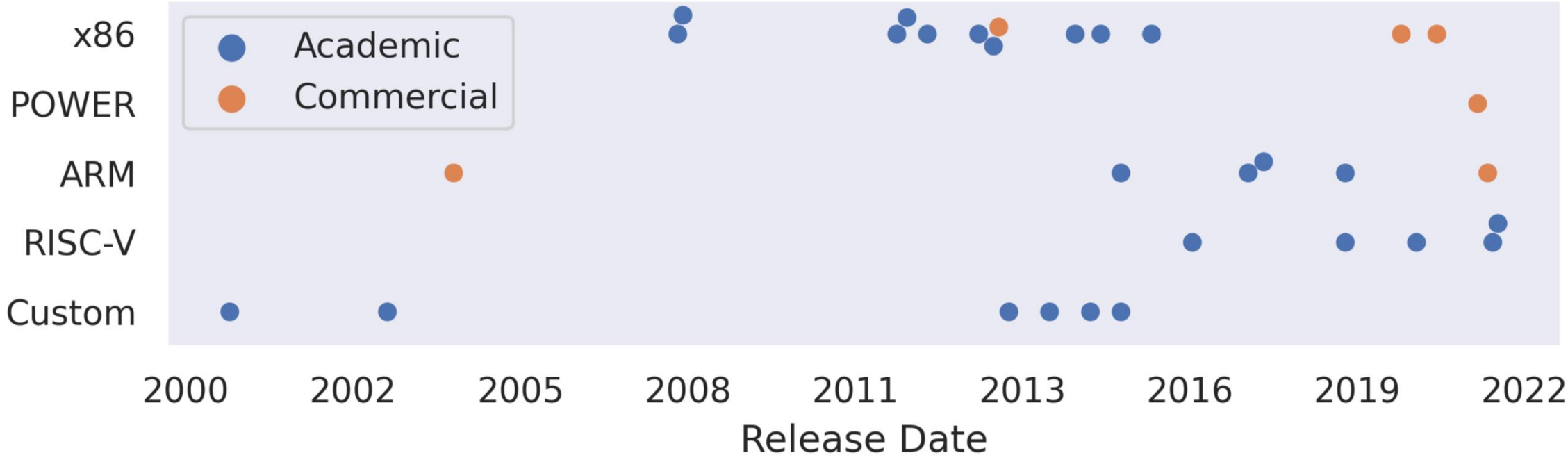
NVIDIA

arm

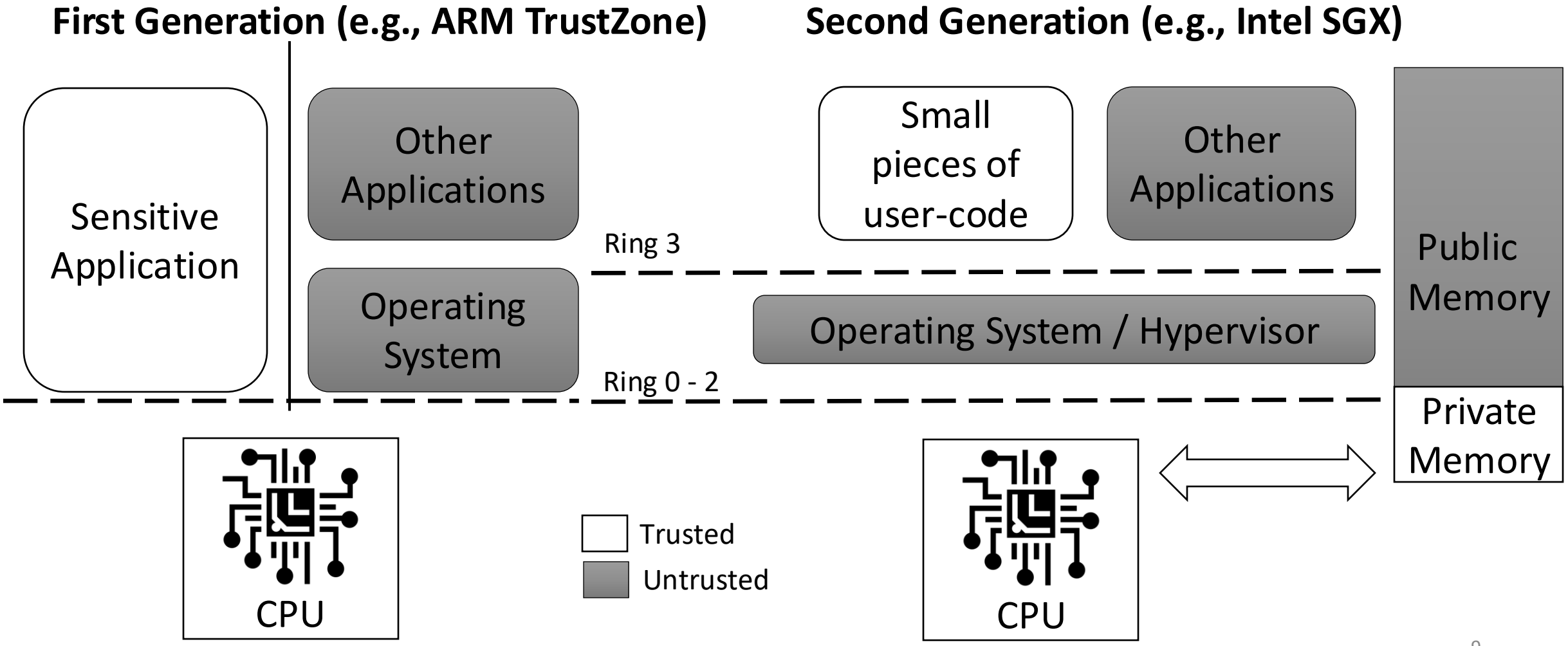
intel

2004	2005	2006	2014	2015	2017	2022	2023	
TPM 1.2	ARM TrustZone	Intel TXT	TPM 2.0	Intel SGX	AMD SEV	NVIDIA CC	Intel TDX	ARM CCA

An incomplete history of the Evolution of TEEs



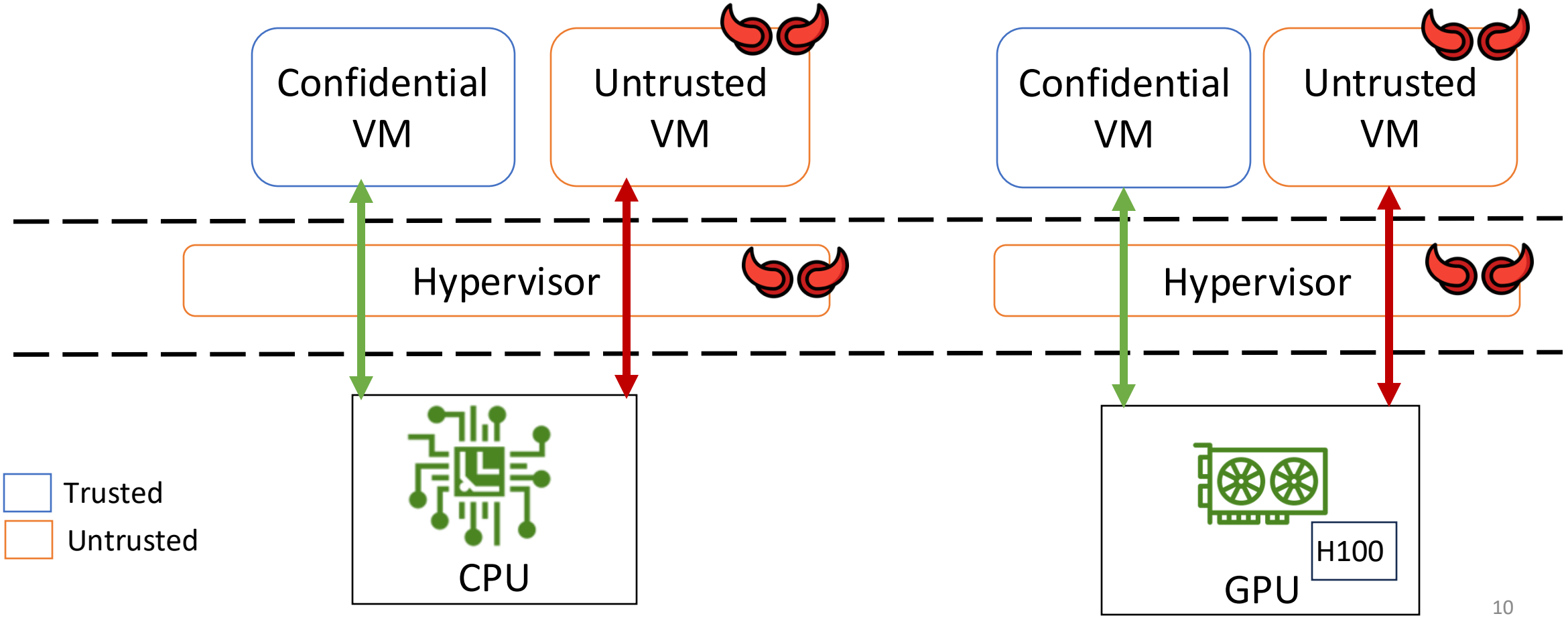
Past Trusted Execution Environments



Present Trusted Execution Environments

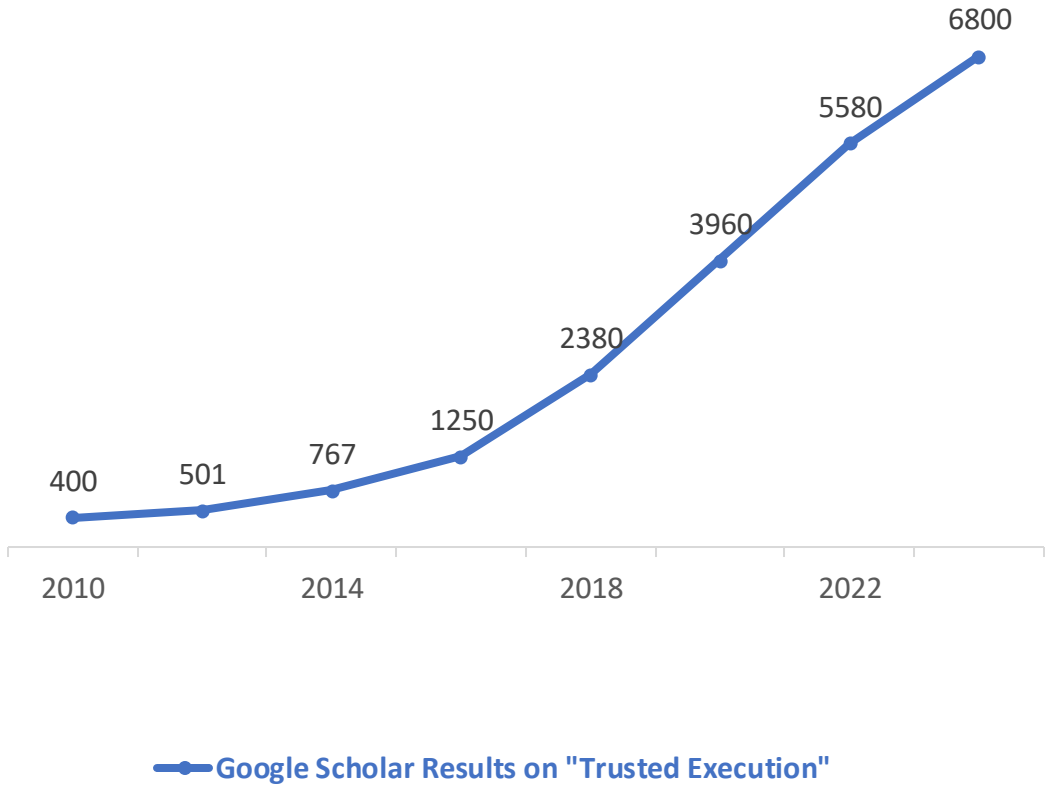
New Generation CPUs
(e.g., AMD SEV-SNP, Intel TDX, Arm CCA)

New Generation GPUs
(e.g., NVIDIA CC)



A Decade of Confidential Computing

- 2015: Intel rolled out SGX, Azure coined the term Confidential Computing
- 2019: Confidential Computing Consortium (CCC) was formed
- 2023: NVIDIA Confidential mode for H100



From Concept to Adoption

For General Compute

- Key Management
- Financial Services
- Data Analytics
- Healthcare
- Secure Blockchain
- Secure Multi-party Computation

For Smartphones

POSTED ON APRIL 29, 2025 TO SECURITY & PRIVACY

Building Private Processing for AI tools on WhatsApp

oppo

Newsroom All Press Release Stories Intellectual Property

How OPPO and Google Are Redefining Mobile AI with Seamless Integration and Enhanced Security

Stories · Mar 04, 2025

Security Research Overview Blog Bounty Research Device Submit a Report

Blog

June 10, 2024

Private Cloud Compute: A new frontier for AI privacy in the cloud

Written by Apple Security Engineering and Architecture (SEAR), User Privacy, Core Operating Systems (Core OS), Services Engineering (ASE), and Machine Learning and AI (AIML)

For AI

Confidential Inference Systems

Design principles and security risks

Version 1.0, June 2025

AZURE CONFIDENTIAL COMPUTING BLOG 14 MIN READ

Azure AI Confidential Inferencing: Technical Deep-Dive

MarkRussinovich MICROSOFT

Sep 24, 2024

May 3, 2024 Security

Reimagining secure infrastructure for advanced AI

OpenAI calls for an evolution in infrastructure security to protect advanced AI

Why did confidential computing take off?

Cloud Computing: Adoption → Security Risks

Key Growth Period when cloud became commercially viable

- **2006:** AWS launched EC2
- **2008–2010:** Google, Microsoft, IBM expanded cloud offerings

Widespread Adoption

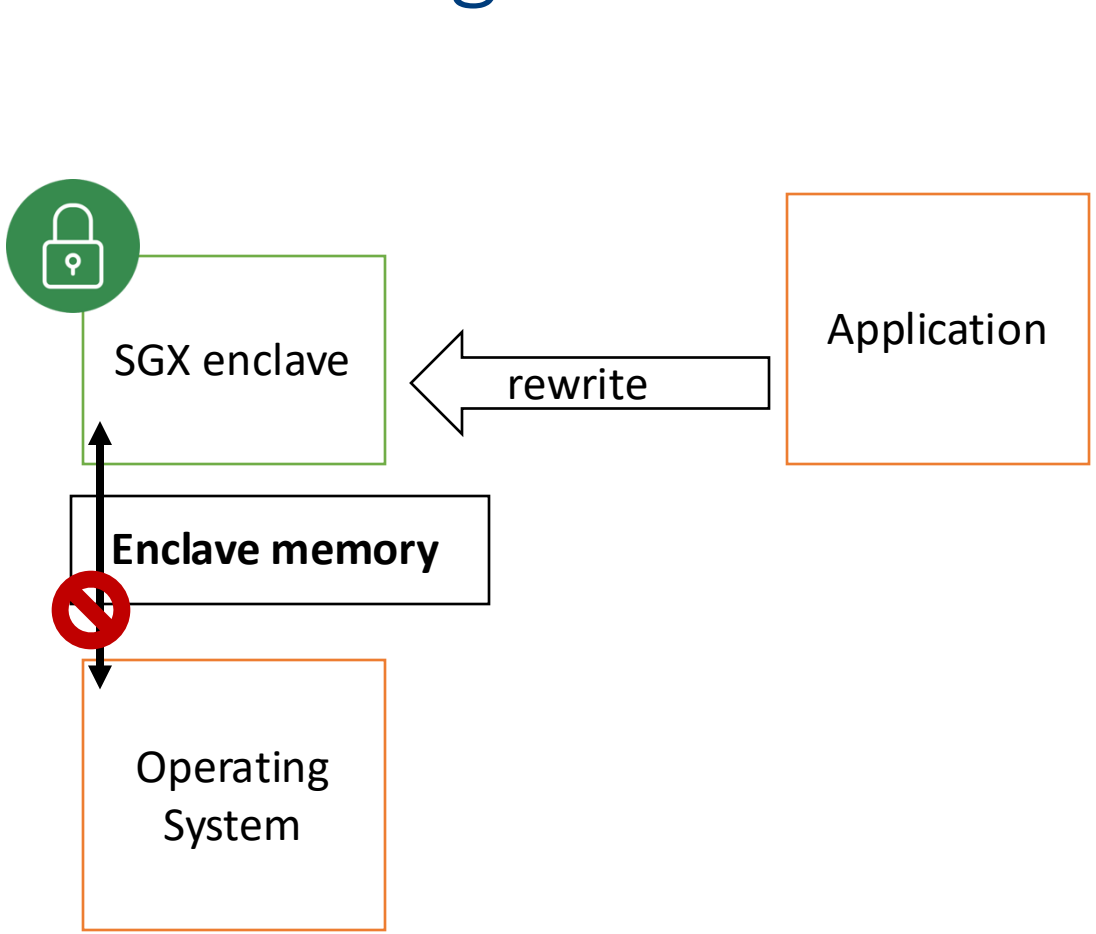
- **2010s – Present:** Cloud became essential for modern businesses
  Cloud-native applications, serverless computing, edge computing, AI
  Increased focus on security and compliance

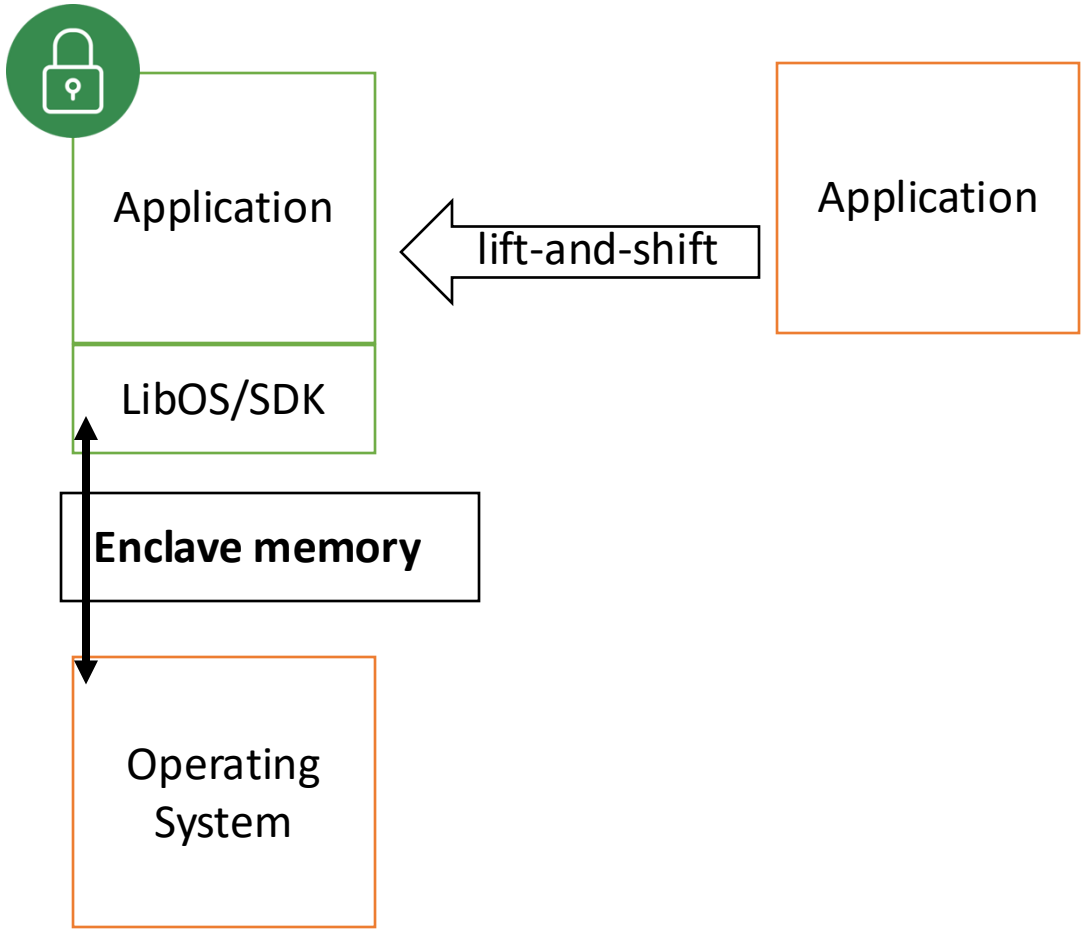
Intel SGX kick-started a revolution

But… SGX was not designed

to protect existing applications	• Library OSes and SDKs • Implementation bugs in the TCB • Iago attacks through untrusted interfaces
to protect against side-channels	• 2015: Controlled-Channel Attacks • 2017: Meltdown, Spectre, and an era of speculative side-channels
for cloud workloads	• Limited to 92MB of physical RAM

Learning from Intel SGX





Virtual Machines: A cloud-native abstraction

Shift from Enclaves to *Confidential VMs*

New Abstractions

Isolate VMs instead of processes

- Easy to lift-and-shift workloads
- Cloud-native abstraction
- Performance

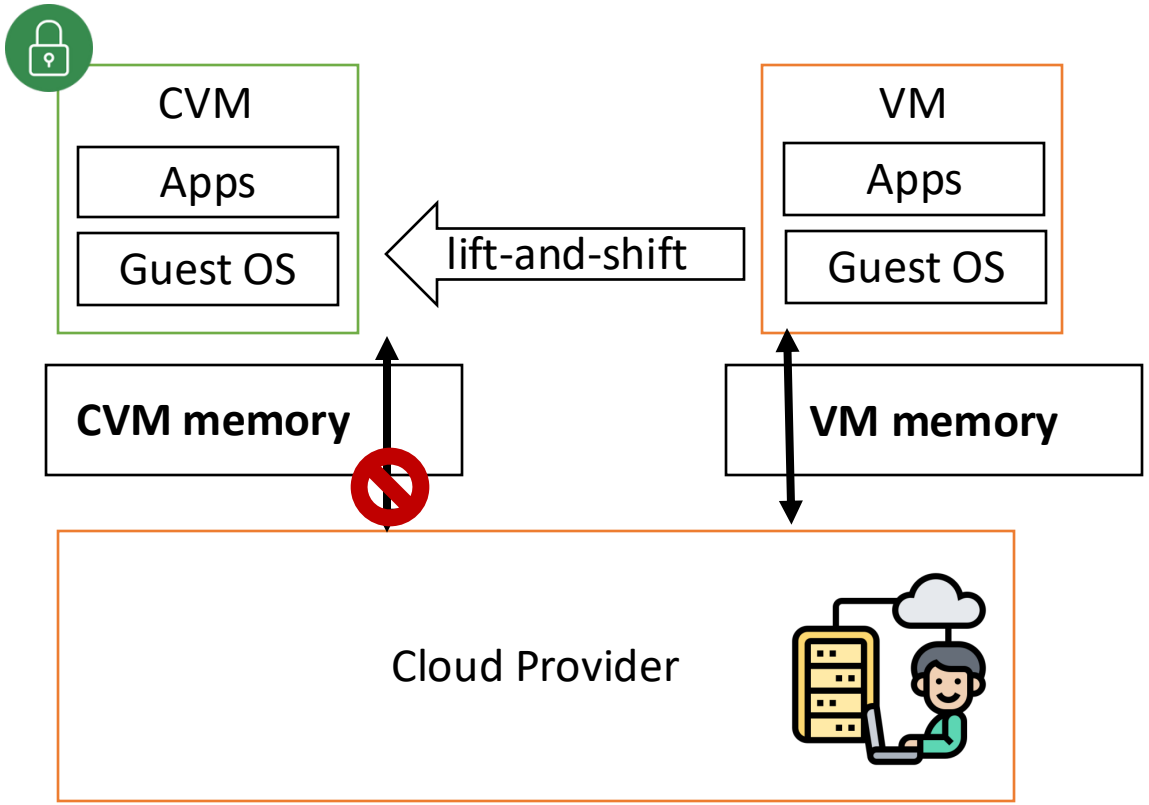
Beyond the CPU

Expand the isolation to devices

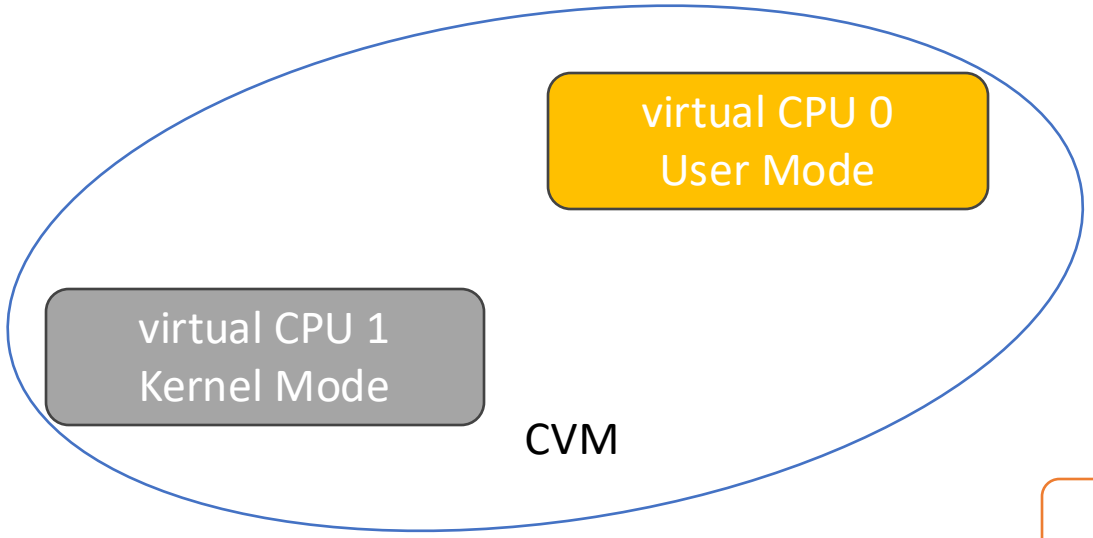
- (Custom) accelerators are vital
- Allocated at VM granularity
- Performance

Confidential VMs are Built for Lift-and-Shift

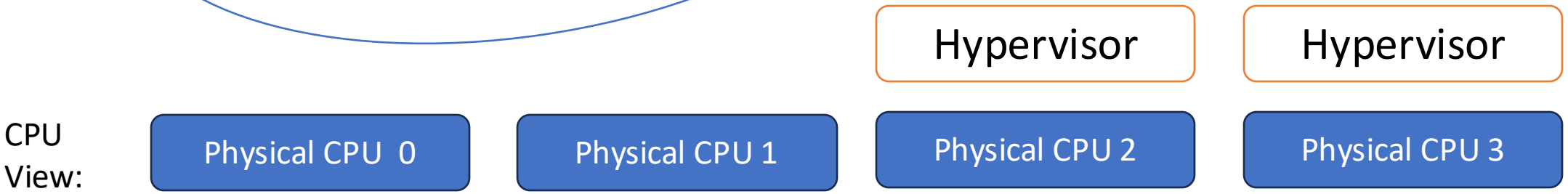
- Offered by AMD SEV-SNP, Intel TDX, Arm CCA, RISC-V CoVE



Confidential VMs: CPU & Memory Abstraction



Hypervisor manages,
Hardware isolates



But… did CVMs solve what Enclaves couldn't?

existing ~~applications~~ VMs

- ~~Library OSes and SDKs~~
- Implementation bugs in the TCB
- Iago attacks through untrusted interfaces

Whole guest OS

Hypervisor

side-channels

- 2015: Controlled-Channel Attacks
- 2017: Meltdown, Spectre, and and era of speculative side-channels

New side-channels

cloud workloads

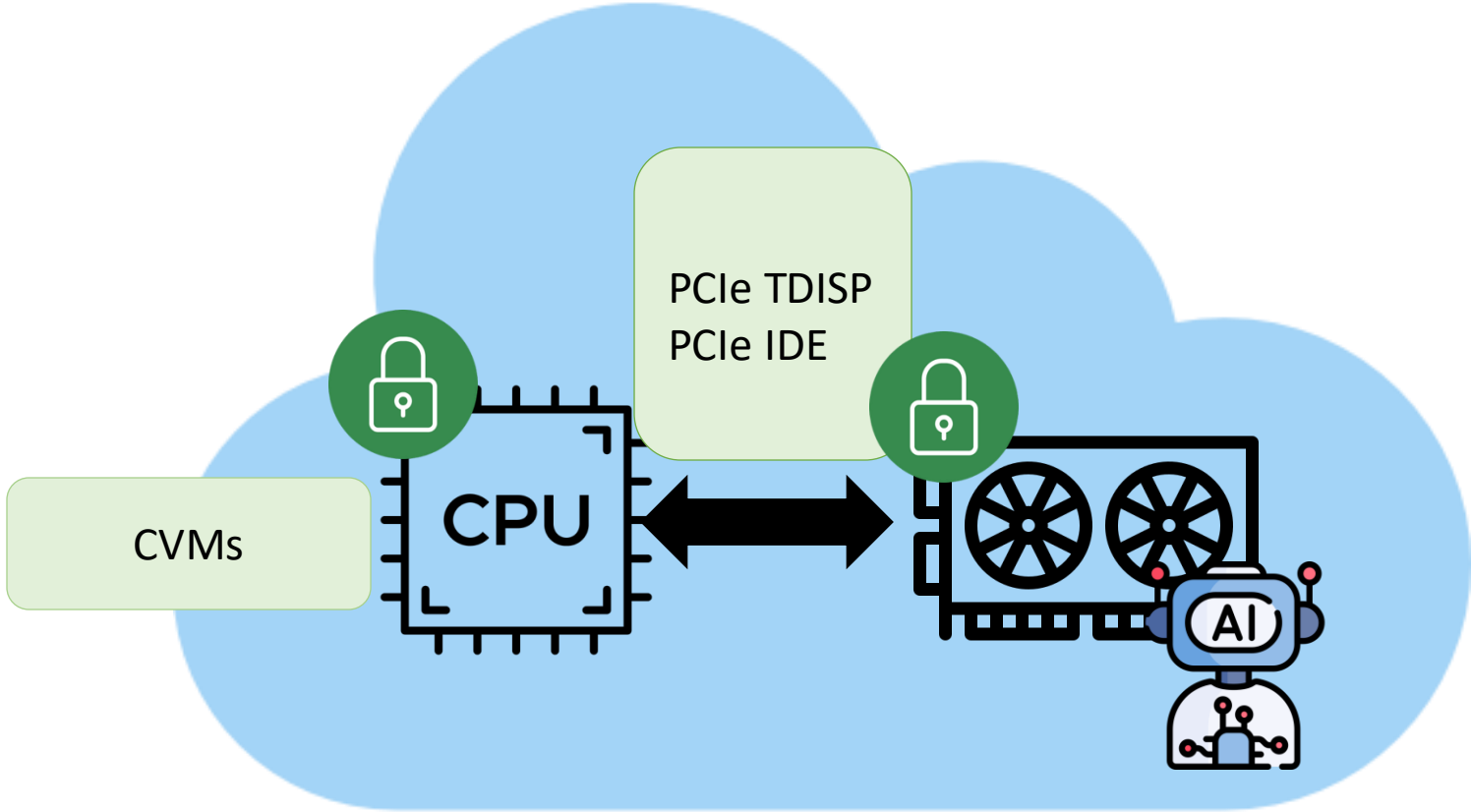
- ~~Limited to 92MB of physical RAM~~

Removed Merkle tree

Not just CPUs

Beyond CPU-based Protection

Adding TEE support to Accelerators



Nvidia’s Confidential Computing on Hopper GPUs

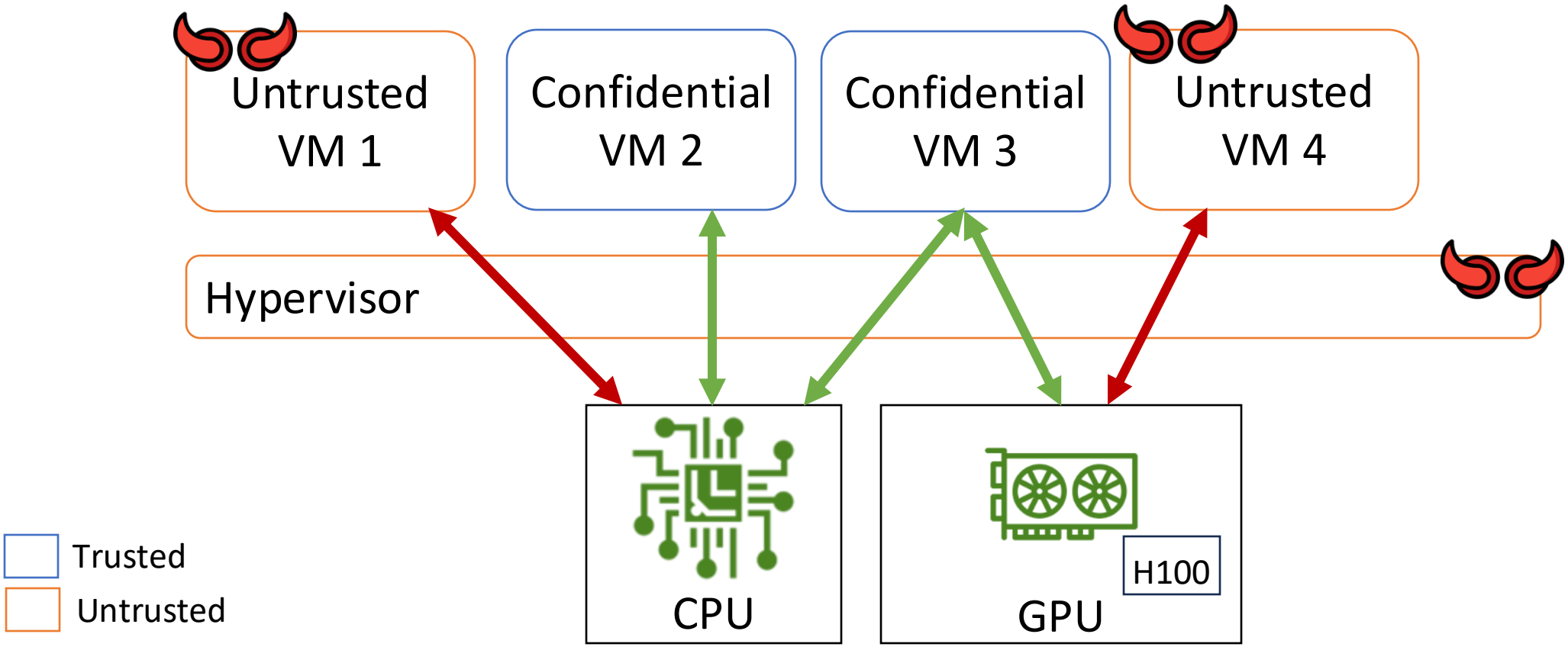
Graviton: TEEs on GPUs

Confidential Machine Learning within Graphcore IPUs

Ascend-CC: CC on Heterogeneous NPU for Emerging Generative AI Workloads

And many more academic works…

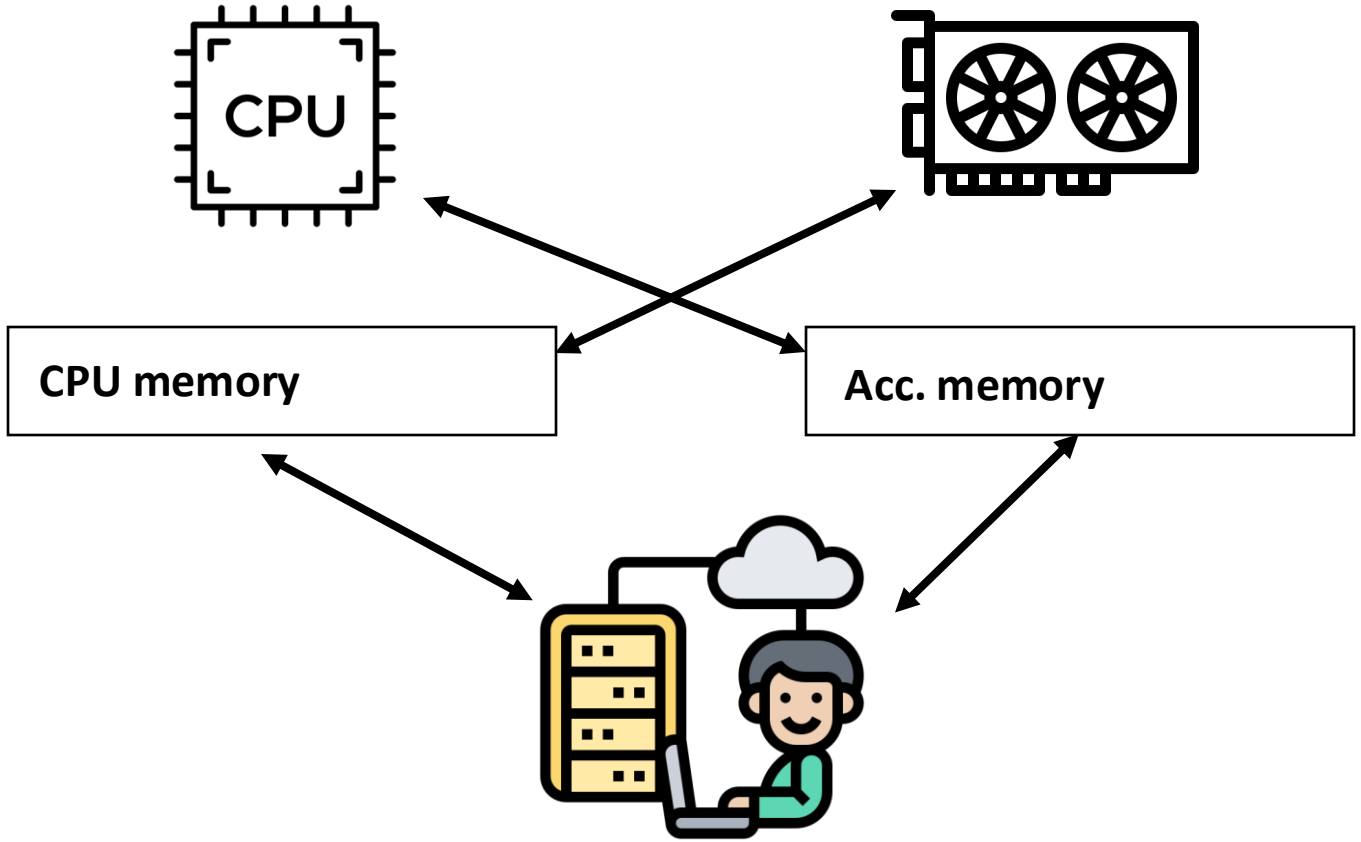
Securely Composing CPU & Accelerator TEEs



Infrastructure is Optimized by the Cloud Provider

Fast interconnects:
CXL, RDMA, etc.

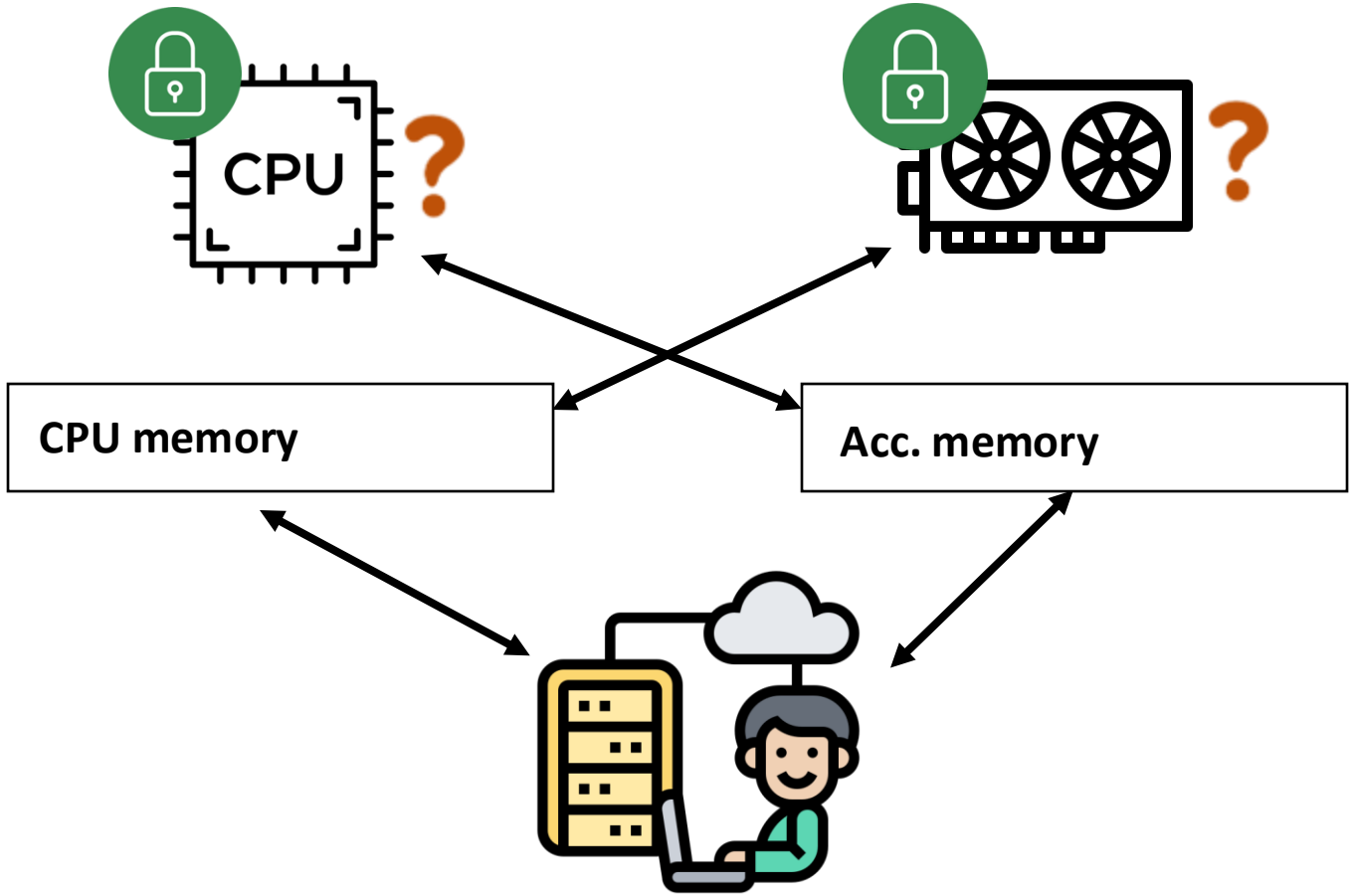
Optimizations:
kernel bypass, copy-on-write, interrupt delivery



Untrusted Cloud Provider and Hardware

Fast interconnects :
CXL, RDMA, etc.

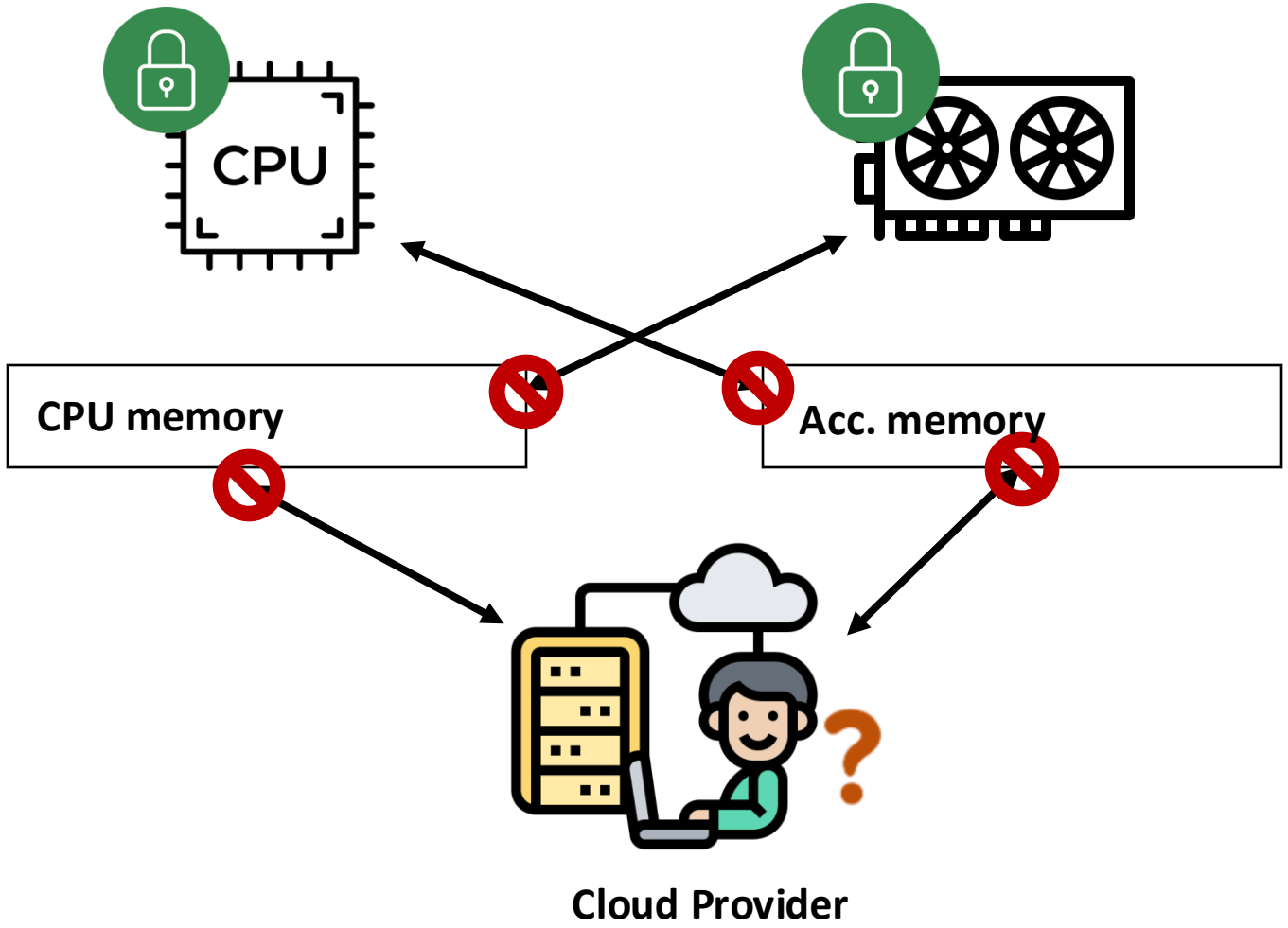
Optimizations:
kernel bypass, copy-on-write, interrupt delivery



Confidential Computing Blocks Memory Accesses

Fast interconnects :
CXL, RDMA, etc.

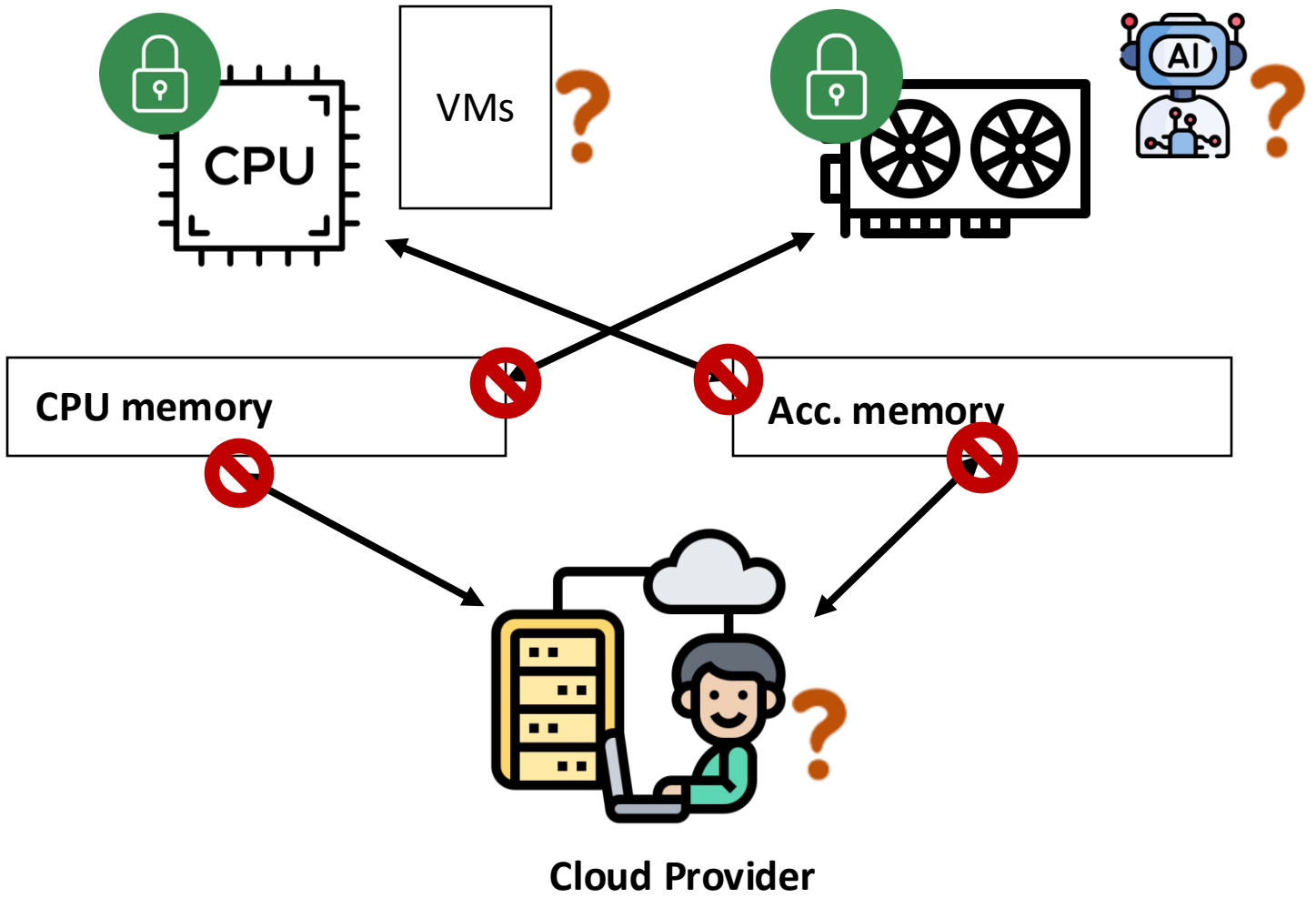
Optimizations:
kernel bypass, copy-on-write, interrupt delivery



What About the Software and Optimizations?

Fast interconnects :
CXL, RDMA, etc.

Optimizations:
kernel bypass, copy-on-write, interrupt delivery



Balancing the Trifecta

What was trusted before, isn't anymore

Lift-and-Shift Can Be Dangerous

Sharing CPU & Accelerator Memory

Naïve life-and-shift

Where do we go from here?

1. Modular, minimal, and verified designs: Scale better and easier to analyze

2. Protect every interface, verify every assumption

3. Formal guarantees are achievable, but rare

4. Confidential Computing is a force for user empowerment

5. Be willing to build for the future, not just patch the present

6. Build big systems as research vessels, not quick publications

7. Extract lessons, not features

8. Papers that educate, backed by systems that encountered real challenges and validated aspiration

Summary

- Confidential computing is a reincarnation of trusted execution environments
- Decades of research and development are coming to fruition
- Right place right time for AI revolution
- Need to proceed with caution and research the nuances

ETH zürich

SECTRS | Secure & Trustworthy Systems Group

Thank you

https://shwetashinde.org/
shweta.shinde@inf.ethz.ch